Vision Research 127 (2016) 165–176

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Crossmodal binding rivalry: A “race” for integration between unequal sensory inputs

Maria Kostaki[a], Argiro Vatakis[a,b,*]

[a] Postgraduate Program Basic and Applied Cognitive Science, Department of Philosophy and History of Science, University of Athens, Athens, Greece
[b] Cognitive Systems Research Institute, Athens, Greece

ARTICLE INFO

Article history:
Received 5 December 2015
Received in revised form 8 August 2016
Accepted 11 August 2016
Available online 6 September 2016

Keywords:
Multisensory integration
Double flash illusion
Fusion illusion
Fission illusion
Crossmodal binding rivalry
Auditory
Visual

ABSTRACT

Exposure to multiple but unequal (in number) sensory inputs often leads to illusory percepts, which may be the product of a conflict between those inputs. To test this conflict, we utilized the classic sound induced visual fission and fusion illusions under various temporal configurations and timing presentations. This conflict between unequal numbers of sensory inputs (i.e., *crossmodal binding rivalry*) depends on the binding of the first audiovisual pair and its temporal proximity to the upcoming unisensory stimulus. We, therefore, expected that tight coupling of the first audiovisual pair would lead to higher rivalry with the upcoming unisensory stimulus and, thus, weaker illusory percepts. Loose coupling, on the other hand, would lead to lower rivalry and higher illusory percepts. Our data showed the emergence of two different participant groups, those with low discrimination performance and strong illusion reports (particularly for fusion) and those with the exact opposite pattern, thus extending previous findings on the effect of visual acuity in the strength of the illusion. Most importantly, our data revealed differential illusory strength across different temporal configurations for the fission illusion, while for the fusion illusion these effects were only noted for the largest stimulus onset asynchronies tested. These findings support that the optimal integration theory for the double flash illusion should be expanded so as to also take into account the multisensory temporal interactions of the stimuli presented (i.e., temporal sequence and configuration).

1. Introduction

Our brain has the ability to integrate information from different modalities originating close in time and space (Stein & Meredith, 1993). Integration for sensory signals that are equal in number (e.g., one visual and one auditory) is usually quite straightforward, resulting in enhanced detectability of a target and/or faster reaction times (Forster, Cavina-Pratesi, Aglioti, & Berlucchi, 2002; Stein, Lagondon, Wilkinson, & Price, 1996). Sensory inputs, however, that are equal in number but have distal origin in time or space often result in perceptual illusions with inputs from one modality distorting the percept of other sensory inputs. For instance, vision may influence the spatial processing of an auditory stimulus (i.e., ventriloquist effect; Alais & Burr, 2004), while audition may affect the temporal processing of the visual stimuli in terms of temporal position, perceived duration, or flickering rate (e.g., temporal ventriloquism effect; Burr, Banks, & Morrone, 2009; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Repp & Penel, 2002; Welch, DuttonHurt, & Warren, 1986). This differential modality dominance has been described by computational models that attempted to minimize the variance (i.e., increase reliability) in the final percept (Ernst & Bülthoff, 2004). These models aim to estimate and weight the variance of the audiovisual incoming inputs using either Bayesian or Maximum Likelihood estimations (Battaglia, Jacobs, & Aslin, 2003; Ernst & Banks, 2002; Shams, Ma, & Beierholm, 2005) and determine the degree to which one modality will dominate over the other under specific circumstances (known also as optimal integration principle).

The account of optimal integration has been proposed by Shams et al. (2005) in order to address the binding of unequal sensory inputs originating from different modalities. One such case of unequal sensory inputs is the well-known example of the sound-induced flash illusion (SIFI), where a single flash in the presence of two beeps is perceived as two distinct flashes (Shams, Kamitani, & Shimojo, 2000) and the fusion illusion, where two

---

* Corresponding author at: Makedonomachou Prantouna 7, 11525 Athens, Greece.
E-mail address: argiro.vatakis@gmail.com (A. Vatakis).

http://dx.doi.org/10.1016/j.visres.2016.08.004
0042-6989/© 2016 Elsevier Ltd. All rights reserved.

flashes presented with one beep are "fused" to a single flash (Andersen, Tiippana, & Sams, 2004). Shams et al. (2005) proposed that the human brain combines unequal in number audiovisual stimuli according to Bayesian rules relying on the most reliable (i.e., with less variance) for the task modality: in this case, audition (see also Apthorp, Alais, & Boenke, 2013; Cuppini, Magosso, Bolognini, Vallar, & Ursino, 2014; Roseboom, Kawabe, & Nishida, 2013).

The SIFI however, has not always been robust across participants between or even within studies. Research has shown that some participants tend to be highly susceptible to the classical presentation of the illusion (i.e., beep presented either in synchrony with the first flash or between the two flashes), while others are less susceptible (Kumpik, Roberts, King, & Bizley, 2014; McGovern, Roudaia, Stapleton, McGinnity, & Newell, 2014; Stevenson, Zemtsov, & Wallace, 2012). Such differential susceptibility has led researchers to: (a) preselect the participants so as to perceive the illusion (Mishra, Martinez, & Hillyard, 2013), (b) preselect individuals based on their visual acuity (Rosenthal, Shimojo, & Shams, 2009), which later revealed that reduced acuity led to higher susceptibility to the illusion (Kumpik et al., 2014), (c) exclude participants with weak illusory percepts from further analysis (Fiedler, O'Sullivan, Schroter, Miller, & Ulrich, 2011), or (d) evaluate illusory performance relative to one's visual acuity baseline (Apthorp et al., 2013). There are also studies that have not treated or considered susceptibility and/or visual acuity differences including all participants in their analysis (e.g., Andersen et al., 2004) and studies that have split their participants in those who could and those who could not perceive the illusion and analyzed the two groups separately (Mishra, Martinez, & Hillyard, 2008; Mishra, Martinez, Sejnowski, & Hillyard, 2007).

Many factors could potential promote this differential participant susceptibility to the SIFI. It could, for instance, be associated with the temporal window of integration (TWI; i.e., the interval in which no disparity in timing is detected and stimuli are integrated; Kerlin & Shapiro, 2015; Stevenson et al., 2012). For instance, Stevenson and colleagues have shown that narrower TWIs result in reduced illusory percepts due to higher discrimination ability for asynchronous inputs. Similarly, Kerlin and Shapiro (2015) have shown longer alpha rhythm wavelength in occipital activity (i.e., longer TWIs) to result in increased susceptibility to the illusion at longer stimulus onset asynchronies (SOAs). Moreover, it has been shown that degraded percepts in one modality affect the unified multisensory percept in a fashion similar to what the optimal integration mechanism would predict (Alais & Burr, 2004; Bresciani, Dammeier, & Ernst, 2006; Ernst & Banks, 2002; Rohe & Noppeney, 2015; Slutsky & Recanzone, 2001). Thus, one's discrimination ability in unisensory stimulation, such as vision, might also be a parameter that affects the degree of the susceptibility to the SIFI and the fusion illusion (something that some researchers have started to investigate more vigorously; e.g., Kumpik et al., 2014). Recently, Odegaard and Shams (2016) posed yet another view that might account for the individual differences in susceptibility to the illusion that points to the individuals' "*binding tendency*", which refers to the brain's probability to assume a common cause for the sensory inputs coming from different modalities and, thus, integrate them.

Mishra et al. (2013) recently posed yet another challenge on the theories accounting for the double flash illusion: the view that temporal positioning and proximity modulates the SIFI. Specifically, Mishra and colleagues showed that two brief sounds can affect the degree of color integration of two successive flashes. Using one red and one green flash accompanied by two brief sounds they found that participants had strong illusory percepts of orange flashes (one or two) instead of a red and a green flash. The percent of orange reports was subject to the temporal proximity of the two flashes as well as the temporal position of the second sound in relation to the flashes (i.e., when the second beep was presented between the two flashes color discrimination increased, while when the second beep followed the second flash discrimination decreased). Such results show, for the first time, that the temporal relation of audiovisual inputs within the TWI may alter the illusory visual percept in crossmodal conditions. To-date, research on the SIFI has not shown evidence of differential illusory strength as a function of the temporal sequence of the audiovisual inputs (i.e., whether the flash is presented simultaneously with the first or the second beep irrespective of the SOA between the two beeps; Apthorp et al., 2013; Shams, Kamitani, & Shimojo, 2002). A closer look at the literature, however, reveals that no research so far has ever directly compared the three possible (and widely used) temporal sequences that auditory and visual inputs can take in the SIFI and the fusion illusion. Such comparison will, however, allow one to clarify how the temporal presentation and temporal sequence of the different sensory inputs modulate the strength of the illusion.

In the present study, therefore, we aim, for the first time, to evaluate the most common temporal sequences used in the SIFI and fusion illusion across the same participants and at different temporal proximities (i.e., SOAs) using the classic experimental set-up of the SIFI (Shams et al., 2000) and the fusion illusion (Andersen et al., 2004). The experienced illusions could potentially be dominated by audition, which is indeed more reliable than vision for temporal tasks (e.g., Andersen et al., 2004; Wada, Kitagawa, & Noguchi, 2003). In such case, one would expect equal (or not significantly different) illusory strengths at all configurations and timings within – at least – the TWI. This dominance account, however, may not be sufficient (as discussed), thus, we aim to examine whether or not additional parameters could also provide a more thorough explanation of the phenomenon. The candidate parameters are adopted from the multisensory integration literature and relate to a: (a) resilient binding when visual stimulation precedes or is in synchrony with the auditory input (Keetels & Vroomen, 2012; van Wassenhove, Grant, & Poeppel, 2007; Vatakis & Spence, 2007, 2008), (b) decreased tolerance of the perceptual system to auditory precedence in an audiovisual stimulus pairing (e.g., Vatakis, 2013), and c) weakened tolerance for larger temporal distances between audiovisual inputs (i.e., as distance increases between a flash and a beep, the less likely we are to treat them as a unified audiovisual pair; e.g., Vatakis & Spence, 2010) – even within the TWI.

We, therefore, hypothesize that in the presence of unequal number of sensory inputs, a rivalry between those inputs will arise, which is dependent on the binding of the first audiovisual stimulus pair and its temporal proximity with the next unisensory stimulation (note that the term 'rivalry' does not refer to bistable percepts but instead to the conflict for binding as described here). That is, stronger binding (i.e., in the case of a visual lead or audiovisual synchrony; see Fig. 1A and B) will lead to an increased rivalry with the upcoming stimulus, while weaker binding (i.e., auditory lead; see Fig. 1C) will lead to a decreased rivalry. Binding rivalry is hypothesized as a determinant of the strength of the SIFI: higher rivalry is expected to result in lower illusory percepts and slow reaction times (RTs), while lower rivalry is expected to result in higher illusory percepts and quicker RTs. Binding is highly dependent on timing, thus, rivalry between the unequal number of stimulus inputs is expected to subside with distal in time presentations. In the case that these parameters lead to differential illusory robustness across different timing presentations, then a potential refinement of the optimal integration theory will be put forward so as to better account for the SIFI and the fusion illusion.