The Effect of Simultaneously Presented Words and Auditory Tones on Visuomotor Performance

Rita Mendonça[1*], Margarida V. Garrido[2], and Gün R. Semin[1,3]

[1]William James Center for Research, ISPA – Instituto Universitário
Rua Jardim do Tabaco, 34, 1149-041 Lisboa, Portugal

[2]Iscte - Instituto Universitário de Lisboa, Cis-Iscte
Av. Das Forças Armadas, 1649-026 Lisboa, Portugal

[3]Faculty of Social and Behavioral Sciences, Utrecht University
3584 CS Utrecht, The Netherlands

[*] Correspondence concerning this article should be addressed to Rita Mendonça, William James Center for Research, ISPA – Instituto Universitário, Rua Jardim do Tabaco, 34, 1149-041 Lisboa, Portugal (e-mail: rmendonca@ispa.pt)

Abstract

The experiment reported here used a variation of the spatial cueing task to examine the effects of unimodal and bimodal attention orienting primes on target identification latencies and eye-gaze movements. The primes were a nonspatial auditory tone and words known to drive attention consistent with the dominant writing and reading direction, as well as introducing a semantic, temporal bias (past-future) on the horizontal dimension. As expected, past-related (visual) word primes gave rise to shorter response latencies on the left hemifield and future-related words on the right. This congruency effect was differentiated by an asymmetric performance on the right space following future words and driven by the left-to-right trajectory of scanning habits that facilitated search times and eye-gaze movements to lateralized targets. Auditory tone prime alone acted as an alarm signal, boosting visual search and reducing response latencies. Bimodal priming, i.e., temporal visual words paired with the auditory tone, impaired performance by delaying visual attention and response times relative to the unimodal visual word condition. We conclude that bimodal primes were no more effective in capturing participants' spatial attention than the unimodal auditory and visual primes. Their contribution to the literature on multisensory integration is discussed.

Keywords

Attention, spatial agency bias, multisensory integration, vision, audition

Introduction

This research was designed to examine how distinct stimuli, namely nonspatial auditory tones and visual temporal words, can modulate visual attention when presented concurrently. Research has produced conflicting reports on whether bimodal cues capture attention more effectively than unimodal cues (see Spence and Santangelo, 2009 for a review). Audiovisual bimodal cues seem to delay visual processing relative to unimodal events (Driver and Spence, 1998; Mcgurk and Macdonald, 1976; Spence and Driver, 1997) because task demands increase and processing is equally impeded in both modalities (Robinson et al., 2018; Shams et al., 2002). However, other research has shown that auditory tones temporally colocalized with visual targets drastically reduce visual search latencies (Dalton and Spence, 2007; Van der Burg et al., 2008; Vroomen and De Gelder, 2000).

The current study examines whether bimodal primes composed of a visual word and an auditory tone facilitate attention and target discrimination over their single unimodal presentation. We used a bimodal cueing task with time words priming horizontal locations consistent with the semantic indication (past-left/future-right in 'Western' languages, Lakens et al., 2011) and biased by the culturally-defined reading and writing direction (Suitner and Maass, 2016), and a spatially neutral auditory prime previously shown to accelerate visual search towards the location of a synchronized visual event (Ngo and Spence, 2010; Vroomen and De Gelder, 2000).

The experimental task is inspired by previous research but introduces a manipulation in which auditory and visual priming stimuli were presented concurrently. We hypothesized that co-occurring modalities would support each other to produce multisensory integration that would enhance motor and gaze performance over single modality trials. In bimodal trials, the co-occurring auditory tone should enhance the attentional and motor effects generated by the spatial grounding of words, particularly for 'rightward' time words (i.e., future-related) that are aligned with the habitual language direction.

In the following, we provide a brief overview of the literature on the spatial biases induced by reading and writing habits and the semantic, temporal bias words. We then turn to the literature on the attention-capturing features of auditory tones and multimodal stimuli. After concluding the introduction, we provide an overview of the research.

Contributors to attentional and cognitive biases

Culturally habitualized routines constitute one of the major influences on how a substantive range of processes comprising representation, attention, and cognition are shaped. In particular, reading and writing are the typical activities that fall into the category of habit formation. The influence of these activities can be in opposing directions as in the case of, for example, European languages (left to right reading-writing) versus Arabic linguistic communities and Hebrew (right to left). Consequently, these cultural habits can give rise to systematic representational, attentional, and cognitive biases. Similarly, specific conceptual categories, such as time (Lakens et al., 2011; Majid et al., 2013; Santiago et al., 2007), but also politics (Farias et al., 2016; Mills et al., 2015), or numbers (Shaki et al., 2009; Zebian, 2005) can exercise an influence in the way attention is directed. In most Western communities, time is anchored on a horizontal dimension evolving from left to right (past/left versus future/right; e.g., Ouellet et al., 2010a; Santiago et al., 2010).

Notably, multisensory primes produce additive effects over unimodal primes. These lead to gains in attention orienting because the conjugation of, for instance, two modalities intensifies attentional effects (Alvarado et al., 2007; Teder-Sälejärvi et al., 2002). One special case is when a spatially neutral auditory signal is paired with a visual target (Talsma et al., 2010; Van der Burg et al., 2008). In short, we have three types of attention driving factors that overlap:

1. The dominant reading and writing system shapes how attentional direction unfolds on a horizontal trajectory influencing among other things psychological variables such as agency and movement (e.g., Spatial Agency Bias, Suitner and Maass, 2016);
2. The grounding of specific abstract concepts such as time on a horizontal space (Fuhrman and Boroditsky, 2010; Lakens et al., 2011; Ouellet et al., 2010b);
3. Nonspatial auditory signals that have been shown to improve detection of a synchronously presented visual target (e.g., Dalton and Spence, 2007; Van der Burg et al., 2008; Vroomen and De Gelder, 2000)

In the following, we detail briefly the empirical evidence for the three attention driving factors, as well as for their integration in multimodal research. Before proceeding with the three sources driving attention, we should briefly mention presumed biological influences on spatial attention. It has been argued that in spatial tasks, the right hemisphere is dominant in generating a predisposition to attend to the left hemispace (Brooks et al., 2014).

For instance, neurologically normal individuals bisect a horizontal line by erring slightly to the left of a line's midpoint (i.e. pseudoneglect) (see Jewell and McCourt, 2000 for a review). In free exploration of images, the initial bias has been found to the left visual field (Ossandón et al., 2014), and asymmetrical scanning of visual space has been reported (Butler and Harvey, 2006). The biologically determined account has however been challenged by studies with native readers of right-to-left languages (Afsari et al., 2018; Rashidi-Ranjbar et al., 2014) or bilingual populations with opposite script directions, who have shown minimal lateralization (Hernandez et al., 2017; Kermani et al., 2018).

Reading and writing systems

The influence of culturally anchored reading and writing direction has been shown across a wide range of psychological processes starting with asymmetrical scanning routines, to ascription of agency. Script direction guides attention because people typically start scanning space where writing begins. Eyes and hand progress together in space which creates a correlation between space and the occurrence of future information. The dominant writing system has been shown to shape the representation of action (Suitner and Maass, 2016). For example, participants from rightward flowing languages typically place the agent of a sentence to the left in thematic-role drawing tasks (Maass and Russo, 2003). These effects are also carried over to artwork aesthetic preferences (Chahboun et al., 2017; Smith et al., 2020), interpretation of films and football games (Maass et al., 2007), stereotypical representation of groups (Maass et al., 2009), gender categorization (Suitner et al., 2017), and memory (Bettinsoli et al., 2019). Reading and writing habits establish an enduring preferential scanning of space (Afsari et al., 2018; Chokron and De Agostini, 2000). Therefore, the anticipation of future information is facilitated when it coincides with script direction because people expect stimuli to flow in accordance with momentum (Hubbard, 2005).

Grounding the abstract concept of time

In contrast to reading and writing habits, the grounding of abstract concepts such as time lacks direct sensory-based experience. However, across most cultures time is anchored spatially (Bergen and Lau, 2012; e.g., Boroditsky et al., 2011; Tversky et al., 1991). In cultures with left-to-right orthography, time is grounded by a rightward movement (e.g., European languages). In cultures with right-to-left orthography (e.g., Arabic/Hebrew-speaking cultures) time moves leftward (Fuhrman and Boroditsky, 2010). Although there are other cultural variations on how time is grounded (e.g., time is mapped in cardinal directions, Boroditsky and Gaby, 2010; time is on a vertical axis, Lai and Boroditsky, 2013; time is uphill or downhill, Núñez et al., 2012), space is the common denominator, and directionality and location of past and future vary as a function of a culture's orthographic or ecological properties. Studies employing verbal information revealed that exposure to past and future terms in left-to-right orthography facilitates motor and attentional responses to the left and right space (Ouellet et al., 2010a; Santiago et al., 2007; Torralbo et al., 2006; Weger and Pratt, 2008). The same effect emerges when time categories are presented auditorily irrespective of orthographic directionality (Ouellet et al., 2010b). This research shows that not only the reading and writing convention for text direction guides temporal mappings, but that temporal words are effective media to prime scanning directions as a function of time-space associations.

Auditory primes as attention drivers

Aside from the two aforementioned attention driving endogenous factors, exogenous ones such as auditory warning signals (Ho and Spence, 2005) have attention capturing properties. The characteristics of exogenous primes is that they are typically non-semantic in nature (e.g., abrupt onset like a flash of light or a sound burst). They barely carry spatial information, but when they are delivered near the location of the impending target (in cued trials), target detection is facilitated. It has also been argued that exogenous primes increase alertness (Fernandez-Duque and Posner, 1997) since conditions with these primes produce faster responses than conditions without them. Additionally, nonspatial auditory tones when concurrently presented with visual targets increase the target's salience and guide attention towards a visual change (Ngo et al., 2012; Ngo & Spence, 2010; Van der Burg et al., 2008).

Multimodal primes and attention capturing

Research on the capacity for multimodal primes to retain attention-capturing capabilities has produced conflicting reports (see Spence and Santangelo, 2009 for a review). Simultaneous input from different sensory modalities seem to support each other to produce multisensory interactions that are additive, and which may alter cueing effects (Alvarado et al., 2007; Laurienti et al., 2005) even when multimodal information is redundant (Selcon et al., 1995). Multimodal events enhance responses of the superior colliculus neurons (a brain structure involved in the overt orienting of the eyes and head) above those evoked by the sum of the unimodal stimuli components (Wallace et al., 1998). However, research has struggled to demonstrate the bimodal advantage in behavioral tasks (Santangelo et al., 2006; Ward et al., 1998). In a seminal study, Ward (1994) presented audiovisual primes peripherally to participants and reported numerically larger (but not statistically significant) cueing effects over the most effective unimodal element. In an elevation discrimination task immune to stimulus-response overlap that Ward did not control for, Spence and Driver (1997) were not able to report an advantage of audiovisual primes to capture attention relative to the best unimodal primes. Others have also reported negligible differences in speeded responses between unimodal and bimodal presentations (Ho et al., 2009, Exp. 1; Santangelo et al., 2006; Santangelo and Spence, 2008).

Stimuli in one modality can alter the processing of other sensory input (Shams et al., 2002; Sloutsky and Napolitano, 2003). Most research has pointed to visual dominance in adults (see Spence et al., 2011 for a review). While vision is the dominant modality by default in spatial tasks (see Welch and Warren, 1986 for a review), there is evidence that audiovisual primes promote auditory dominance in auditory responses (Barnhart et al., 2018). The asymmetry in modality performance can owe it to task demands and the preparedness they instigate in the observer to process either modality (Robinson et al., 2018). This may cause distinct sensory input to compete for attentional resources and preclude super-additive effects to emerge.

The present study

The current research investigates whether bimodal, relative to unimodal, primes can efficiently capture attention and facilitate speeded responses to lateral targets. The novelty of this study is that it brings together distinct but interrelated research traditions on attention-orienting primes showing that: a) movement (i.e. agency) representation is affected by consistent exposure to script direction in a given culture, as are other categories grounded in the same left-right continuum (i.e. time) (Suitner and Maass, 2016); b) temporal language induces a semantic bias in visuomotor responses on the left/right coordinates (e.g., Ouellet et al., 2010a; Torralbo et al., 2006); c) nonspatial auditory tones co-occurring with visual events enhance visual search (Van der Burg et al., 2008); and d) multisensory stimuli interact to product additive cueing effects (e.g., Alvarado et al., 2007; Teder-Sälejärvi et al., 2002).

In a spatial cueing task, we examined the effects of unimodal (auditory tone alone; visual time-related word alone) and bimodal primes (concurrent presentation of visual word and auditory tone) on lateralized target identification. A condition without prime was also included. From the integration of the aforementioned research on the factors that guide attention, we have derived the following specific hypotheses:

H1: The spatial information of temporal words should drive attention to and produce faster identification of target positions congruent with their horizontal grounding (past word-left target, future word-right target).

H2: The pressure exerted by the rightward orthographic directionality (in European linguistic communities) should produce a unidirectional influence on attention orienting. This unidirectional influence, when combined with words that have a rightward connotation (i.e., future) is expected to benefit from this script habitualization. These combined biases are expected to accelerate responses on the right hemifield to a greater extent than past words on the left. That is, we should observe a semantic congruency effect (H1) but overall, a precedence of the orthographic directionality that differentiates between congruent presentations.

H3: The nonspatial auditory tone, when presented in isolation in the unimodal condition, would act as a warning signal and facilitate attention and target identification over the absence of target presentation.

H4: Without prejudice to asymmetric performance (H2), bimodal presentation (i.e., concomitant auditory and word primes) relative to unimodal visual presentation (i.e., word) is expected to produce an enhanced multisensory interaction that facilitates attention and speeds up target identification.

This study advances prior research in several respects. The bimodal prime elements are very different in their attention-grabbing properties and are not frequently paired in multimodal research. The bimodal primes combine a typical exogenous prime (i.e., auditory nonspatial tone) and a typical endogenous or symbolic prime (i.e., directional time word). The auditory tone was presented binaurally creating an essentially unlocalizable source (see Spence and Driver, 1997 for a similar example with loudspeakers). The words were meaning-based but not conventional like other spatial pointers. Aside from their modality, these primes could not conflict as the auditory tone is devoid of any inherent spatial information. This circumvents problems encountered in prior studies, for instance with conflicting directional information conveyed by the two simultaneous modalities. Thus, the study makes an empirical contribution to the literature on multimodal cueing, which has yet not been able to show whether there is a clear advantage of bimodal over unimodal primes.

Method

Participants

Participants were fifty-four undergraduate students who were compensated with course credit (M_{age} = 23.70, SD_{age} = 5.50). Sample size was determined based on the effect size proposed for temporal priming tasks (von Sobbe et al., 2019) for a repeated measures within-subjects ANOVA using BUCSS R method (Anderson et al., 2017) adjusting effects for uncertainty and publication bias (desirable level of assurance = .90; statistical power = .80). Participants were screened for normal visual acuity and reported no hearing problems. Five participants were excluded because they were nationals of countries other than Portugal. All the remaining participants (n = 49) were Portuguese native speakers and had not spent any significant time in right-to-left speaking countries. The experiment was approved by the Ethics Committee of the host institution and participants gave their informed written consent.

Primes

Attention orienting primes were four distinct kinds: no prime, auditory tone alone, visual time-related word alone, and visual time-related word co-occurring with an auditory tone. Visual stimuli consisted of 24 words from a list of Spanish time-relevant words previously used by other authors (e.g., Ouellet et al., 2010b; Torralbo et al., 2006). Words comprised 8 past- and 8 future-related words. The remaining 8 words were neutral and generated by us. Neutral words carried temporal meaning but lacked temporal directionality (see Appendix 1 for the extensive word list). The future and past word sets comprised 1 verb inflected in either the past (“it was” – “foi”) or future tense (“it will” – “será”), and 7 past (e.g., “before” – “antes”) or future temporal adverbs (e.g., ‘‘after’’ – “depois”). The neutral time-related word set was composed of nouns (e.g., “month” – “mês”). Four additional words were used for the practice trials. To establish their temporal meaning, the words were translated and piloted on a Portuguese speaking sample (n = 99). The words were individually and randomly presented to the participants at the center of the screen. Below each word, there was a horizontal line and a slider bar that participants were asked to place where they thought best represented the temporal meaning conveyed by the word (0 = *far* *past*; 100 = *far future*). Participants did not see the polar labels. They were presented with the horizontal line and a slider bar positioned on the scale's midpoint just below the word "present", which symbolized the present moment. Importantly, the words were selected such that they reflected different gradations of the temporal continuum, that is, they ranged from far past/future to immediate past/future. The words were grouped according to their mean ratings and their temporal orientation was found to be significantly different, F (2, 21) = 79.281, $p < .001$, ($M_{\text{past-oriented}} = 22.64$, $SE_{\text{past-oriented}} = 3.83$; $M_{\text{neutral}} = 56.73$, $SE_{\text{neutral}} = 1.49$; $M_{\text{future-oriented}} = 75.81$, $SE_{\text{future-oriented}} = 3.25$; p's $\leq .001$ between groups). Each word was presented four times throughout the experiment: twice alone, and twice paired with the auditory tone.

As auditory stimuli, we used a short non-informative auditory tone of 500-Hz tone (44.1 kHz sample rate, 16 bit, mono) with a duration of 60 ms (Van der Burg et al., 2008). The tone was binaurally presented via headphones either alone or concurrently with the visual word primes at the center of the screen, that is, auditory prime and visual prime had the same onset.

Targets

The targets were two strings of five letters simultaneously presented to the left and the right sides of the screen midpoint, at ±13.31° of eccentricity. This means the target sets were in the near peripheral left and right visual fields which ensured that participants could not discriminate the target unless gaze movements were made (i.e. overt attention). The target letter strings subtended 4.77° of visual angle. The target letter was either a *p* or a *q* embedded in one of the two letter strings (1 target and 4 distractors on one side, 5 distractors on the opposite side). The distractors were kept constant across the experiment. The distractors' and the targets' positions within the stream were varied randomly across trials. As there were two possible targets but only one correct answer per trial, the task required discrimination rather than mere detection. This setup allows us to rule out cueing effects as emerging from stimulus-response compatibility, a correspondence between target location and response effector. This is because the location of the target key on the gamepad had a 50:50 chance to coincide with the location of the target letter on the left or right sides of the screen. Additionally, there were two bilateral letter sets instead of a single lateralized target. By doing so, we ensured that participants gazed towards both sides to actively look for the target, instead of systematically looking at one side and inferring the target from there.

Apparatus and Display

The task was programmed using Experiment Builder (Version 1.10.1630, SR Research, 2016). An Eyelink 1000 plus eye tracker (SR Research) with a sampling rate of 1000 Hz was calibrated to the participants' dominant eye, but viewing was binocular. Calibration was performed with a five-point procedure, with 0.5° average accuracy for all points. Calibration was repeated whenever the error at any point was higher than 1°. A chinrest was set at 60 cm to the screen to prevent participants from moving their head. Responses were collected through a standard gamepad with keys marked *p* and *q*. The auditory stimulus was administered via headphones.

Procedure

A modification of a spatial cueing paradigm was used to combine both visual (words) and auditory (single tone) modalities. The task was run individually at the university's laboratory. Instructions informed participants that both the auditory tone and the visual words were uninformative of target location. They were told their task was simply to discriminate the target letters embedded in the letter strings, which would appear after the prime presentation. The general instruction was a speed-accuracy one (See Fig. 1 for a trial example). Participants placed their index fingers on the response keys and were asked to press them as soon as they detected a *q* or a *p*. Each trial began with a gaze-contingent fixation cross (0.3° × 0.3°), therefore the next screen would only be triggered after a minimum fixation of 1000 ms. This procedure ensured that the starting point of eye movements was at the center of the display, thus preventing attention to be oriented elsewhere prior to trial onset. A blank screen was presented for 500 ms which was followed by the presentation of one of the four possible prime combinations for 700 ms: no prime, auditory tone, visual word, or visual word + auditory tone (the auditory tone lasted 60 ms and tone and word had the same onset). Both in the prime-absent and the auditory tone conditions, a fixation cross was presented at the center of the display to anchor the initial gaze movement of participants for visual search. An interstimulus interval screen of 150 ms followed. The next screen was the target display where the two letter strings (one of which containing the *p* or *q*) appeared on the left and right visual fields for 1000 ms. A screen with a feedback message followed for 800 ms. After the feedback message disappeared, a blank screen was presented for 500 ms and a new trial began.

Each block consisted of 96 trials randomly presented, divided by 4 groups of 24 trials composed of the 4 prime categories. Within each 24 trial subset, the target was counterbalanced for letter (*p* or *q*) and location (left or right) making the target location non-predictive. Thus, the ratio of valid and invalid trials was 50:50. This means that participants could not infer the target location by attending to the meaning of the words. All factors were equiprobable and presented in a counterbalanced order. The experiment comprised 2 blocks resulting in a total of 192 trials. Participants took a self-paced break between blocks followed by a recalibration. Twelve practice trials following the same counterbalancing schema preceded the main experiment. On average, the experiment was completed in 45 minutes.

(Fig. 1 HERE)

Results

Preliminary data treatment

We flagged outliers based on values surpassing 2.5 median absolute deviations (MAD) (Leys et al., 2013). No participant was excluded as mean responses did not exceed 25% of outlier values (maximum outlier values per participant ≤ 17%). The outlier values (< 1% of data across all the analyzed variables) were then replaced by one unit above the next extreme score on that variable. Two rectangular areas of interest (AOI) were designed around the left and right target letter strings but were not visible to the participants.

Statistical analysis

We compared both manual performance (i.e., response times), and gaze movement (i.e., direction of the first saccade, and time to first fixation) in within-subjects repeated-measures analyses of variance (ANOVAs). Post-hoc comparisons were performed using the Bonferroni correction procedure. The four types of primes are fundamentally distinct and therefore it would not be suitable to compare them directly. We split them into two clear clusters and analyzed them accordingly: primes without semantic content (i.e. no prime; auditory tone) and primes with semantic content (i.e. visual word primes; visual word prime + auditory tone).

Reaction time

To directly tackle our main claim that orthographic directionality, temporal language and multimodal presentation interact to shape motor performance, we start by reporting the analysis for primes with semantic content. These factors were compared within-subjects in a 2x3x2 design: prime type (visual word primes vs. visual word primes + auditory tone) x word category (past vs. neutral vs. future) x target letter location (left vs. right).

First, we report main effects of prime type, F (1, 48) = 32.563, p < .001, η_p^2 = .40, and word category, F (2, 96) = 5.587, p = .005, η_p^2 = .10. Contrary to what we predicted (H4), visual words presented alone gave rise to faster response times (M = 696.86, SE = 5.60) than when concurrently presented with the auditory prime (M = 732.90, SE = 5.61). As for the temporal category of the word primes, future-related words generated significantly faster target identification (M = 699.03, SE = 6.65) than past-related (M = 723.24, SE = 6.23; p = .004) and temporally neutral words (M = 722.37, SE = 6.98; p = .007). Past-related and neutral words did not generate different response times (p = 1.000). This latter main effect confirms our predictions regarding the rightward pressures exerted by the orthographic system (H2).

The significant interaction of word category x target location confirms the time-space congruency hypothesis (H1), F (2, 96) = 10.937, p < .001, η_p^2 = .19. An inspection of the means reveals the typical congruency effect, in that identification of targets was systematically faster when the target location was consistent with temporal connotation of the word (Fig. 2). However, this tendency was only significant following future-related words; targets on the right (M = 677.79, SE = 9.82), that is on the side consistent with the word indication, were identified faster than targets on the left (M = 720.26, SE = 10.66; p = .009). The remaining target identification latencies on the left and right sides were not significantly different following past-related (left: M = 711.14, SE = 9.62; right: M = 735.35, SE = 12.26; p = .189) and neutral words (left: M = 718.63, SE = 8.80; right: M = 726.11, SE = 12.91; p = .664). This means that congruency effects were only significant following rightward words, although a similar but statistically nonsignificant pattern can also be observed for past-related words.

Following from the previous second-order interaction, and to explore whether there was an asymmetric performance in responding favoring words with rightward connotation (H2), we compared congruent conditions of past word – left target and future word – right target. A paired samples t-test revealed that future-related words produced marginally faster detections of the right target than past-related words of the left target, t (48) = -1.964, p = .055).

The third-order interaction between prime type x word category x target location, that brings together the research models known to drive attention (H1 + H2 + H4), was significant, F (2, 96) = 3.682, p = .029, η_p^2 = .07. We start by reporting results for the unimodal prime condition, i.e., time words alone. Identification latencies were significantly shorter for the side of space that was congruent with the implied directionality of the time-relevant word. Targets on the left (M = 675.10, SE = 12.04) relative to the right (M = 723.81, SE = 16.93; p = .049) were detected faster after a past-relevant word was presented. Similarly, but with an amplified mean difference, response times were facilitated following a future-related word for targets on the right (M = 644.56, SE = 11.68) than on the left (M = 701.22, SE = 12.46; p = .002). Neutral words presented alone did not produce different response times across the left and right targets (p = .715). Finally, we report the cueing effects following bimodal primes, i.e., time-relevant words presented jointly with the auditory tone. Against our predictions (H4), identification latencies in audiovisual prime conditions did not vary for left and right targets as a function of the category of the words (all p's > .126).

Finally, we report the analysis of the primes without semantic content. To address the hypothesis that the auditory tone serves an alertness purpose (H3), correct response times were compared in a 2 (prime type: no prime vs. auditory tone) x 2 (target letter location: left vs. right) repeated measures ANOVA. A significant main effect of prime type was observed, F (1, 48) = 70.043, p < .001, η_p^2 = .59. As predicted, exposure to the auditory tone led to significantly shorter response latencies in detecting the target letter (M = 735.02, SE = 6.43) compared to the silent condition, that is in the absence of a prime (M = 786.84, SE = 6.83) (Fig. 3). The target location did not generate significantly different identification latencies, F (1, 48) = .114, p = .735. Similarly, response times were not significantly different across left and right hemifields as a function of the type of prime, F (1, 48) = 3.333, p = .074.

(Fig. 2 HERE)

(Fig. 3 HERE)

To understand how the direction of first saccades (i.e., correct: congruent with the target location; incorrect: incongruent with the target location) may explain response times in target discrimination, we conducted a 2 (word category: past; future) x 2 (target location: left target; right target) x 2 (first saccade: correct; incorrect) repeated-measures ANOVA. The analysis was performed on the averaged data of the primes with semantic content only given that we aimed to show that, overall, shorter response latencies are associated with an initial gaze movement that is consistent with the target location (and vice-versa). The main effect of first saccade was significant, F (1, 48) = 55.970, $p < .001$, $\eta_p^2 = .54$, confirming that initially correct saccades (i.e., the direction of which coincides with the target location) result in shorter response latencies (M = 706.75, SE = 5.18) than initially incorrect saccades (i.e., the direction of which is opposite to the target location) (M = 747.97, SE = 4.91). More relevant, the significant third-order interaction, F (1, 48) = 5.997, $p = .018$, $\eta_p^2 = .11$, explains the pattern of response time results reported above. Specifically, in trials with a past word and the target located on the left, initially correct saccades (i.e., to the left hemispace) gave rise to shorter response times (M = 715, SE = 11.19) than initially incorrect saccades (i.e., to the right hemispace) (M = 749.95, SE = 9.11; $p = .006$). Likewise, in trials with a future word and the target located on the right, initially correct saccades (i.e., to the right hemispace) produced shorter response times (M = 654.66, SE = 12.64) than initially incorrect saccades (i.e., to the left hemispace) (M = 714.87, SE = 11.87; $p < .001$).

Finally, we provide a supplementary analysis to investigate whether response times with the *q* key (responded with the left index finger) and response times with the *p* key (responded with the right index finger) were significantly different. This analysis overrules the possibility that the above-mentioned cueing effects are driven by the handedness of our participants. A paired samples t-test revealed that reaction time was not different when responded with the left and right hands, t (48) = 1.279, $p = .207$.

Eye-tracking data

We excluded trials in which the observer’s gaze at fixation could not be verified to be within 1° of visual angle or for a minimum of 1000 ms (1.1% across all trial fixations). Fixations under 80 ms were excluded as were trials in which the tracker lost eye position (0.8%). Eye movement data concerning one participant was not included in the analyses due to software issues in the recording session. Eye movement was recorded from target onset.

Direction of the first saccade

We analyzed the percentage of first saccades made in each trial to the left and right sides of space as a function of the prime type and category of the time-related words. For obvious reasons, the latter factor was included only for the analysis of primes with semantic meaning. By analyzing the direction to which the first saccade was launched we were able to test whether the word primes triggered attention towards their implied direction (H1), validating the results obtained for the previous reported measures. We only report the percentage of saccades made to the right side of space (right saccade = 1), since reporting both percentages would be redundant.

The interesting comparison lies in the primes carrying semantic content (time-related words; time-related words + auditory tone). We obtained a clear main effect of word category, F (2, 94) = 22.121, p < .001, η_p^2 = .32 (Fig. 4). Irrespective of the prime type, future-related words triggered significantly more initial saccades to the right side of space (M = 57.98%, SE = 1.47%) compared to past-related (M = 46.63%, SE = 1.35*; p* < .001) or neutral words (M = 49.06%, SE = 1.59%; p < .001). Thus, we can conclude that the initial gaze movement of the participants seems to have responded to the informational content of the word prime as we predicted (H1), therefore attesting that time-related words can serve as attention-orienting primes. The interaction of word category and saccade direction did not yield significant differences in the percentage of right saccades generated, F (2, 94) = .034, p = 967.

Finally, the percentage of saccades towards the right side of space, and evidently to the left as well, was virtually the same following primes without semantic content (no prime; auditory tone), F (1, 47) = .165, p = .687, η_p^2 = .003. Roughly half of the first saccades were made to the right following the trials without prime (M = 50.08%, SE = 4.9%), and the same applies for the trials with the auditory tone as a prime (M = 50.37%, SE = 4.7%). These primes do not carry any informational content thus they are not able to induce expectations on target location.

(Fig. 4 HERE)

Time to first fixation

After having established that the word primes induce the expected orientation of attention, we further tested whether there was a right-sided asymmetry in oculomotor movement derived from scanning habits (H2) and further facilitated by bimodal stimuli (H4). To this end, we analyzed the average time elapsed until the first fixation in each trial landed on the left and right AOI's as a function of the prime type and word category. For primes without semantic content (no prime; auditory tone), the time to first fixation was analyzed across left and right AOI's.

We obtained a main effect of both prime type, F (1, 47) = 24.457, $p < .001$, $\eta_p^2 = .34$, and word category, F (2, 94) = 8.644, $p < .001$, $\eta_p^2 = .16$. Although the third-order interaction did not reveal the effect of prime type on fixation latencies (see below), first fixations following the visual words alone did occur earlier (M = 211.16, SE = 4.76), that is reached the AOI's earlier in the trial, than those following bimodal primes of word primes + auditory tone (M = 227.75, SE = 5.23). This is in line with the previous results but contrary to our prediction (H4). Rightward oriented words, those future-related, generated earlier first fixations (M = 209.28, SE = 4.39) than past-related (M = 222.24, SE = 5.48; $p = .013$), and neutral words (M = 226.95, SE = 6.08; $p < .001$). Past and neutral terms did not produce different times to first fixation ($p = .922$). This evidence reinforces that words with underlying rightward connotation do facilitate gaze movement, more so than words sharing the same category (i.e., time) but implying the opposite directionality (past-related) (H2).

A second-order interaction effect of word category x AOI emerged, F (2, 94) = 11.925, $p < .001$, $\eta_p^2 = .20$ (Fig. 5). As hypothesized (H1) and replicating the previous analyses, first fixations were facilitated when landing in the AOI congruent with the direction implied by the word prime. Thus, first fixations following past related words generated earlier fixations on the left AOI (M = 212.21, SE = 7.10) compared to the right AOI (M = 232.26, SE = 7.36; $p = .039$). In contrast, words pertaining to the future launched shorter first fixations to the right AOI (M = 192.36, SE = 4.56) than to the left AOI (M = 226, SE = 6.63; $p < .001$). Time taken to first fixation following neutral words did not vary across left (M = 225.46, SE = 7.27) and right AOI's (M = 228.45, SE = 7.91; $p = .744$). To check for our claim of asymmetrical attentional performance favoring right space (H2), we ran a direct comparison between both congruency conditions. That is, we directly compared the time to first fixation between past-related words – left AOI and future-related words – right AOI, for both prime types. We confirmed that first fixations were significantly shorter following rightward than leftward prime-target pairs. This is true for when words were presented alone, t (47) = 2.273, $p = .028$, and concurrently with the auditory event, t (47) = 2.356, $p = .023$.

The third-order interaction was not statistically significant, F (2, 94) = .145, $p = .865$, hence the uni- or bimodality of the primes did not interact with the remaining factors to affect the time to first fixation.

(Fig. 5 HERE)

The comparison between the primes without semantic content (no prime, auditory tone) revealed a main effect of prime type, F (1, 47) = 13.257, p = .001, η_p^2 = .22. Supporting what was observed in the reaction time and gaze measures, the auditory tone gave rise to faster first fixations landing on the AOI's (M = 199.63, SE = 5.91) than the absence of prime (M = 218.17, SE = 5.29) (Fig. 6). The remaining effects were not statistically significant (all p's > .279).

(Fig. 6 HERE)

To understand how the direction of first saccades (i.e., correct: congruent with the word prime; incorrect: incongruent with the word prime) may explain the time taken to first fixation in a trial (the fixation that resulted from the initial gaze movement of participants), we conducted a 2 (word category: past; future) x 2 (first saccade: correct; incorrect) repeated-measures ANOVA. This analysis did not include primes without semantic content nor neutral words because neither carry directional content. Therefore, first saccades resulting from these primes cannot be computed as correct or incorrect. Additionally, the analysis was performed on the averaged data for the primes with semantic content because the previous analysis on time to first fixation revealed that prime type did not interact with the remaining factors.

The main effect of initial saccade, F (1, 47) = 57.075, p < .001, η_p^2 = .55, confirms that first fixations were significantly shorter after saccades that were correct, or congruent with the word prime (M = 183.38, SE = 3.15), than incorrect, or incongruent (M = 218.67, SE = 5.05). As expected, the direction of the first saccade explains the pattern of time to first fixation, F (1, 47) = 8.189, p = .006, η_p^2 = .15. The significant interaction reveals that past words followed by correct saccades (i.e., left) resulted in shorter first fixations (M = 199.73, SE = 4.33) than when followed by incorrect saccades (i.e., right; M = 223.98, SE = 6.49, p < .001). Likewise, future words followed by correct saccades (i.e., right) produced shorter first fixations (M = 167.03, SE = 3.11) than by incorrect saccades (i.e., left; M = 213.26, SE = 5.38). This demonstrates that initial gaze movement consistent with the prime predicts shorter time to first fixation which then affects response latencies, as we have reported above. Further, and confirming that words aligned with script-direction (i.e., future-related) bias visuomotor attention accordingly, first fixations after correct saccades were significantly shorter after future than past word primes (p < .001).

Finally, a correlation was performed to investigate the relationship between response time and time to first fixation. We observed a significant positive correlation between response time and time to first fixation (r (48) = .304, p = .036), such that when the time participants take to land the first fixation on either interest area increases so does the time taken to identify the target.

Discussion

The aim of the present study was to compare the attentional and motor effects elicited by unimodal and bimodal primes. Specifically, we hypothesized that a noninformative auditory event co-occurring with a temporal visual word suffices to increase the salience of the visual prime and consequently to enhance visual search relative to the visual word presented alone. In unimodal visual trials, future-related words preferentially biased attention and facilitated performance on the right hemifield due to asymmetrical routines of reading and writing. In unimodal auditory trials, the tone boosted visual search and identification of the lateralized targets. However, our results indicate that bimodal primes were not more advantageous in capturing participants' attention than the unimodal primes.

Retrospective and prospective word primes have successfully directed attention toward their implied location, be it left or right. This was revealed by shorter response latencies, higher percentage of first saccades, and shorter time to first fixation on the congruent locations. The proposition that people make use of horizontal properties of space when processing temporal language is not new (Boroditsky et al., 2011; Fuhrman and Boroditsky, 2010; Núñez and Cooperrider, 2013; Weger and Pratt, 2008). An entirely different question is whether the time-space convergence translates into benefits for visuospatial orienting. Previous research showing that activation of past- and future-referent cues prime behavioral responses to left and right space seems to support this hypothesis (Lakens et al., 2011; Ouellet et al., 2010a; Santiago et al., 2007; Torralbo et al., 2006). Although motor outputs resulting from cueing tasks are often taken as a signal for attention allocation, they may not necessarily substantiate that attention has been deployed. Our results go one step further than prior reports by replicating the time-space overlap in motor performance while tracking participants' eye movements. The three collected measures of gaze movement systematically showed that the meaning of the temporal words was truly embodied by participants because it affected their visual attention. The fact that word category (past, future) has generated earlier first fixations on congruent areas of interest (left, right) irrespective of whether they contained the target is an indicator that participants gazed spontaneously as a response to the word prime. This suggests that, together with the first saccade direction which was congruent with the words' implied spatial information, participants gaze movement responded to the content of the primes. This spatial activation occurred despite participants being informed that the words were unrelated to target location, and thus attending to their meaning would not necessarily ensure correct target identification.

In addition, the proposed motor rightward asymmetry was also observed. The habitualized left-to-right eye trajectory appears to have promoted (marginally) faster right-sided target identification following prospective words in the condition of unimodal visual words. The significant imbalance in responding between congruent prime-target pairs on time taken to land the first fixation (a behavior that precedes motor responses) further attests that attention is not homogenously distributed but follows a preferential pattern - otherwise no differences between congruencies would have emerged. These results suggest that the precedence of the rightward scanning direction must be due, at least in part, to habitualized regularities brought upon by the convention for text direction (Bergen and Lau, 2012; Bettinsoli et al., 2019; Bulf et al., 2017; Flath et al., 2019). In fact, by acknowledging that universal genetic proclivities for a left anchoring of attention exist (Brooks et al., 2014), the right-sided advantage observed with time words in unimodal trials (which counters leftward biological predispositions) shows that reading and writing regularities must exert some form of attention control. This asymmetric spatial performance between prospective and retrospective terms is a novel contribution. It is important to note that these words are a) not overlearned (relative to terms such as ‘left’ and ‘right’, Hommel et al., 2001) and b) derived from an abstract concept (i.e. time) lacking concrete sensory-motor basis. Distinct tasks have shown that the reading and writing direction of one’s native language preferentially maps human motion (Maass et al., 2009; Maass and Russo, 2003; Suitner and Maass, 2016) and by extension many other abstract concepts (e.g., politics, Farias et al., 2016; time, Lakens et al., 2011; number line, Zebian, 2005). This study is the first to show an imbalance in true attentional performance which favors prime-target combinations aligned with script direction.

Our main prediction was that the concurrent presentation of visual and auditory primes would produce an additive effect on the unidirectional spatial asymmetry described above. We expected bimodal primes, relative to their unimodal counterpart, to speed up visual search (Ngo & Spence, 2010; Van der Burg et al., 2008) and produce benefits for rightward target detection. Our results revealed the opposite. Bimodal audiovisual primes (versus unimodal visual primes) seemed to have hampered performance in response times as well as time taken to first fixation. The unimodal visual condition was substantially more efficient in grabbing attention compared to the audiovisual condition. A very different result can be seen in unimodal conditions of what we labelled as ‘primes without semantic content’.

When presented in isolation, the auditory tone gave rise to faster responses than trials absent of prime stimuli. It is important to note that the auditory event contained no information as to the targets' impending location. While participants could not extract any information nor rely on top-down strategies on both unimodal conditions lacking semantic content, they have performed far better in trials presenting a single tone, than in trials without a prime. Likely, exposure to the auditory event created a sense of general alertness (Coull and Nobre, 1998). The alerting signal led participants to mobilize attentional resources to the forthcoming target screen. The recruitment of these resources might have benefited the encoding of target information (Matthias et al., 2010), although no relation between tone and target timing or location could be established. Therefore, the auditory tone triggered two simultaneous processes: an alertness one that boosted motor performance, and an attentional preparedness one, that enhanced perceptual processing of targets (Correa et al., 2004; Kusnir et al., 2011).

Although we expected the same phenomena to manifest in bimodal cueing conditions, namely that the auditory tone paired with the temporal word would foster performance, we observed otherwise. Several explanations may be put forward. First, one should always consider that, in general, the visual modality takes precedence over the auditory modality (see Bertelson and De Gelder, 2004 for a review). However, if the preferred modality in our task were vision, then it would have had dominated bimodal primes and attenuated processing in the non-dominant modality, audition. If this were the case, no costs should have been observed in bimodal primes and we would have obtained the same effects across bimodal (auditory + visual) and unimodal primes (visual). Second, the bimodal stimuli in this study were employed as primes, rather than targets as in other studies demonstrating enhanced visual perception with audiovisual stimuli (Dalton and Spence, 2007; Frassinetti et al., 2002; Van der Burg et al., 2008). This means that in our study, auditory and visual events preceded target onset. Therefore, the pop-out effect that the co-occurring single tone could have induced on target discrimination was potentially invalidated.

The prime candidate explanation for the impairment in cueing effects with bimodal primes is that the co-occurring auditory tone acted as a distractor, preventing the semantic processing of the words to take place. An alternative explanation may be that the auditory tone did increase the salience of the word primes (or the fixation cross in trials without words) and the participants' focus on them. Consequently, this could delay attention to be released to the target screen, which would result in slower responses (Robinson and Sloutsky, 2010). In our view, this hypothesis is less likely because if time words were made salient, then we should have observed identification of targets on the left and right space congruent with the words' spatial information in bimodal trials.

Projective word cues such as ours impose the activation of a complex frame of reference (the retrieval of the spatial properties associated with time) (Gibson and Kingstone, 2006). The visual and semantic processing was likely interrupted by the introduction of another, albeit semantically non-competing, conjoint stimulus. This would explain why effects on the two conditions containing the auditory tone rendered opposite results: while performance was enhanced for the auditory tone alone, it was impaired when presented synchronously with the word prime, failing to recruit the necessary attentional resources. Arguably, it is also possible that audiovisual primes might benefit from slight desynchronization of its auditory and visual components in order to capture attention more efficiently (Spence and Driver, 1999). By delaying the auditory event by about 55 ms relative to the onset of the visual event, both signals would arrive synchronously at the superior colliculus. This would increase the likelihood for multisensory integration (see Spence and Santangelo, 2009 for a review). Several studies have reported enhancement in multisensory over unisensory settings (Talsma et al., 2010) particularly when modalities provide complementary information (e.g., a voice and a moving mouth Mcgurk and Macdonald, 1976; visual and auditory apparent motion streams, Soto-Faraco et al., 2002). In contrast, the introduced auditory tone was inherently nonspatial and did not provide any supplementary information to decode the target location. Thus, what might have impeded the attentional cueing effects was precisely the nonspatial feature of the auditory tone we deemed advantageous because it represented no conflicting content. It is likely that multisensory integration was not achieved because the visual and auditory properties of the stimulus were not bound into a single coherent percept (Oruc et al., 2008; Talsma and Woldorff, 2005). Hence, multiple senses did not support each other to produce an enhanced, concerted response above that evoked by the single unimodal elements (Stein and Stanford, 2008). The goal of multisensory integration is to gather information across the senses that is relevant for the self to effectively deal with the surroundings. If integration of information belonging to the same object fails, then one modality can compromise processing in the second modality (Robinson et al., 2018; Shams et al., 2002; Sloutsky and Napolitano, 2003).

It is important to note that the asymmetric performance that favored target detection on the right following unimodal future primes was observed in a sample of participants who read from left-to-right. Although there is previous research confirming that the mapping of time onto space is reversed in leftward speaking communities (Fuhrman and Boroditsky, 2010; Ouellet et al., 2010b), future research would benefit from testing the proposed asymmetry in the representation of time concepts in participants that read from right-to-left.

These findings reveal a broader picture in terms of what contributes to the triggering of visuospatial attention. They confirm that temporal language orients attention, but not equally: prospective word primes have benefits because they activate motoric and attentional processes similar to those instilled by scanning habits. The co-occurrence of an auditory, inherently nonspatial, tone (but not its single presentation) undermines motor and visual output, impairing cueing effects. The results find support in other studies that failed to report attention capturing advantages for bimodal over unimodal cueing (Santangelo et al., 2006; Ward et al., 1998) and therefore constitute an important advance to the multisensory cueing literature.

References

Afsari, Z., Keshava, A., Ossandón, J.P., König, P., 2018. Interindividual differences among native right-to-left readers and native left-to-right readers during free viewing task. Vis. cogn. 26, 430–441.

Alvarado, J.C., Vaughan, J.W., Stanford, T.R., Stein, B.E., 2007. Multisensory versus unisensory integration: Contrasting modes in the superior colliculus. J. Neurophysiol. 97, 3193–3205.

Anderson, S.F., Kelley, K., Maxwell, S.E., 2017. Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. Psychol. Sci. 28, 1547–1562.

Barnhart, W.R., Rivera, S., Robinson, C.W., 2018. Different patterns of modality dominance across development. Acta Psychol. (Amst). 182, 154–165.

Bergen, B.K., Lau, T.T.C., 2012. Writing direction affects how people map space onto time. Front. Psychol. 3, 109.

Bertelson, P., De Gelder, B., 2004. The psychology of multimodal perception. In: Charles, S., Driver, J. (Eds.), Crossmodal Space and Crossmodal Attention. Oxford University Press, Oxford, England, pp. 141–179.

Bettinsoli, M.L., Maass, A., Suitner, C., 2019. The first, the least and the last: Spatial asymmetries in memory and their relation to script trajectory. Mem. Cogn. 47, 229–239.

Boroditsky, L., Fuhrman, O., McCormick, K., 2011. Do English and Mandarin speakers think about time differently? Cognition 118, 123–129.

Boroditsky, L., Gaby, A., 2010. Remembrances of times east: Absolute spatial representations of time in an australian aboriginal community. Psychol. Sci. 21, 1635–1639.

Brooks, J.L., Sala, S. Della, Darling, S., 2014. Representational pseudoneglect: A review. Neuropsychol. Rev. 24, 148–165.

Bulf, H., De Hevia, M.D., Gariboldi, V., Cassia, V.M., 2017. Infants learn better from left to right: A directional bias in infants' sequence learning. Sci. Rep. 7.

Butler, S.H., Harvey, M., 2006. Perceptual biases in chimeric face processing: Eye-movement patterns cannot explain it all. Brain Res. 1124, 96–99.

Chahboun, S., Flumini, A., Pérez González, C., McManus, I.C., Santiago, J., 2017. Reading and writing direction effects on the aesthetic appreciation of photographs. Laterality Asymmetries Body, Brain Cogn. 22, 313–339.

Chokron, S., De Agostini, M., 2000. Reading habits influence aesthetic preference. Cogn. Brain Res. 10, 45–49.

Correa, Á., Lupiáñez, J., Milliken, B., Tudela, P., 2004. Endogenous temporal orienting of attention in detection and discrimination tasks. Percept. Psychophys. 66, 264–278.

Coull, J.T., Nobre, A.C., 1998. Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. J. Neurosci. 18, 7426–7435.

Dalton, P., Spence, C., 2007. Attentional capture in serial audiovisual search tasks. Percept. Psychophys. 69, 422–438.

Driver, J., Spence, C., 1998. Cross–modal links in spatial attention. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 353, 1319–1331.

Farias, A.R., Garrido, M. V., Semin, G.R., 2016. Embodiment of abstract categories in space. . . grounding or mere compatibility effects? The case of politics. Acta Psychol. (Amst). 166, 49–53.

Fernandez-Duque, D., Posner, M.I., 1997. Relating the mechanisms of orienting and alerting. Neuropsychologia 35, 477–486.

Flath, M.E., Smith, A.K., Elias, L.J., 2019. Cultural differences in lateral biases on aesthetic judgments: The effect of native reading direction. Cult. Brain 7, 57–66.

Frassinetti, F., Bolognini, N., Làdavas, E., 2002. Enhancement of visual perception by crossmodal visuo-auditory interaction. Exp. Brain Res. 147, 332–343.

Fuhrman, O., Boroditsky, L., 2010. Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. Cogn. Sci. 34, 1430–51.

Gibson, B.S., Kingstone, A., 2006. Visual attention and the semantics of space: beyond central and peripheral cues. Psychol. Sci. 17, 622–7.

Hernandez, M.D., Wang, Y., Sheng, H., Kalliny, M., Minor, M., 2017. Escaping the corner of death? An eye-tracking study of reading direction influence on attention and memory. J. Consum. Mark. 34, 1–10.

Ho, C., Santangelo, V., Spence, C., 2009. Multisensory warning signals: When spatial correspondence matters. Exp. Brain Res. 195, 261–272.

Ho, C., Spence, C., 2005. Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. J. Exp. Psychol. Appl. 11, 157–174.

Hommel, B., Pratt, J., Colzato, L., Godijn, R., 2001. Symbolic control of visual attention. Psychol. Sci. 12, 360–5.

Jewell, G., McCourt, M.E., 2000. Pseudoneglect: A review and meta-analysis of performance factors in line bisection tasks. Neuropsychologia 38, 93–110.

Kermani, M., Verghese, A., Vidyasagar, T.R., 2018. Attentional asymmetry between visual hemifields is related to habitual direction of reading and its implications for debate on cause and effects of dyslexia. Dyslexia 24, 33–43.

Kusnir, F., Chica, A.B., Mitsumasu, M.A., Bartolomeo, P., 2011. Phasic auditory alerting improves visual conscious perception. Conscious. Cogn. 20, 1201–1210.

Lai, V., Boroditsky, L., 2013. The immediate and chronic influence of spatio-temporal metaphors on the mental representations of time in English, Mandarin, and Mandarin-English speakers. Front. Psychol.

Lakens, D., Semin, G.R., Garrido, M. V., 2011. The sound of time: Cross-modal convergence in the spatial structuring of time. Conscious. Cogn. 20, 437–43.

Laurienti, P.J., Perrault, T.J., Stanford, T.R., Wallace, M.T., Stein, B.E., 2005. On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. In: Experimental Brain Research. Exp Brain Res, pp. 289–297.

Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. 49, 764–766.

Maass, A., Pagani, D., Berta, E., 2007. How beautiful is the goal and how violent is the fistfight? Spatial bias in the interpretation of human behavior. Soc. Cogn. 25, 833–852.

Maass, A., Russo, A., 2003. Directional bias in the mental representation of spatial events: Nature or culture? Psychol. Sci. 14, 296–301.

Maass, A., Suitner, C., Favaretto, X., Cignacchi, M., 2009. Groups in space: Stereotypes and the spatial agency bias. J. Exp. Soc. Psychol. 45, 496–504.

Majid, A., Gaby, A., Boroditsky, L., 2013. Time in terms of space. Front. Psychol. 4, 554.

Matthias, E., Bublak, P., Müller, H.J., Schneider, W.X., Krummenacher, J., Finke, K., 2010. The influence of alertness on spatial and nonspatial components of visual attention. J. Exp. Psychol. Hum. Percept. Perform. 36, 38–56.

Mcgurk, H., Macdonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746–748.

Mills, M., Smith, K.B., Hibbing, J.R., Dodd, M.D., 2015. Obama cares about visuo-spatial attention: Perception of political figures moves attention and determines gaze direction. Behav. Brain Res. 278, 221–225.

Ngo, M.K., Pierce, R.S., Spence, C., 2012. Using multisensory cues to facilitate air traffic management. Hum. Factors 54, 1093–103.

Ngo, M.K., Spence, C., 2010. Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. Attention, Perception, Psychophys. 72, 1654–1665.

Núñez, R., Cooperrider, K., 2013. The tangle of space and time in human cognition. Trends Cogn. Sci. 17, 220–229.

Núñez, R., Cooperrider, K., Doan, D., Wassmann, J., 2012. Contours of time: Topographic construals of past, present, and future in the Yupno valley of Papua New Guinea. Cognition 124, 25–35.

Oruc, I., Sinnett, S., Bischof, W.F., Soto-Faraco, S., Lock, K., Kingstone, A., 2008. The effect of attention on the illusory capture of motion in bimodal stimuli. Brain Res. 1242, 200–208.

Ossandón, J.P., Onat, S., König, P., 2014. Spatial biases in viewing behavior. J. Vis. 14, 1–26.

Ouellet, M., Santiago, J., Funes, M.J., Lupiáñez, J., 2010a. Thinking about the future moves attention to the right. J. Exp. Psychol. Hum. Percept. Perform. 36, 17–24.

Ouellet, M., Santiago, J., Israeli, Z., Gabay, S., 2010b. Is the future the right time? Exp. Psychol. 57, 308–314.

Rashidi-Ranjbar, N., Goudarzvand, M., Jahangiri, S., Brugger, P., Loetscher, T., 2014. No horizontal numerical mapping in a culture with mixed-reading habits. Front. Hum. Neurosci. 8, 1–5.

Robinson, C.W., Moore, R.L., Crook, T.A., 2018. Bimodal presentation speeds up auditory processing and slows down visual processing. Front. Psychol. 9, 1–10.

Robinson, C.W., Sloutsky, V.M., 2010. Development of cross-modal processing. Wiley Interdiscip. Rev. Cogn. Sci. 1, 135–141.

Santangelo, V., Spence, C., 2008. Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies. Conscious. Cogn. 17, 989–1015.

Santangelo, V., Van Der Lubbe, R.H.J., Belardinelli, M.O., Postma, A., 2006. Spatial attention triggered by unimodal, crossmodal, and bimodal exogenous cues: A comparison of reflexive orienting mechanisms. Exp. Brain Res. 173, 40–48.

Santiago, J., Lupiáñez, J., Pérez, E., Funes, M.J., 2007. Time (also) flies from left to right. Psychon. Bull. Rev. 14, 512–516.

Santiago, J., Román, A., Ouellet, M., Rodríguez, N., Pérez-Azor, P., 2010. In hindsight, life flows from left to right. Psychol. Res. 74, 59–70.

Selcon, S.J., Taylor, R.M., McKenna, F.P., 1995. Integrating multiple information sources: Using redundancy in the design of warnings. Ergonomics 38, 2362–2370.

Shaki, S., Fischer, M.H., Petrusic, W.M., 2009. Reading habits for both words and numbers contribute to the SNARC effect. Psychon. Bull. Rev. 16, 328–331.

Shams, L., Kamitani, Y., Shimojo, S., 2002. Visual illusion induced by sound. In: Cognitive Brain Research. Elsevier, pp. 147–152.

Sloutsky, V.M., Napolitano, A.C., 2003. Is a picture worth a thousand words? Preference for auditory modality in young children. Child Dev. 74, 822–833.

Smith, A.K., Duerksen, K.N., Gutwin, C., Elias, L.J., 2020. Lateral biases in aesthetic and spatial location judgments: differences between tasks and native reading directions. Laterality Asymmetries Brain, Behav. Cogn. 25, 5–21.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., Kingstone, A., 2002. The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. In: Cognitive Brain Research. Elsevier, pp. 139–146.

Spence, C., Driver, J., 1997. Audiovisual links in exogenous covert spatial orienting. Percept. Psychophys. 59, 1–22.

Spence, C., Driver, J., 1999. A new approach to the design of multimodal warning signals. Eng. Psychol. Cogn. Ergon. 4, 455–461.

Spence, C., Parise, C., Chen, Y.C., 2011. The Colavita visual dominance effect. In: McDonald, J.J., Green, J.J., Störmer, V., Hillyard, S.A. (Eds.), The Neural Bases of Multisensory Processes. CRC Press/Taylor & Francis, pp. 529–556.

Spence, C., Santangelo, V., 2009. Capturing spatial attention with multisensory cues: A review. Hear. Res. 258, 134–142.

Stein, B.E., Stanford, T.R., 2008. Multisensory integration: Current issues from the perspective of the single neuron. Nat. Rev. Neurosci. 9, 255–266.

Suitner, C., Maass, A., 2016. Spatial agency bias: Representing people in space. Adv. Exp. Soc. Psychol. 53, 245–301.

Suitner, C., Maass, A., Ronconi, L., 2017. From spatial to social asymmetry: Spontaneous and conditioned associations of gender and space. Psychol. Women Q. 41, 46–64.

Talsma, D., Senkowski, D., Soto-Faraco, S., Woldorff, M.G., 2010. The multifaceted interplay between attention and multisensory integration. Trends Cogn. Sci.

Talsma, D., Woldorff, M.G., 2005. Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. J. Cogn. Neurosci. 17, 1098–1114.

Teder-Sälejärvi, W.A., McDonald, J.J., Di Russo, F., Hillyard, S.A., 2002. An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. In: Cognitive Brain Research. Elsevier, pp. 106–114.

Torralbo, A., Santiago, J., Lupiáñez, J., 2006. Flexible conceptual projection of time onto spatial frames of reference. Cogn. Sci. 30, 745–57.

Tversky, B., Kugelmass, S., Winter, A., 1991. Cross-cultural and developmental trends in graphic productions. Cogn. Psychol. 23, 515–557.

Van der Burg, E., Olivers, C.N.L., Bronkhorst, A.W., Theeuwes, J., 2008. Pip and pop: Nonspatial auditory signals improve spatial visual search. J. Exp. Psychol. Hum. Percept. Perform. 34, 1053–1065.

von Sobbe, L., Scheifele, E., Maienborn, C., Ulrich, R., 2019. The space-time congruency effect: A meta-analysis. Cogn. Sci. 43, e12709.

Vroomen, J., De Gelder, B., 2000. Sound enhances visual perception: Cross-modal effects of auditory organization on vision. J. Exp. Psychol. Hum. Percept. Perform. 26, 1583–1590.

Wallace, M.T., Meredith, M.A., Stein, B.E., 1998. Multisensory integration in the superior colliculus of the alert cat. J. Neurophysiol. 80, 1006–1010.

Ward, L.M., 1994. Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. Can. J. Exp. Psychol. 48, 242–259.

Ward, L.M., McDonald, J.J., Golestani, N., 1998. Cross-modal control of attention shifts. In: Wright, R.D. (Ed.), Vancouver Studies in Cognitive Science. Oxford University Press, pp. 232–268.

Weger, U.W., Pratt, J., 2008. Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. Psychon. Bull. Rev. 15, 426–430.

Welch, R.B., Warren, D.H., 1986. Intersensory interactions. In: Boff, K.R., Kaufman, L., Thomas, J.P. (Eds.), Handbook of Perception and Performance: Sensory Processes and Perception. Wiley, New York, NY, pp. 25–36.

Zebian, S., 2005. Linkages between number concepts, spatial thinking, and directionality of writing: The SNARC effect and the REVERSE SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. J. Cogn. Cult. 5, 165–190.

Appendix 1

Table 1. Temporal prime words and mean ratings

Past		Neutral		Future	
Word	**Mean rating**	**Word**	**Mean rating**	**Word**	**Mean rating**
Antigamente *(previously)*	6.06	Dia *(day)*	48.63	De seguida *(next)*	66.13
Passado *(past)*	9.12	Hoje *(today)*	52.84	Posteriormente *(subsequently)*	68.60
Foi *(it was)*	19.83	Mês *(month)*	56.58	Em breve *(soon)*	69.26
Antes *(before)*	23.31	Trimestre *(quarter)*	57.38	Depois *(after)*	73.28
Anteontem *(before yesterday)*	24.91	Semana *(week)*	57.63	Amanhã *(tomorrow)*	73.78
Ontem *(yesterday)*	26.92	Semestre *(semester)*	59.32	Depois de amanhã *(after tomorrow)*	78.38
Há pouco tempo *(not long ago)*	34.19	Ano *(year)*	59.35	Será *(it will)*	82.90
Recentemente *(recently)*	36.76	Temporada *(season)*	62.07	Futuro *(future)*	94.11

Acknowledgements

The authors gratefully acknowledge the financial support from the Portuguese Foundation for Science and Technology (UIDB/04810/2020) with grants awarded to the first author (SFRH/BD/118845/2016).

Figure Captions

Figure 1. Panel A shows and example of a trial with bimodal priming (time word + auditory tone) and panel B shows an example of a trial with unimodal auditory priming (auditory tone).

Figure 2. Mean response time (in milliseconds) as a function of primes with semantic content, time-related word category, and target location. Error bars represent the standard error from the mean.

Figure 3. Mean response time (in milliseconds) as a function of primes without semantic content, and target location. Error bars represent the standard error from the mean.

Figure 4. Mean percentage of first saccades in each trial to the right hemifield as a function of prime type, and time-related word prime category. Error bars represent the standard error from the mean.

Figure 5. Average time to first fixation (in milliseconds) as a function of primes with semantic content, time-related word category, and AOI location. Error bars represent the standard error from the mean.

Figure 6. Average time to first fixation (in milliseconds) as a function of primes without semantic content, and AOI location. Error bars represent the standard error from the mean.