National and Kapodistrian
University of Athens
Department of Philosophy and History of Science
Postgraduate Program Basic and Applied Cognitive Science

"Crossmodal Binding Rivalry:
A "race" for integration between unequal sensory inputs"

by

Maria Kostaki

Student Registration Number: 12M07

Master Thesis

Supervisory Committee

Athanassios Protopapas

Argiro Vatakis

Konstantinos Moutoussis

Athens

June, 2015

“Crossmodal Binding Rivalry:

A “race” for integration between unequal sensory inputs”

2015

Supervisory Committee

Athanassios Protopapas
Associate Professor
Department of Philosophy and History of Science

Argiro Vatakis
Researcher
Cognitive Systems Research Institute

Konstantinos Moutoussis
Associate Professor
Department of Philosophy and History of Science

Maria Kostaki

“Crossmodal Binding Rivalry:

A “race” for integration between unequal sensory inputs”

*

Master Thesis

Postgraduate Program Basic and Applied Cognitive Science

National & Kapodistrian
University of Athens

This page was intentionally left blank.

Acknowledgments

I would like to express my sincere thanks to Prof. Athanassios Protopapas, Dr. Argiro Vatakis, and Prof. Konstantinos Moutousis for their valuable comments on my thesis and the excellent collaboration we had during my Master's studies. My special thanks go to Dr. Argiro Vatakis as this thesis would not have been possible without her unwavering support, guidance, and encouragement. She was always eager to discuss any difficulties I encountered and help me solve them. She is a great person and scientist and working with her was an amazing experience that I will always remember with the warmest feelings of gratitude and respect. I sincerely thank her for the opportunity she gave me to work with her and her true desire to keep me on track. I would also like to thank my parents, my sister Matoula, and my friends for their patience and constant support. Last but not least, I would like to thank all the participants who took part in the experiments of this thesis for their invaluable help.

Part of this work was presented as a poster at the *International Conference on Timing and Time Perception (ICTTP2014),* Corfu, Greece and the *International Multisensory Research Forum 2014,* Amsterdam, The Netherlands.

The completion of this project was co-funded by the Project of "SSF Scholarships" from the resources of the OP of "Education and Lifelong Learning", by the European Social Fund (ESF) of NSRF, 2007-2013.

This page was intentionally left blank.

Table of Contents

Abstract

We have used the classic experimental set-up of the sound induced visual fission and fusion illusions with multiple temporal configurations and timing presentations to test whether a rivalry between the unequal number of sensory inputs could refine the theory of optimal integration for these illusions. This ‘crossmodal binding rivalry’ hypothesis depends on the binding of the first audiovisual stimulus pair and its temporal proximity with the upcoming unisensory stimulus and is based on the multisensory integration rules. According to these rules binding has differential robustness depending on the sensory inputs and their spatiotemporal relationship within the temporal window of integration (e.g., binding is more resilient to visual leads as compared to auditory ones). We, therefore, expect that strong binding of the first audiovisual pair will lead to strong rivalry with the upcoming unisensory stimulus and as a result weaker illusory percepts while weak binding will lead to less intense rivalry and stronger illusory percepts. The data revealed differential illusory strength across different temporal configurations for the fission illusion while we also replicated and extended previous findings on the effect of visual acuity in the strength of the illusion, showing that poor visual acuity (poor discrimination of two flashes presented) resulted in enhanced illusory percepts.

Keywords: multisensory integration, double flash illusion, crossmodal binding rivalry

This page was intentionally left blank.

1. Introduction

Our brain has the ability to integrate information from different modalities originating close in time and space (e.g., Stein & Meredith, 1993). Integration for sensory signals that are simple (e.g., flash of an LED) and equal in number (e.g., one visual and one auditory) is usually quite straightforward resulting in enhanced detectability of a target or faster reaction times (e.g., Forster, Cavina-Pratesi, Aglioti, & Berlucchi, 2002; Stein, Lagondon, Wilkinson, & Price, 1996). The story, however, becomes more complicated when the sensory signals are unequal in number. It has been hypothesized that the brain resolves the binding of unequal sensory inputs according to Bayesian rules relying on the modality that has the most reliable information at hand (Ernst & Bülthoff, 2004). Thus, for instance, the use of a spatial or temporal task will result in visual or auditory dominated percept, respectively (e.g., Alais & Burr, 2004; Repp & Penel, 2002). A well-known example of auditory dominance over vision is the sound-induced flash illusion (SIFI), where a single flash in the presence of two beeps is perceived as two distinct flashes (e.g., Shams, Kamitani, & Shimojo, 2000), as well as the fusion illusion, where two flashes presented with one beep are “fused” to a single flash (e.g., Andersen, Tiipana, & Sams, 2004). Shams, Mab, and Beierholm (2005) proposed that these illusory percepts can be accounted for by optimal integration, where the reliability of audition in time dominates the visual percept (see also Apthorp, Alais, & Boenke, 2013; Cuppini, Magosso, Bolognini, Vallar, & Ursino, 2014; Roseboom, Kawabe, & Nishida, 2013).

The optimal integration account has provided an explanation for a number of multisensory phenomena, however, in regards to SIFI some findings cannot be fully covered under the wide ‘umbrella’ of optimality in integration. Firstly, SIFI has not always been robust across participants between or even within studies. Some participants tend to be highly susceptible to the classical presentation of the illusion (i.e., where the beep is 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 research showed 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 ones visual acuity baseline (Apthorp et al., 2013). There are also studies that have not treated or considered susceptibility and/or visual acuity differences and have included all participants in their analysis (e.g., Andersen et al., 2004) and studies that have split their participants in those who could perceive the illusion and those who could not and analyzed their results in groups (Mishra, Martinez, Sejnowski, & Hillyard, 2007; Mishra, Martinez, & Hillyard, 2008). It is as yet unclear what promotes this differential participant susceptibility, it could, however, be potentially 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 have shown longer alpha rhythm wavelength in occipital activity (i.e., longer TWIs) to result to increased susceptibility to the illusion at longer stimulus onset asynchronies (SOAs).

Moreover, optimal integration (as currently defined) does not take into account the source of origin of the inducing beeps (i.e., assumes that beeps come from the same source) and as a result cannot explain why two auditory beeps of different pitch (e.g., noise and pure-tone, 300 and 3500 Hz tone) when presented with a single flash, enhance the detection of the single flash instead of creating a second illusory flash (Roseboom et al., 2013). Also unclear is how stimulus complexity and familiarity lead to a decline of the experienced illusion. For example, when abstract visual (i.e., visual pattern formed by squares; Takeshima & Gyoba, 2013) or familiar stimuli (such as faces or buildings) are compared with simple (i.e., flashes) or unfamiliar ones, respectively (Setti & Chan, 2011), illusory percepts decrease. Thus, optimal integration -in general- may not provide a full explanation of the SIFI and fusion illusions or may need to be further refined by taking into consideration other aspects related to multisensory integration and sensory interaction.

Apart from the above-mentioned challenges to the optimal integration theory, Mishra, Martinez, and Hillyard (2013) recently posed yet another one: temporal positioning and proximity modulates 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 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 segregation increased, while when the second beep followed the second flash, segregation decreased). Thus, Mishra et al.'s results suggest, for the first time, that the temporal relation of audiovisual inputs within the TWI may also alter the illusory visual percept of crossmodal conditions. To-date, research on the SIFI has not found evidence of differential illusory strength as a function of the temporal relation of the audiovisual input (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 on the literature, however, reveals that no research so far has ever directly compared all the temporal positions that auditory and visual inputs can take in the SIFI and fusion illusions. Such comparison will allow one to clarify how the temporal presentation and temporal positioning of the different sensory inputs modulate the strength of the illusion.

In the present study, therefore, we aim, for the first time, to evaluate all potential configurations that SIFI and fusion illusion can take across the same participants and at different proximities (i.e., SOAs) using the classic experimental set-up of the SIFI (Shams et al., 2000) and fusion illusions (Andersen et al., 2004). The integration of these incongruent percepts could potentially be dominated by audition, which is indeed more reliable than vision for this temporal task (e.g., Andersen et al., 2004; Wada, Kitagawa, & Noguchi, 2003). In such a case, one would expect equal 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 "rules" could also provide a more thorough explanation of the phenomenon. The candidate "rules" 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 pairing (Vatakis, 2013), and c) weakened tolerance for larger temporal distances between audiovisual inputs (i.e., the farther apart a flash and a beep are the less likely we are to treat them as an audiovisual pair; e.g., Vatakis & Spence, 2010) – even within the TWI.

We, therefore, hypothesize that in the presence of unequal numbers of sensory input, a rivalry between those inputs will arise depended on the binding of the first audiovisual stimulus pair and its temporal proximity with the next unisensory stimulation. That is, strong binding (i.e., in the case of visual lead or audiovisual synchrony; see Figure 1A, B) will lead to a strong rivalry with the upcoming stimulus, while weak binding (i.e., auditory lead; see Figure 1C) will lead to lower rivalry levels. Binding rivalry is hypothesized as a determinant of the strength of the SIFI: strong rivalry is expected to result in weak illusory percepts and slow reaction times (RTs), while weaker rivalry is expected to result in strong illusory percepts and quicker RTs. Binding is highly dependent on timing, thus, rivalry between the unequal number of stimulus input is expected to subside with distal in time presentations. If these rules apply and differential illusory robustness across different timing presentations is observed, then a possible refinement of the optimal integration theory will be put forward so as to incorporate additional rules accounting for the SIFI and fusion illusions.

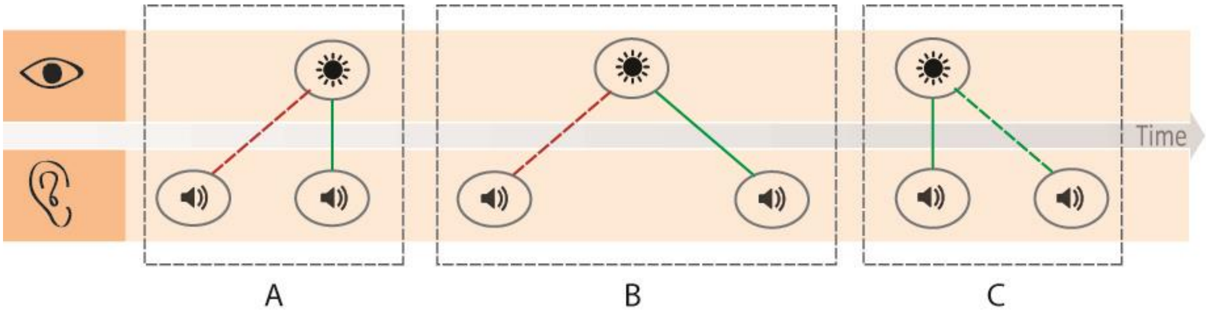

Figure 1. A schematic presentation of the binding between unequal numbers of sensory inputs. Dashed lines represent weak binding (binding weakness is graded with red dashed lines indicating weaker binding as compared to green dashed lines), while straight lines represent strong binding between auditory and visual stimuli. A) An example of a strong rivalry between the initially presented auditory input and the synchronous audiovisual pair. B) An example of a weak rivalry between the initially presented auditory input and the asynchronous audiovisual pair. C) An example of a strong rivalry between the initially presented synchronous audiovisual pair and the subsequent auditory input.

2. Methods

2.1 Participants

Thirty-seven naïve volunteers (age range: 18-35, M = 26.3 years of age, 25 females) took part in the experiment. All participants reported normal or corrected-to-normal vision and normal hearing. Three participants were excluded from all further analysis due to inappropriate completion of the task (i.e., continuous pressing of one response type for all conditions).

2.2 Stimuli and apparatus

The experiment was comprised of the classic experimental conditions of the double flash illusion (Shams et al., 2000, 2002) with multiple stimulus onset asynchronies (SOAs) between the visual and auditory presentations. The visual stimulus consisted of a uniform white disk (from now on ‘flash’ or F) on a black background, subtending 3° of the visual field at 3° eccentricity below the fixation point. The flash duration was approximately 23 ms and the SOA between two successive flashes was constant at approximately 73 ms. The tone (1850Hz; from now on ‘beep’ or B) was 7 ms in duration.

The experiment was composed of four unimodal conditions (i.e., 0F1B, 0F2B, 1F0B, 2F0B) and four multimodal conditions (i.e., 1F1B, 2F2B, 1F2B, 2F1B). In the unimodal conditions, 0F were accompanied by either 1B or 2B (see Figure 2A). The SOA between the two successive beeps was constant at 50 ms. In the multimodal conditions, stimulus presentation varied by SOAs of 0, ±25, ±50, ±100 ms (with 0 ms indicating that the first flash was in synchrony with the first beep and negative SOAs indicating beep-first presentations; see Figure 2B, 2C). For example, in the case of 1F2B, the stimulus temporal sequence could be as follows: a) the first beep presented before the flash and the second beep in synchrony with the flash (e.g., -25 ms | 0 ms, from now on ‘left’ or L condition), b) the first beep presented before the flash, while the second after the flash (e.g., -25 ms | +25 ms, from now on ‘middle’ or M condition), or c) the first beep presented in synchrony with the flash and the second beep after the flash (e.g., 0 ms | +25 ms, from now on ‘right’ or R condition; Figure 2B).

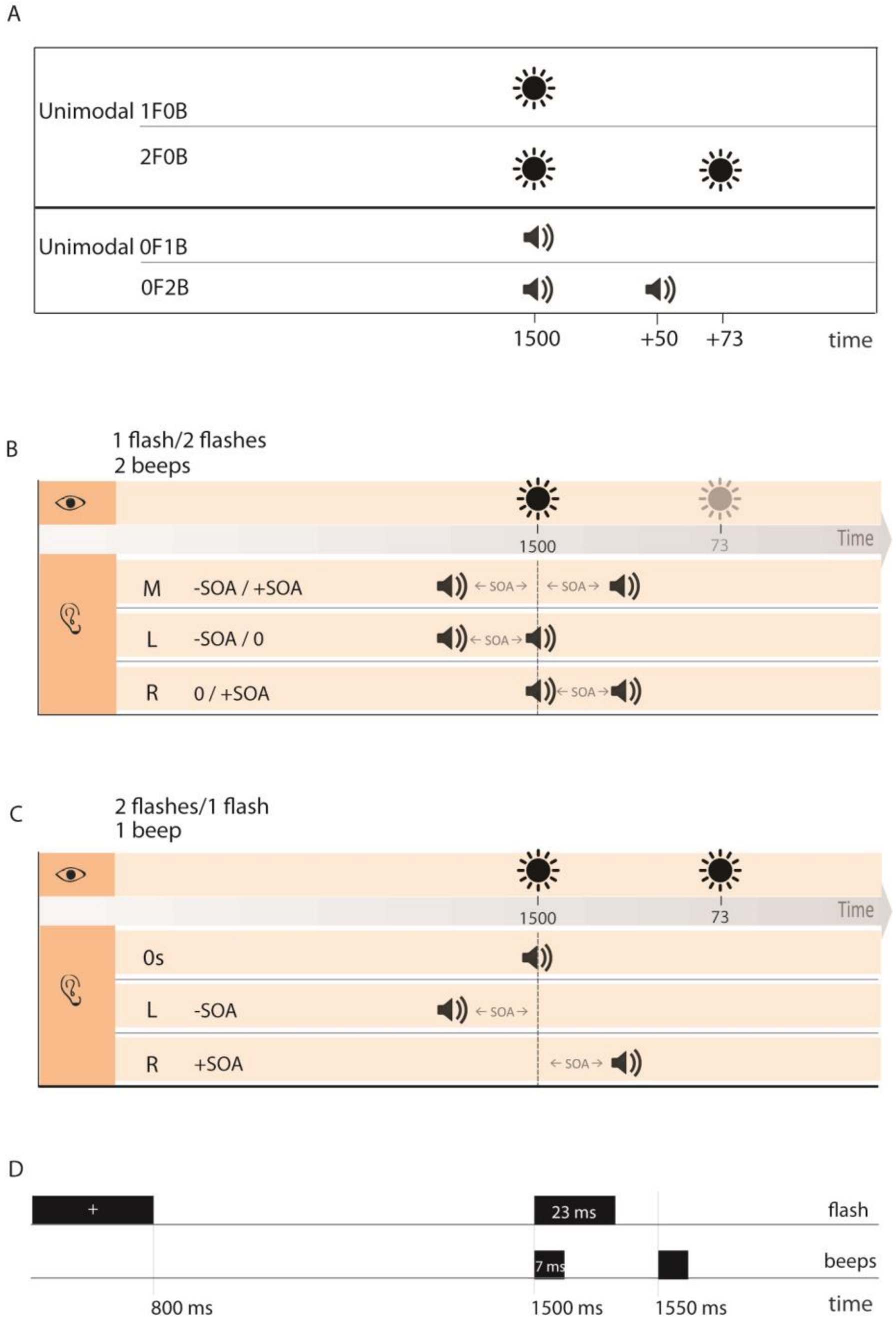

Figure 2. A schematic presentation of the various audiovisual timing conditions utilized in the experiment: A) 1F1B and 1F2B illusion conditions with multiple SOA presentations. B) 2F1B illusion and 2F2B conditions with multiple SOA presentations. C) 1F0B, 2F0B, and 0F1B, 0F2B unimodal conditions. D) An example of the timing presentation of the 1F2B-50R illusion condition. Note that the SOAs were measured from the onset of the first flash.

A total of 36 different conditions of unimodal and multimodal conditions were presented with 30 repetitions for each. The experiment was divided in two blocks with breaks in between. The experiment was performed using Presentation (Version 16.4, Neurobehavioral systems, Inc.). A CRT monitor of 75Hz refresh rate was used for visual stimulation, while the auditory stimuli were presented through headphones. The order of stimulus presentation was randomized.

2.3 Experimental procedure

Participants were seated approximately 60 cm from the screen in a dedicated dimly-lit room. They were instructed to fixate on the fixation point on the center of the screen and report the number of flashes (0, 1 or 2) presented by pressing the corresponding key on the PC keyboard. Participants were explicitly told to keep their eyes on the fixation point until the end of the trial and respond only to the number of flashes presented while ignoring the auditory stimulation. Written instructions were also provided on the computer screen before the start of the experiment.

The initiation of the experiment was self-paced. A ‘Ready’ screen was presented and participants were instructed to press ‘Enter’ for the experiment to start. Each trial began with the fixation point which was presented for 800 ms, a 700 ms blank screen, and, subsequently, the presentation of a given condition (see Figure 2D). Participants had to respond in order to advance to the next trial, thus there was no time limit for responding; they were, however, instructed to respond as quickly as possible. Before the main experiment, a short practice block was given to the participants in order to familiarize them with the process. The experiment lasted approximately 40 minutes (20 min per block).

3. Results

3.1 Analysis

A large portion of the participants performed poorly in the detection of two flashes (i.e., 2F), while an equally large portion performed accurately. Given the large number of participants in both groups and the fact that multisensory research has demonstrated enhanced detection of a stimulus in the presence of another sensory (even if it's irrelevant) stream (e.g., Noesselt et al., 2010; Perez-Bellido, Soto-Faraco, & Lopez-Moliner, 2013), no participants were rejected. So far researchers either rejected participants that fused the 2F condition into one flash (see Fiedler et al., 2011; Rosenthal, Shimojo, & Shams, 2009) or ignored performance at this condition excluding participants with poor 1F2B illusory percepts (see Mishra, Martinez, & Hillyard, 2013). In our study, participants were divided in two separate groups for analysis with the expectation that those who had difficulty in detecting the two unimodal flashes (i.e., poor visual acuity) would experience higher illusory percepts in the multimodal conditions with unequal number of stimulation. Thus, two groups were created with Group 1 including participants whose performance in 2F was over 50% (N = 17; age range: 18-35; M = 27; 11 females), while Group 2 had participants whose performance was below 50% (N = 17; age range: 18-35; M = 25.6; 13 females). The percentage of correct responses and reaction times (RTs) were analyzed using a repeated measures analysis of variance (ANOVA) and Bonferroni corrected t-tests were used for all post-hoc comparisons.

3.2.1 Unimodal - multimodal (non illusory) conditions

Participant accuracy was analyzed with the between-subjects factor of Group (Group 1 vs. 2) and the within-subjects factors of Modality (auditory, visual, audiovisual) and Stimulus presentation (single or double). Significant main effects of Modality [$F(2,64) = 99.78$, $p < 0.001$, $\eta^2 = 0.76$], Stimulus presentation [$F(1,32) = 48.07$, $p < 0.001$, $\eta^2 = 0.60$], and Group [$F(1,32) = 31.14$, $p < 0.001$, $\eta^2 = 0.49$] were obtained. That is, participants were more accurate in detecting the absence of visual stimulation (i.e., auditory-only presentations; M = 0.96) as compared to its presence (M = 0.75) or when audiovisual (M = 0.91), and for audiovisual as compared to visual only presentations (i.e. sound enhances visual target detection as expected from multisensory integration). Additionally, accuracy was higher for the detection of single as compared to double stimulus presentations (M = 0.94 and 0.81, respectively) and for Group 1 (M = 0.93) as compared to Group 2 (M = 0.82), thus reflecting the differences mentioned above.

A significant triple interaction of Modality, Stimulus presentation, and Group [F(2,64) = 46.30, $p < 0{,}001$, $\eta^2 = 0.59$] verified that Group 1 was more accurate in reporting 2F presentations (M = 0.86) as compared to Group 2 (M = 0.29; see Figure 3). These results were also depicted in the significant interactions of Modality and Group [F(2,64) = 53.53, $p < 0.001$, $\eta^2 = 0.63$], Stimulus presentation and Group [F(1,32) = 33.53, $p < 0.001$, $\eta^2 = 0.51$], and Modality and Stimulus presentation [F(2,64) = 78.12, $p < 0.001$, $\eta^2 = 0.71$].

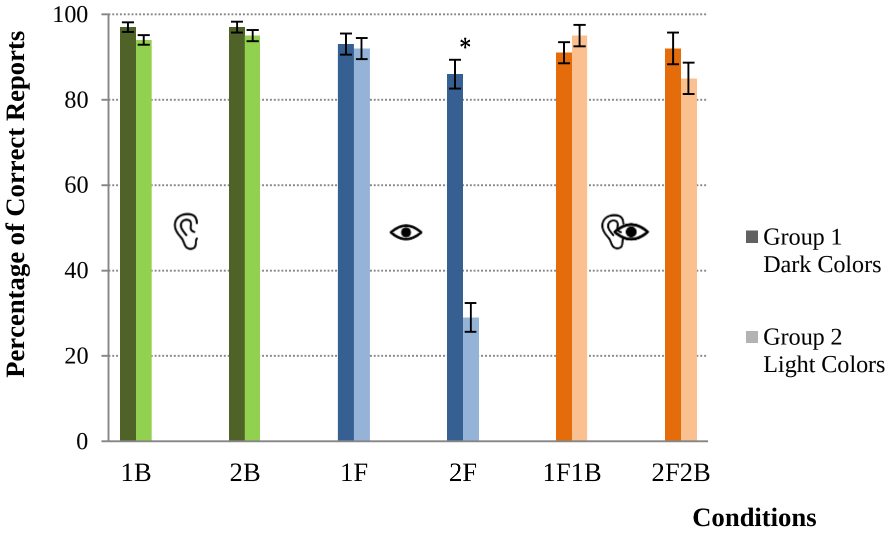

Figure 3. Mean percentage of correct responses in unimodal and multimodal conditions are plotted as a function of Modality (auditory - A, visual - V, audiovisual - AV), Stimulus Number (1 or 2), and Group (Groups 1 and 2). Significant differences between groups ($p<0.05$) are highlighted by an asterisk.

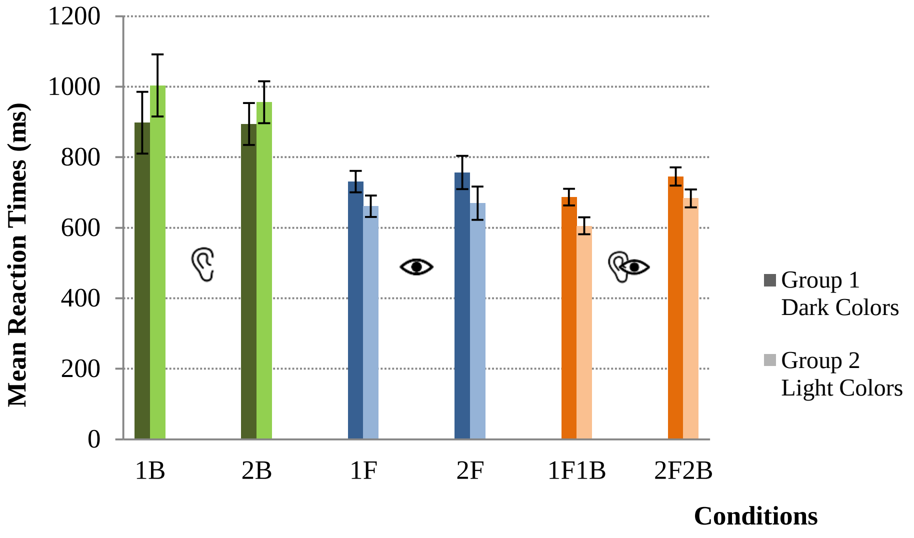

Figure 4. Mean reaction times (ms) in unimodal and multimodal conditions plotted as a function of Modality, Stimulus Number, and Group.

Similar analysis of the participant RTs resulted in a main effect of Modality [F(2,64) = 23.00, p < 0.001, η^2 = 0.42; Greenhouse-Geisser correction], with participants being faster in the visual (M = 704.36 ms) and audiovisual conditions (M = 680.01 ms) as compared to the auditory condition (M = 937.93 ms; see Figure 4). The slower performance in the auditory condition could probably be due to the absence of visual stimulation, thus participants reacting slower by waiting for a potential visual presentation. No main effect of Stimulus presentation [F(1,32) = 1.46, p = 0.23, η^2 = 0.04] or Group [F(1,32) = 0.22, p = 0.64, η^2 = 0.007] was obtained. A significant interaction between Modality and Stimulus presentation was found [F(2,64) = 3.34, p = 0.04, η^2 = 0.09] with participants responding faster in the 1F1B condition as compared to the 2F2B (M = 645.93 and 714.09 ms, respectively). All other interactions were not significant.

3.2.2 1F2B illusion condition

Participant illusory reports (i.e., reporting two flashes instead of one) were analyzed with the between-subjects factor of Group (Group 1 vs. 2) and the within-subjects factors of

SOA (i.e., the timing between auditory and visual presentations; 25, 50, and 100 ms) and Sound position (in relation to the single flash; M, L, and R; see Figure 2B). Significant main effects of SOA [F(2,64) = 7.36, $p = 0.005$, $\eta^2 = 0.187$; Greenhouse-Geisser correction] and Sound position [F(2,64) = 9.008, $p = 0.001$, $\eta^2 = 0.22$; Greenhouse-Geisser correction] were obtained, and more importantly, a marginal main effect of Group was found [F(1,32) = 3.91, $p = 0.057$, $\eta^2 = 0.11$] with Group 2 experiencing higher illusory percepts as compared to Group 1 (M = 0.64 and 0.47 respectively). The analysis showed that participants had significant higher illusory reports in the 50 and 100 ms conditions (M = 0.59 for both) as compared to the 25 ms (M = 0.48; see Figure 5). Moreover, participants were more susceptible to the illusion in the M and L condition than in the R (M = 0.59, 0.56, and 0.51, respectively), showing that the temporal position of the stimuli *does* affect the strength of the illusion. Significant interactions of SOA and Sound position [F(4,128) = 14.88, $p < 0.001$, $\eta^2 = 0.32$; Greenhouse-Geisser correction], SOA and Group [F(2,64) = 4.84, $p = 0.011$, $\eta^2 = 0.13$], and Sound position and Group [F(2,64) = 8.02, $p = 0.001$, $\eta^2 = 0.20$] were also obtained. Specifically, in the 25 ms presentations higher illusory percepts were obtained in the M condition (M = 0.60) as compared to the L and R conditions (M = 0.43 and 0.41, respectively). For the SOA of 50 ms, a marginally significant difference ($p = 0.069$) was obtained with the illusion being higher in the M condition (M = 0.64) as compared to the R condition (M = 0.54). For the 100 ms presentations, participants had significantly higher illusory percepts in the L condition as compared to the M and R positioning (M = 0.65, 0.55, and 0.58, respectively). Additionally, Group 2 (i.e., with poor 2F performance) reported higher illusory percepts in the 100 ms (M = 0.73) as compared to Group 1 (M = 0.46) and Group 2 experienced higher illusion in the M condition (M = 0.72) as compared to Group 1 (M = 0.46), thus, in some cases, the influence of the sound is greater for those with poor performance in the 2F presentation. Possibly, due to low visual resolution, participants in

Group 2 are more susceptible to the auditory stimulation they receive but this phenomenon does not seem to be consistent across the different conditions. All other interactions did not reach significance. RT analysis showed no significant main effects or interactions.

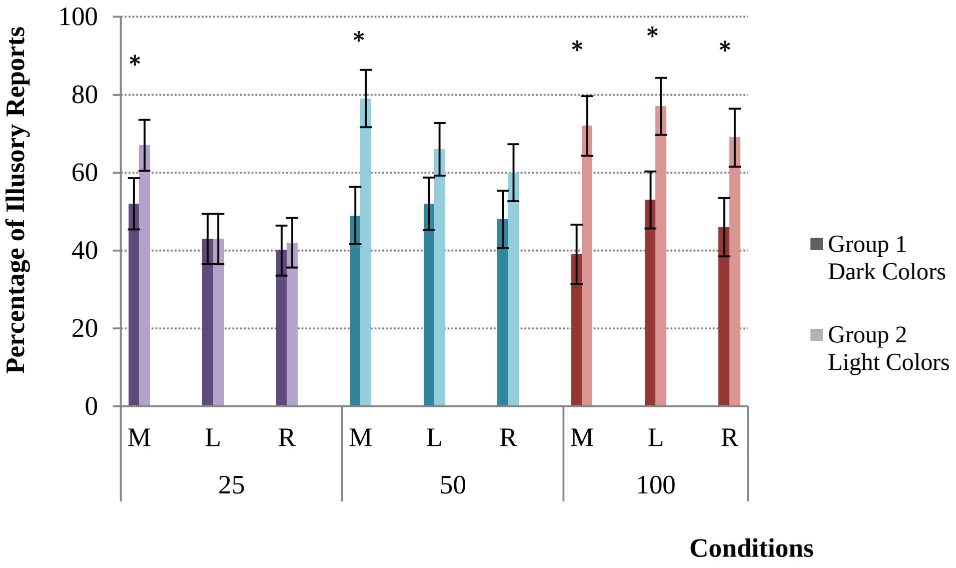

Figure 5. Mean percentage of illusory percepts in 1F2B condition plotted as a function of SOA (25, 50, 100 ms), Sound position (M, L, R) and Group (Groups 1 and 2). Significant differences between groups ($p<0.05$) are highlighted by an asterisk.

3.2.3 2F1B Illusion Condition

Participant illusory reports (i.e., reporting one flash instead of two) were analyzed with the between-subjects factor of Group (Group 1 vs. 2) and the within-subjects factor of SOA (25, 50, and 100 ms) and Sound position (L, R; see Figure 2B). Significant main effects of SOA [$F(2,64) = 72.34$, $p < 0.001$, $\eta^2 = 0.69$], Sound position [$F(1,32) = 54.56$, $p < 0.001$, $\eta^2 = 0.63$], and Group [$F(1,32) = 42.67$, $p < 0.001$, $\eta^2 = 0.57$] were obtained. That is participants' illusory reports were higher in 25 and 50 ms ($M = 0.66$ and 0.67, respectively) as compared to 100 ms ($M = 0.51$; see Figure 6). Moreover, participants' illusory reports were higher in the L conditions ($M = 0.67$), where the beep preceded the first flash than in R conditions (M

= 0.56), where the beep was following the first flash (note that in R conditions the beep could be presented either between the two flashes -25R and 50R- or after the two flashes -100R; see Figure 2C). Also, participants in Group 2 (M = 0.84) had significantly higher illusory percepts as compared to those in Group 1 (M = 0.39). A significant interaction of SOA and Sound position [F(2,64) = 91.41, $p < 0.001$, $\eta^2 = 0.74$] was also obtained with participants reporting higher illusory percepts in the L (M = 0.66) as compared to the R condition in 100 ms (M = 0.37). No other significant interactions were obtained.

We also checked whether the differences between the two groups for this illusory condition were due to decision bias (β) rather than perceptual sensitivity (d′). A signal detection analysis was performed (Macmillan & Creelman, 1991) for the 2F1B condition. Sensitivity [$d' = z(H) - z(FA)$] and response bias [$\beta = 0.5*(z(H) + z(FA))$] (where z(p) denotes the inverse of the cumulative Normal distribution) were calculated in each group for each 2F1B condition. Response rates of correct detection of a single flash were categorized as ‘hits’ (H) and response rates of incorrect detection of a single flash as ‘false-alarm’ (FA). Incidents of $p = 0$ and $p = 1$ were approximated by 1/N and 1 - (1/N), respectively, where N is the number of trials tested. Repeated measures analysis of the results revealed a main effect of Group both in sensitivity [F(1,32) = 27.63, $p < 0.001$, $\eta^2 = 0.46$] and criterion [F(1,32) = 27.72, $p < 0.001$, $\eta^2 = 0.46$] showing that participants in Group 2 had difficulty (d′ = 0.35) in discriminating the two flashes and that their criterion was shifted towards audition (β = -1.63; negative criterion values indicate bias towards audition) as compared to those in Group 1 who could more easily discriminate between the two flashes (d′ = 1.80) and were less biased by audition (β = -0.71).

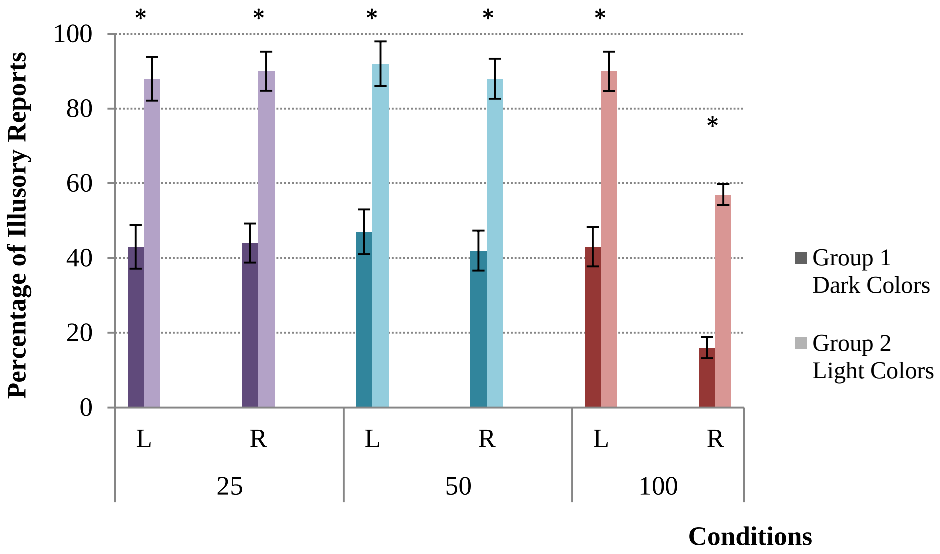

Figure 6. Mean percentage of illusory percepts in 2F1B condition plotted as a function of SOA, Sound position (L, R), and Group (Groups 1 and 2). Significant differences between groups (p<0.05) are highlighted by an asterisk.

The RTs analysis showed a significant main effect of Sound position [F(1,32) = 9.63, p = 0.004, η^2 = 0.23], with faster responding in the L condition (M = 644.47 ms) as compared to the R condition (M = 689.22 ms). No main effects of SOA [F(2,64) = 2.19, p = 0.14, η^2 = 0.06; Greenhouse-Geisser correction] or Group [F(1,32) = 0.58, p = 0.45, η^2 = 0.02] were found. The only significant interaction was that of SOA and Group [F(2,64) = 4.13, p = 0.02, η^2 = 0.11], but no other differences were found.

3.2.4 Illusory vs. Actual Presence of Two Flashes (1F2B vs. 2F2B)

To test whether the illusory double flash percept is equivalent (in terms of performance) to the veridical percept of two flashes across the two participant groups, participant illusory versus veridical reports were analyzed with the between-subjects factor of Flash presence (Illusory vs. Actual), SOA (25, 50, and 100 ms), and Sound position (M, L, and R) and the within-subjects factor of Group (Group 1 vs. 2). Significant main effects of Flash presence [F(1,32) = 55.64, p < 0.001, η^2 = 0.63], SOA [F(2,64) = 9.04, p = 0.003, η^2 = 0.22;

Greenhouse-Geisser correction], and Sound position [F(2,64) = 10.10, p = 0.001, η^2 = 0.26; Greenhouse-Geisser correction] were obtained, while no main effect of Group [F(1,32) = 0.10, p = 0.75, η^2 = 0.003] was found. Specifically, participants reported higher percentage of two flashes in 2F2B conditions (M = 0.83) as compared to the 1F2B illusion conditions (M = 0.55), showing that the illusory percepts of two flashes (i.e., 1F2B conditions) are not as robust as the veridical percepts of the two flashes (i.e., 2F2B conditions; see Figure 7). Moreover, participants reported higher percentage of two flashes, irrespective of the condition they belonged to (illusory or veridical), when the SOA between the visual and auditory presentations was 50 and 100 ms (M = 0.72 and 0.73, respectively) as compared to 25 ms (M = 0.63), as well as in the M and L conditions (M = 0.73 and 0.69, respectively) as compared to the R condition (M = 0.66).

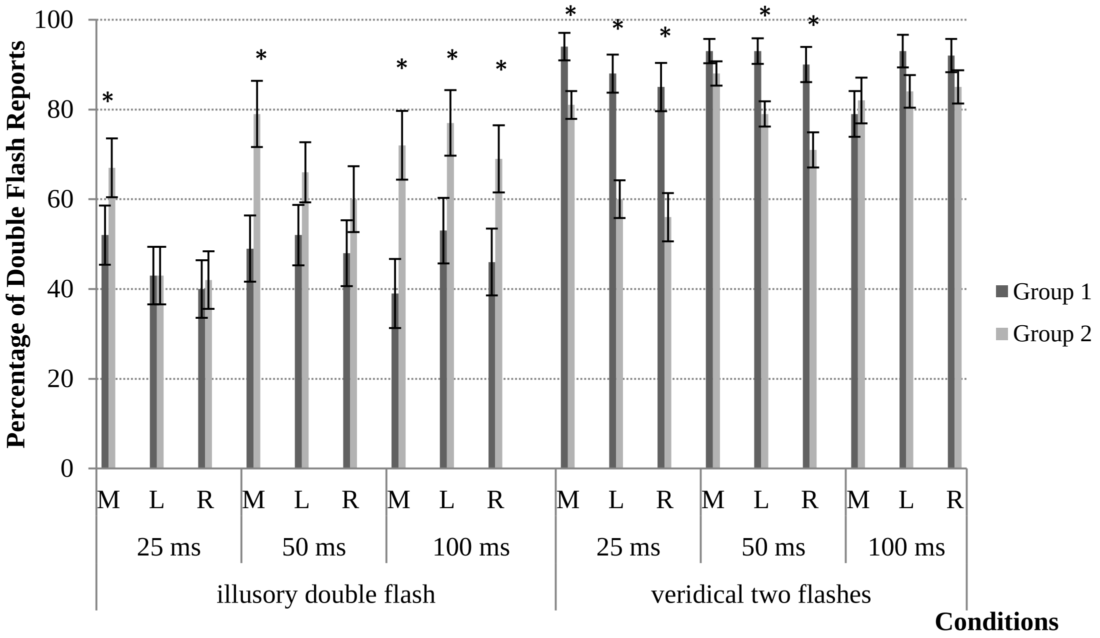

Figure 7. Mean percentage of double flash reports plotted as a function of Flash presence (illusory vs. veridical presentations; i.e., 1F2B vs. 2F2B conditions, respectively), SOA, Sound position (M, L, R), and Group (Groups 1 and 2).

3.2.5 Illusory vs. Actual Presence of a Single Flash (2F1B vs. 1F1B)

Similar to the previous analysis, we were also interested to test whether the illusory percept of a single flash is equivalent (in terms of performance) to the veridical percept of a single flash across the two participant groups. Thus, participant illusory versus veridical reports were analyzed with the between-subjects factor of Flash presence (Illusory vs. Actual), SOA (25, 50, and 100 ms), and Sound position (L vs. R) and the within-subjects factor of Group (Group 1 vs. 2). Significant main effects of Flash presence [F(1,32) = 83.57, $p < 0.001$, $\eta^2 = 0.72$], SOA [F(2,64) = 57.69, $p < 0.001$, $\eta^2 = 0.64$], Sound position [F(1,32) = 66.77, $p < 0.001$, $\eta^2 = 0.68$], and Group [F(1,32) = 37.60, $p < 0.001$, $\eta^2 = 0.54$] were obtained. The results showed that participants reported higher single flash percentage in the 1F1B conditions (M = 0.94) as compared to the 2F1B illusion conditions (M = 0.62; see Figure 8), a result similar to the previous analysis. Moreover, participants had higher percentage of single flash reports when the SOA was 25 and 50 ms (M = 0.80 and 0.80, respectively) as compared to 100 ms (M = 0.73) and in the L as compared to the R condition (M = 0.81 and 075, respectively). Group 2 had higher single flash reports as compared to Group 1 (M = 0.90 and 0.66, respectively).

A significant interaction of Flash presence and Group [F(1,32) = 34.67, $p < 0.001$, $\eta^2 = 0.52$] was obtained with participants in Group 1 being less susceptible in reporting one flash in the illusion condition as compared to those in Group 2 (M = 0.39 and 0.84, respectively). A significant triple interaction of Flash presence, SOA, and Sound position [F(2,64) = 55.62, $p < 0,001$, $\eta^2 = 0.63$] verified that participant single flash reports were greater for veridical flashes (i.e., 1F1B) as compared to the illusory fusion of the two flashes (i.e., 2F1B) both in the L and R conditions across all SOAs (M = 0.94 and 0.65 ms in 25 L; 0.93 and 0.67 ms in 25 R; 0.94 and 0.69 ms in 50 L; 0.93 and 0.65 ms in 50 R; 0.95 and 0.66 ms in 100 L; 0.92 and 0.37 ms in 100 R, for Actual and Illusory single flash percepts, respectively). These results were also depicted in the significant interactions of Flash presence and Sound position [F(1,32) = 23.80, $p < 0.001$, $\eta^2 = 0.43$], Flash presence and SOA [F(2,64) = 61.56, $p < 0.001$, $\eta^2 = 0.66$], and SOA and Sound position [F(2,64) = 84.76, $p < 0.001$, $\eta^2 = 0.73$].

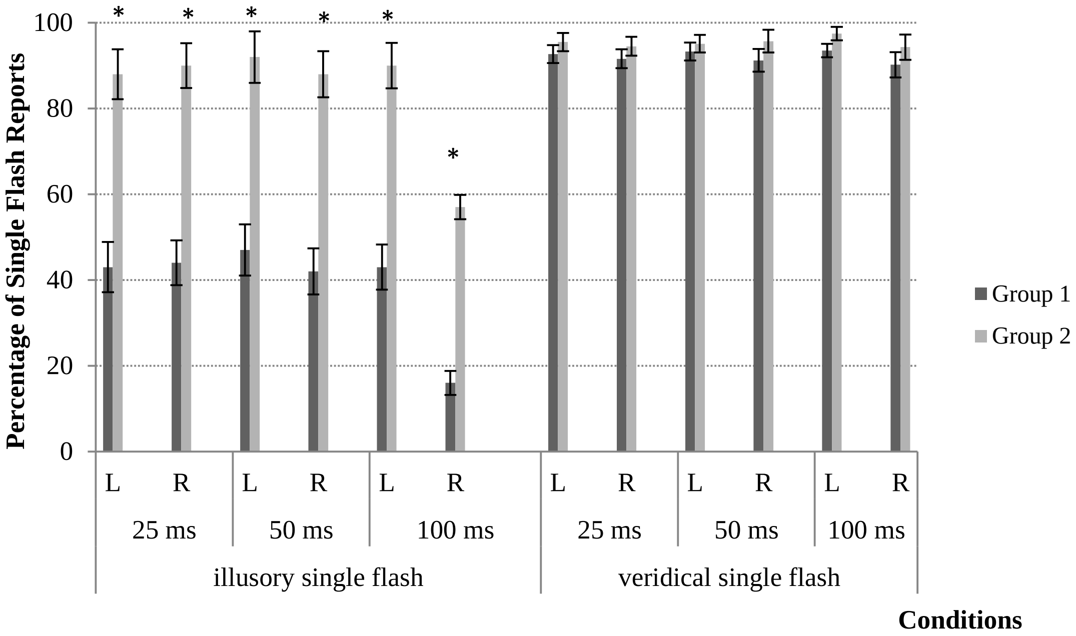

Figure 8. Mean percentage of single flash reports plotted as a function of Flash presence (illusory vs. veridical presentations; i.e., 2F1B vs. 1F1B conditions, respectively), SOA, Sound position (M, L, R), and Group (Groups 1 and 2).

4. Discussion

In the present study, we evaluated, for the first time, all potential temporal configurations of the SIFI and fusion illusion across the same participants and at different proximities (i.e., SOAs) in order to examine whether a ‘crossmodal binding rivalry’ (i.e., a rivalry between unequal number of sensory inputs that depends on the binding of the first audiovisual stimulus pair and its temporal proximity with the upcoming unisensory stimulus) contributes in ones susceptibility to the illusions. Our results showed that indeed the temporal positioning and temporal proximity of the auditory and visual stimuli affect the strength of the illusion experienced in both phenomena (i.e., SIFI, fusion illusion). Moreover, we examined how ones ability to discriminate between two flashes affects the strength of the SIFI and fusion illusions across different temporal configurations and proximities. Our results are in line with previous research, supporting that poor discrimination capabilities results in higher susceptibility in the fusion illusion (Mishra et al., 2008), and also extend these findings for the SIFI, where in some conditions participants' susceptibility to the illusion was affected by poor visual acuity (Kumpik et al., 2014; McGovern et al., 2014). Overall, our results showed that 'crossmodal binding rivalry' and visual acuity are two issues that one has to consider when investigating the optimality of integration in the SIFI and fusion illusion. Optimal integration in these illusions may not only be driven by auditory dominance (e.g., Shams et al., 2005) but also by one's visual acuity and the binding rivalry of the presented sensory inputs.

Manipulation of the temporal positioning of the auditory and visual stimuli resulted in higher illusory percepts, for both illusions, when the auditory stimulus preceded the visual (i.e., M and L conditions; see Figure 2B, 2C) as compared to the cases where the visual stimulus or the synchronous audiovisual pair lead (i.e., R condition). Thus, according to the crossmodal binding rivalry hypothesis, auditory leads (i.e., L condition) result in weaker binding and, thus, weaker rivalry with the upcoming audiovisual pair and higher susceptibility to illusory percepts. The effect, although along the lines of our hypothesis (i.e., weaker binding rivalry leads to higher illusory percepts), was not consistent for all timings tested or across participant groups. That is, in the SIFI, we found higher illusory percepts for the M condition at the SOAs of 25 and 50 ms, however, the same was not true for 100 ms, where the highest illusory reports were noted in the L condition. This latter finding could potentially be accounted for by some kind of temporal ventriloquism effect (Morein-Zamir, Soto-Faraco, & Kingstone, 2003), where the sound attracts the visual stimulus closer to its temporal position possibly enhancing the visual illusion. So far, previous research that have tested the standard configurations of the SIFI (i.e., L and R conditions) across different SOAs (e.g., Foss-Feig et al., 2010; Neufeld et al., 2012; Shams et al., 2002) have found a symmetry in the strength of the illusion depending on the temporal positioning of the stimuli. Our results replicate these findings for the SOAs of 25 and 50 ms, where we also found a symmetric strength of the illusion in the L and R conditions. However, what other studies have missed is the measure of performance in the M condition in comparison with the L and R conditions, which we tested and found to result in higher illusory percepts for the SOAs of 25 and 50 ms. It could be argued that the temporal distance of the two beeps in the M condition is twice as much as the temporal distance in the L and R conditions and, thus, these conditions are not directly comparable. However, based on our hypothesis (i.e., the rivalry between the unequal number of inputs) what matters is not the temporal distance of the two auditory stimuli (as long as one is within the TWI), but the temporal distance of each auditory stimuli from the flash[1]. Thus, temporal positioning remains a candidate factor that modulates illusory susceptibility, thus further specifying optimality in integration. Early neurophysiological data also support the possibility of different underlying causes between the various SIFI configurations. Simon and Gabor (2015), for example, have tested the M and R configurations in an ERP experimental set-up and have found correlation between the physiological and behavioral data only for the M condition suggesting that the underlying mechanisms of the two SIFI configurations may be different.

Concerning the fusion illusion, our results showed that the temporal positioning of audiovisual stimuli affected the strength of the illusion only for the SOA of 100 ms, where

---

[1] To completely eliminate doubts on how the temporal distance between the two beeps in the M condition could affect the strength of the illusion, we conducted a control experiment (N = 8) comparing the M condition with temporal distances at 12 and 25 ms SOA and with mixed SOA at 25 and 12 ms (before and after the flash respectively). Results although they showed a main effect of SOA [$F(2,14) = 5.09$, $p < 0.022$, $\eta^2 = 0.73$] no difference in the strength of the illusion as a function of the temporal distance of the beeps was found in post-hoc comparisons ($M = 0.44$, $M = 0.70$, $M = 0.52$ respectively).

participants showed higher susceptibility to the illusion for auditory leads. This was not the case for 25 and 50 ms. This latter finding may suggest that participants' visual acuity is a stronger determinant for auditory dominance in the fusion illusion irrespective of the temporal positioning of the stimuli at least up to the SOA of 100 ms (where vision is lagging as compared to audition; e.g., Wada, Kitagawa, & Noguchi, 2003). For instance, we found a strong association between participants' visual temporal resolution and susceptibility to the fusion illusion. That is, participants with low visual acuity were almost two times more willing to report one flash instead of two as compared to those with high visual acuity (an effect not observed in SIFI). These results suggest that in the 'race' for integration in the fusion illusion, the "winner" modality is the one with the highest reliability (i.e., auditory modality).

According to previous research (Apthorp et al., 2013; Kumpik et al., 2014; McGovern et al., 2014; Mishra et al., 2008) and the multisensory enhancement obtained for unisensory stimulation (e.g., Stein et al., 1996), we evaluated both participant groups with high and low accuracy in two flash detection. The study of participants with poor visual discrimination ability (that has mostly being ignored in the literature) resulted in valuable information on the relation of visual acuity and the level of the perceived illusion. Specifically, we found that visual acuity was a strong determinant for the fusion percepts rather than for the SIFI[2], where only specific conditions were affected. That is, in fusion illusion participants with low visual acuity were more susceptible across all conditions, while in the SIFI illusion participants with low visual acuity experienced stronger illusory percepts in the temporal configurations of 100 ms and in the M condition across SOAs. Mishra, Martinez, and Hillyard (2008) have also

---

[2] SIFI: Main effect of Group [$F(1,32) = 3.91$, $p = 0.057$, $\eta^2 = 0.11$] with Group 2 experiencing higher illusory percepts as compared to Group 1 ($M = 0.64$ and 0.47 respectively).
Fusion: Main effect of Group [$F(1,32) = 42.67$, $p < 0.001$, $\eta^2 = 0.57$] with Group 2 experiencing higher illusory percepts as compared to Group 1 ($M = 0.84$ and 0.39 respectively)

noted the association of higher illusory percepts in the fusion illusion for participants with flash discrimination difficulty, while this was not the case for those who had higher susceptibility to the SIFI (Mishra et al., 2007). Kumpik and colleagues (2014), however, have proposed that lower visual acuity leads to higher susceptibility for both illusions. The difference between the present results and those of Kumpik et al. may lie to the conditions tested, with the latter study testing spatial characteristics of the visual acuity for an M-type condition in the SIFI, while we tested a multitude of conditions with the flash appearing in the same spatial position. Overall, these results suggest either that poor visual acuity is enhanced in the presence of a single sound -as we hypothesized based on the principles of multisensory integration (e.g., Noesselt et al., 2010)- or that audition becomes the most reliable modality when vision is not reliable. Poor visual acuity and enhanced illusory experiences may also be related to enlarged TWIs. For instance, Stevenson et al. (2012) have shown that the length of the TWI affects the strength of the illusion across participants, with narrow TWIs leading to reduced illusory effects and large TWIs to enhanced illusory effects. McGovern et al. (2014) have also proposed that participants' temporal acuity affects illusory strength by showing that older adults unable to discriminate between two flashes experienced strong illusory percepts in the SIFI for larger SOAs as compared to young adults, while no such differences were observed for the fusion illusion. Thus, although visual acuity seems to affect both the SIFI and fusion illusion, the results across studies are as yet inconsistent with some studies reporting visual acuity to affect both illusions, while others reporting modulations in only one of the two illusions. In the present study, we found that visual acuity promotes several fusion illusion effects, while SIFI is affected only in specific temporal configurations. It is necessary to further elucidate this relationship of visual acuity and perceived illusion in future experimentation.

According to the crossmodal binding rivalry hypothesis, we expected an increase in the strength of the illusion in higher SOAs, given the weaker binding expected at higher temporal distances, which, in turn, would result in weaker stimulus binding rivalries. So far, research has shown that stronger illusory percepts are obtained in lower SOAs (i.e., 25-75 ms) between audiovisual stimuli and that as the SOAs increase (i.e., 75-112 ms) the illusory percepts tends to drop (Shams et al., 2002). In contradiction with previous findings but in alignment with our hypothesis, our data in the SIFI showed that participants –irrespective of group and stimulus configuration- were more susceptible to the illusion for the SOAs of 50 and 100 ms as compared to those of 25 ms. A closer look at the data, however, reveals that participant Group 2 (i.e., those with low visual acuity) is the one that pulls the strength of the illusion to higher percentages in higher SOAs affecting the general trend of illusion across SOAs. Concerning the effect of SOA in the strength of the fusion illusion, we found that for 100 ms the illusory percepts declined when the auditory stimulus followed the presentation of two flashes for both groups (i.e., R condition). This latter result was expected, however, we also expected decreased illusory reports in the L condition of 100 ms as compared to the illusory reports in smaller SOAs since the temporal distance between the beep and the flash was above the average length of the TWI for simple audiovisual stimuli (e.g., Zampini, Shore, & Spence, 2003). This finding may indicate the presence of other effects taking place in the illusion, such as a temporal ventriloquism effect, where a sound slightly preceding the visual input alters the latter’s perceived temporal location (Morein-Zamir et al., 2003) possibly leading to enhanced illusory percepts. This hypothesis warrants further investigation.

The overall participant performance in terms of accuracy in flash detection, indicate –as we hypothesized- that the temporal positioning of audiovisual stimuli within the TWI as well as visual acuity affects the strength of the SIFI and fusion illusion. Our hypotheses on the reaction time data where not, however, verified. That is, for the most part, reaction times did not differ across participants and illusory conditions. The sole exception was noted in the fusion illusion, where higher reaction times were found in the L condition (i.e., auditory lead) as compared to the R, which was opposite to our hypothesized lower reaction times in conditions with less rivalry (i.e., L condition). Mishra et al. (2007) have also measured reaction times across the different illusory and non-illusory conditions, but they also have not reported any significant differences. In general (and due to the limited data currently available), it may not be safe to extract any conclusions in terms of the effect of temporal positioning and rivalry on reaction times. That is, reaction times represent a psychophysical measure of perceptual experience (measuring the time between stimulus and response) and, thus, we can infer that the crossmodal binding rivalry will be depicted in the participants' response time. However, in that time interval between the stimulus presentation and the participants' response, a number of different stages of lower and/or higher level processing may also be involved, resulting in a "blurring" of the actual reaction times recorded (McDonald, Green, Störmer, & Hillyard, 2012; Pachella, 1974). Moreover, we utilized an appearance-based task (i.e., measures of the perceived rather the correct response), thus there is no actual measure of the trade-off between participants' accuracy and the quickness of their response. Therefore, the reaction times recorded may be a "*mixture of fast guess responses that happen to be correct by chance, and stimulus controlled responses*" (Pachella, 1974). That is, the reaction times from "correct", "wrong", or "by chance correct or wrong" responses are averaged resulting in a mixed measure, which is safer to avoid for the interpretation of the potential rivalry between the unequal number of inputs from different modalities.

In terms of the "realness" of the illusory flashes, we found that illusory percepts are not as robust as the veridical percepts of an equal number of audiovisual stimuli. However, participants in Group 2 experienced the illusory (i.e., 2F1B) and the veridical (i.e., 1F1B) single flash at the same level of robustness, which was not the case for the SIFI, thus further demonstrating the asymmetry in how visual acuity affects the two illusions. Contrary to our results, Shams and colleagues (2000, 2002) have reported that the strength of the illusory sound-induced flash was experienced as a veridical flash. Subsequent behavioral research, however, have shown that illusory percepts (both for the SIFI and fusion illusion) are not as robust as veridical percepts (e.g., Andersen et al., 2004; Neufeld et al., 2012; Rosenthal et al., 2009; although we are not aware of the statistical significance of these comparisons). Given that behavioral testing focus on the mean illusory experience, recent imaging studies have compared activations between trials that participants experienced the illusion and baseline conditions (i.e., trials with equal number of audiovisual input -1F1B and 2F2B for fusion illusion and SIFI, respectively) and reported similar activation patterns in V1 (Watkins, Shams, Josephs, & Rees, 2007; Watkins, Shams, Tanaka, Haynes, & Rees, 2006), thus arguing for the similarity of the experienced and the actual flash in V1 an area that could reflect subjective perception of visual stimuli. Concerning the non-illusory conditions (see Figure 2B, 2C for non-illusory configurations), our results showed that depending on the temporal positioning of the audiovisual stimuli the percentage of correct flash responses varied below the perfect score specially for Group 2 (e.g. in the L condition of 25 ms, Group 2 reported 2 flashes only in 60% of trials). This may be an indication that the rivalry between asynchronous audiovisual inputs may not be restricted to unequal number of inputs but also extends to the presence of equal in number audiovisual inputs. However, this proposition needs further examination.

Overall, the present study showed, for the first time, that the temporal positioning of the audiovisual stimuli in the SIFI and fusion illusion affects the strength of the illusion via differential crossmodal binding rivalries. Weaker binding of the initially presented audiovisual pair leads to higher susceptibility to the illusion, while stronger binding of this pair leads to a decrease of the experienced illusion. These results, together with those of Mishra et al. (2013), reveal a new set of parameters that affect the strength of the illusions and have to be taken into consideration when talking about the dominance of audition over vision in unequal stimulus presentations. Visual acuity is also a parameter that cannot be discarded when investigating phenomena such as the SIFI and fusion illusion. The present study extends this research on how visual acuity affects the strength of the illusion showing that difficulties in discriminating between two flashes enhance participants' susceptibility to the SIFI and fusion illusion. Optimality in the SIFI and fusion illusion is defined, therefore, by the complex relationship of auditory dominance in temporal tasks, the temporal positioning and proximity of the audiovisual inputs presented, and the visual acuity of the perceiver. Future studies on this relationship will allow its further refinement and the potential inclusion of other known multisensory phenomena (e.g., temporal ventriloquism).

5. References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257-262.

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research, 21*, 301-308.

Apthorp, D., Alais, D., & Boenke, L. (2013). Flash illusions induced by visual, auditory, and audiovisual stimuli. *Journal of Vision, 13*, 1-15.

Cuppini, C., Magosso, E., Bolognini, N., Vallar, G., & Ursino, M. (2014). A neurocomputational analysis of the sound-induced flash illusion. *NeuroImage, 92*, 248-266.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*, 162-169.

Fiedler, A., O’Sullivan, J. L., Schroter, H., Miller, J., & Ulrich, R. (2011). Illusory double flashes can speed up responses like physical ones: evidence from the sound-induced flash illusion. *Experimental Brain Research, 214*, 113-119.

Forster, B., Cavina-Pratesi, C., Aglioti, S. M., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research, 143*, 480-487.

Foss-Feig, J., Kwakye, L., Cascio, C., Burnette, C., Kadivar, H., Stone, W., & Wallace, M. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research, 203*, 381-389.

Keetels, M., & Vroomen, J. (2012). *Perception of synchrony between the senses*. In M. M. Murray and M. T. Wallace (Eds.), Frontiers in the neural basis of multisensory processes (pp. 147-177). London: Taylor and Francis Group.

Kerlin, J., & Shapiro, K. (2015). Multisensory integration: How sound alters sight. *Current Biology, 25*. R76-R77.

Kumpik, D., Roberts, H., King, A., & Bizley, J. (2014). Visual sensitivity is a stronger determinant of illusory processes than auditory cue parameters in the sound-induced flash illusion. *Journal of Vision, 14*, 1-14.

McDonald, J. J., Green, J. J., Störmer, V. S., & Hillyard, S. A. (2012). *Cross-modal spatial cueing of attention influences visual perception*. In M. M. Murray and M. T. Wallace (Eds). The neural bases of multisensory processes (Chapter 26). Boca Raton (FL): CRC Press.

McGovern, D., Roudaia, E., Stapleton, J., McGinnity, T., & Newell, F. (2014). The sound-induced flash illusion reveals dissociable age-related effects in multisensory integration*.* *Frontiers in Aging Neuroscience, 6*, 1-22.

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *The Journal of Neuroscience, 27*, 4120-4131.

Mishra, J., Martinez, A., & Hillyard, S. A. (2008). Cortical processes underlying sound-induced flash fusion. *Brain Research, 1242*, 102-115.

Mishra, J., Martinez, A., & Hillyard, S. A. (2013). Audition influences color processing in the sound-induced visual flash illusion. *Vision Research, 93*, 74-79.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. F. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17*, 154-163.

Neufeld, J., Sinke, C., Zedler, M., Emrich, H., & Szycik, G. (2012). Reduced audio-visual integration in synaesthetes indicated by the double-flash illusion. *Brain research, 1473*, 78-86.

Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H.-J., & Driver, J. (2010). Sound-induced enhancement of low-intensity vision: Multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *The Journal of Neuroscience, 30*, 13609-13623.

Pachella, R. G. (1974). *The interpretation of reaction time in information processing research*. In B. Kantowitz (Ed.), Human information processing (pp. 41-82). Potomac, Md: Lawrence Erlbaum Associates.

Perez-Bellido, A., Soto-Faraco, S., & Lopez-Moliner, J. (2013). Sound-driven enhancement of vision: Disentangling detection-level from decision-level contributions. *Journal of Neurophysiology, 109*, 1065-1077.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 1085-1099.

Roseboom, W., Kawabe, T., & Nishida, S. (2013). The cross-modal double flash illusion depends on featural similarity between cross-modal inducers. *Scientific Reports, 3*, 1-5.

Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-induced flash illusion is resistant to feedback training. *Brain Topography, 21*, 185-192.

Setti, A., & Chan, J. (2011). Familiarity of objects affects susceptibility to the sound-induced flash illusion. *Neuroscience Letters, 492*, 19-22.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature, 408,* 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14*, 147-152.

Shams, L., Mab, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport, 16*, 1923-1927.

Simon, J., & Csifcsák, G. (2015). *Early electrophysiological correlates of susceptibility to the double-flash illusion*. Poster presented at the Second Workshop and Lecture Series on Cognitive neuroscience of auditory and cross-modal perception, Košice, Slovakia.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA, US: The MIT Press.

Stein, B. E., Lagondon, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, *8*, 497-506.

Stevenson, R., Zemtsov, R., & Wallace, M. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology Human Perception and Performance, 38*, 1517-1529.

Takeshima, Y., & Gyoba, J. (2013). Complexity of visual stimuli affects visual illusion induced by sound. *Vision Research, 91*, 1-7.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*, 598-607.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the ‘unity assumption’ using complex audiovisual stimuli. *Perception & Psychophysics, 69*, 744-756.

Vatakis, A., & Spence, C. (2008). Evaluating the influence of the ‘unity assumption’ on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica, 127*, 12-23.

Vatakis, A. (2013). *The role of stimulus properties and cognitive processes in the quality of the multisensory perception of synchrony*. In L. Albertazzi (Ed.), The Wiley-Blackwell Handbook of Experimental Phenomenology. Visual Perception of Shape, Space and Appearance. Wiley-Blackwell.

Wada, Y., Kitagawa, N., & Noguchi, K. (2003). Audio-visual integration in temporal perception. *International Journal of Psychophysiology, 50*, 117-124.

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage, 31*, 1247-1256.

Watkins, S., Shams, L., Josephs, O., & Reesa, G. (2007). Activity in human V1 follows multisensory perception. *NeuroImage*, *37*, 572-578.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research, 152*, 198-210.