

RESEARCH ARTICLE

Crossmodal interference on counting performance: Evidence for shared attentional resources

Claudia Del Gatto^{1*}, Allegra Indraccolo¹, Tiziana Pedale^{2,3}, Riccardo Brunetti¹

1 Experimental and Applied Psychology Laboratory, Department of Human Sciences, Università Europea di Roma, Rome, Italy, **2** Department of Physiology and Pharmacology, Sapienza University of Rome, Rome, Italy, **3** Functional Neuroimaging Laboratory, Fondazione Santa Lucia, IRCCS, Rome, Italy

* claudia.delgatto@unier.it

Abstract

During the act of counting, our perceptual system may rely on information coming from different sensory channels. However, when the information coming from different sources is discordant, such as in the case of a de-synchronization between visual stimuli to be counted and irrelevant auditory stimuli, the performance in a sequential counting task might deteriorate. Such deterioration may originate from two different mechanisms, both linked to exogenous attention attracted by auditory stimuli. Indeed, exogenous auditory triggers may infiltrate our internal "counter", interfering with the counting process, resulting in an overcount; alternatively, the exogenous auditory triggers may disrupt the internal "counter" by deviating participants' attention from the visual stimuli, resulting in an undercount. We tested these hypotheses by asking participants to count visual discs sequentially appearing on the screen while listening to task-irrelevant sounds, in systematically varied conditions: visual stimuli could be synchronized or de-synchronized with sounds; they could feature regular or irregular pacing; and their speed presentation could be fast (approx. 3/sec), moderate (approx. 2/sec), or slow (approx. 1.5/sec). Our results support the second hypothesis since participants tend to undercount visual stimuli in all harder conditions (de-synchronized, irregular, fast sequences). We discuss these results in detail, adding novel elements to the study of crossmodal interference.

OPEN ACCESS

Citation: Del Gatto C, Indraccolo A, Pedale T, Brunetti R (2023) Crossmodal interference on counting performance: Evidence for shared attentional resources. PLoS ONE 18(11): e0294057. <https://doi.org/10.1371/journal.pone.0294057>

Editor: Sergei Tugin, Stanford University School of Medicine, UNITED STATES

Received: March 28, 2023

Accepted: October 16, 2023

Published: November 10, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0294057>

Copyright: © 2023 Del Gatto et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting Information](#) files.

Introduction

Event counting, such as figuring out how many drops of medicine we are putting in a glass of water or how many times cards of a specific suit appeared during a card game, is a common everyday activity used in many different situations. Logie and Baddeley [1] suggested that event counting engages three types of cognitive processes: firstly, we perceive the item to be counted; then, we retrieve a proper counting system from long-term memory; and, finally, the working memory support the process of maintaining the running total. More specifically, they proposed the phonological loop to be the working memory sub-system responsible for both

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

the process of subvocal articulation of numbers and the maintenance of the running total [2]. However, the process of counting can also rely specifically on attentional resources, especially when it comes to selective counting (e.g., telling apart target items to be counted from distractors; [3, 4]).

The first cognitive process involved in the process of counting, as stated by Logie and Baddeley [1] is the perception of the event or stimulus to be counted. There are many situations where we are simultaneously presented with information from multiple sensory modalities and crossmodal congruence or incongruence has a great impact on perception [5]. Many studies have demonstrated the existence of crossmodal interplay between auditory features and visual dimensions, such as spatial elevation [6–11], brightness [12, 13] or lightness [13–17], size [8, 17, 18], angularity of shape [13], direction of movement [19], synchrony, and temporal congruence [20].

Although a wide range of research has focused on crossmodal interactions across all five senses, at a perceptual, mnemonic, or motor level (see for example [21–25]), to the best of our knowledge, no studies have directly investigated crossmodal interactions in event counting.

As mentioned, the second system involved in the counting processes is memory [1]. In this case as well, the impact of crossmodality seems to be relevant: for example, it has been shown that irrelevant background speech disrupts serial short-term memory for materials presented visually [26]. Other studies have shown the impact of crossmodal effects on working memory performances. For example, a study by Brunetti et al. [27] used an audiovisual two-back task with different kinds of audiovisual correspondences in congruent and incongruent conditions, showing better memory performances in congruent conditions. Interestingly, it seems that when stimuli from different sensory modalities offer additional information, they can facilitate processing [28–32]. However, there are also situations where stimuli presented simultaneously in different modalities can impair performance: When sensory modalities offer conflicting information, stimuli presented in one modality can alter or attenuate processing in the other modality [33–37].

Sensory dominance

Much research of this kind, directed for the most part in the domain of attention, pointed out visual dominance. For instance, in the Colavita visual dominance task [33], participants were asked to press a button in response to a sound and another one in response to a light. Interestingly, participants often press only the visual button when both stimuli are presented together, suggesting that the visual modality dominates when responding to multisensory information [30, 38] (see [39] for a review). On the other hand, further research shows that when temporal judgment is required auditory processing can dominate over visual processing. Shams and colleagues [35] and also Burr and Bank [40] showed that the perception of visual stimuli can be distorted by the simultaneous presentation of sounds. In a classic experiment examining the flash illusion, participants see flashes of light while also hearing sounds presented in fast temporal sequences. When participants are presented with a single flash and two sounds, they tend to perceive more than one flash (fission illusion). Interestingly, although weaker, the opposite phenomenon was also observed, i.e., the so-called fusion illusion, where the combination of two visual stimuli and a single auditory stimulus might induce the perception of a single visual stimulus (see, for reviews, [41, 42]). In the first case, the final effect is an overcount of the visual stimuli, while in the second scenario the visual stimuli tend to be undercounted. Studies focusing on the neural mechanisms underlying the fission illusions (EEG, ERP, and fMRI studies) evidenced that this phenomenon is associated with increased gamma band power [43], increased evoked responses [44, 45], and increased bold activity in the primary

visual cortex, and in the superior temporal sulcus, a key region involved in multisensory integration [46, 47]. As for the primary visual cortex, it is interesting that when irrelevant auditory stimuli effectively produced a fission illusion, suggesting an “overcount mechanism”, the occipital visual cortex appeared to increase its activity [46, 47]. On the contrary, during the fusion illusion, suggesting an “undercount” effect, a reduced BOLD signal in the same occipital visual regions was observed [47]. All these findings support an auditory dominance during crossmodal processing, which appear to dramatically impact on the visual perception at behavioral and neural level. The same asymmetry was found in the context of the counting process. Abadi and Murphy [48], studied both fission and fusion illusions by examining counting performance with stimuli that could be spatially and temporally variable. They found sound-induced fusion to be more robust than sound induced fission and that sound-induced fusion is the most likely outcome when visual and auditory stimuli are incongruent (e.g. presented at different time intervals). In this case, therefore, the visual dominance is variable. These results have been interpreted by the authors with the concept of ‘auditory driving’, in which task-irrelevant auditory stimuli can influence the number of visual targets either by excess or by defect [49–52]. In line with this, Ljung and colleagues [53] also investigated the effects of auditory stimuli in a counting task. Specifically, the authors examined if the meaning of a background speech could affect event counting. They suggested that when background speech is similar in meaning to the target task process, it contributes to the disruption. The topic of auditory dominance is further debated, indeed, in a study by Ngo, Cadieux et al. [54]: while some of their results highlight auditory dominance, other ones reveal visual dominance. In the field of sensorimotor literature, research suggests that rhythmic motor behavior is indeed attracted more strongly to auditory than to visual sequences. When there is interference between auditory and visual stimuli, participants’ evaluations are usually more biased toward the auditory than the visual temporal information [35, 55–64].

A study by Dunifon et al. [65] reported a more complex pattern of modality dominance, finding evidence for both auditory and visual dominance. Results showed that crossmodal presentation deteriorated auditory discrimination but also visual reaction times: The authors interpret these outcomes suggesting that auditory stimuli automatically engage attention and attenuate visual processing.

Mechanisms of crossmodal interference

When stimuli are presented bimodally, we can find many examples of crossmodal interference: e.g., different task-irrelevant auditory stimuli seem to have detrimental effects on the immediate recall of visually presented materials, by interfering with working memory processes [26, 66–71]. In a study by Sinnett and Kingstone [72], participants had to evaluate the direction of a tennis ball thrown in two different conditions, without any sounds or paired with a grunting noise. Results showed that the grunting sound condition slowed reaction time and increased errors rates compared with the silent condition. The detrimental effect of task-irrelevant auditory stimuli was also corroborated by Robinson and Parker [73]: They showed that response times to visual stimuli were slower when the visual sequences were presented together with tones. Also, Laughery, Pesina and Robinson [74] examined response times and eye tracking variables while participants completed variations of unimodal and crossmodal serial response time tasks. They found that participants were slower to fixate on the visual stimuli in the crossmodal condition compared to the unimodal condition.

Another study found crossmodal interference particularly in the process of counting. Lo and Lay [75] used a counting task with the presence of either congruent or incongruent sensory signals in the background. Participants were asked to count the number of flashes of a

black circle or the presentation frequency of a sound. Conjointly background sensory cues were presented which could be congruent or incongruent with respect to the targets. Results demonstrated that participants generally performed better with congruent than with the incongruent stimuli.

On the other hand, other studies go in the opposite direction and demonstrate increased visual processing inducted by sounds [76, 77]. For example, McDonald et al. [78] provides psychophysical evidence that a task-irrelevant sound facilitates the detectability of a following flash appearing at the same position. Also, Stein et al. [79] asked participants to rate the intensity of a visual stimuli that could be accompanied by an auditory stimulus or presented alone. Results showed that the auditory stimulus improved the evaluation of visual stimuli.

In the cases seen so far, auditory stimuli appear to interact in complex ways with visual stimuli. In some cases, it seems that visual dominance emerges [33], in others instead it emerges auditory dominance [35, 40], in other cases a more complex pattern of dominance come up [65] and again, in the sensorimotor field it seems that when movement or its planning is involved, auditory dominates over visual stimuli [60–63]. These complex patterns of results are also dependent by the way in which different modalities may capture our attention.

Attention in multisensory processing

It is important to underline the role of attention in the counting process as well as its role in the processing of multisensory stimulation. Mainly, attention facilitates working memory as a buffer for both the encoding and retrieval of information necessary for performing cognitive tasks like counting [3]. Selective attention is in charge of managing new information during various phases of processing, moreover, can act as a supervisor for working memory by enhancing the encoding of the information more relevant to the specific task at hand.

Regarding the role of attention on multisensory stimuli, a review [80] highlighted the effect of top-down and bottom-up processes. Indeed, some studies have reported that multisensory bottom-up processes have the capability to capture our attention [81]. On the other hand, other studies [31, 82] focused on how other higher-level factors can impact on integration of information across the senses. Talsma et al. [80], propose that multisensory integration tends to be pre-attentive when the amount of competition between resources is low. In contrast, when stimuli between different modalities are competing for resources, and thus the saliency is low, top-down selective attention could be necessary [83, 84]. A corollary of this explanation is that auditory and visual modalities share the same pool of attentional resources and compete for them: a supramodal attention that can be allocated to whatever sensory modality in bottom-up or top-down way (see [85–87]).

The supramodal approach relies on executive functions: The system dedicated to managing conflicting multisensory messages coming from irrelevant and relevant sources [88–92]. This system has been shown to be supramodal by integrating stimuli coming from different modalities [91–94].

In contrast to a supramodal nature point of view, each sensory modality could have its own separate pool of attentional resources (the modal view: [95]). Therefore, separate tasks performed in two different sensory modalities would not interfere. Evidence from behavioral studies support this assumption [96]. Keitel et al. [97], for example, investigated the consequences of a shift from simultaneous bimodal (visual and auditory) to unimodal selective attention. Their results suggest a modality-specific resources account. Given the conflicting experimental evidence, the debate on modality specific or supramodal nature of attentional resources is still open. However, in the specific context of the counting process it is possible to hypothesize that a type of supramodal attention intervenes since it has already been

demonstrated that there are crossmodal interferences in tasks involving counting: for example when visual target stimuli are presented simultaneously with incongruent auditory stimuli [75].

Factors affecting selective counting

The present study investigates how the visual and auditory modalities interact in a longer, continuous task involving attentional and ideomotor simulations aspects. Specifically, a crossmodal counting task, although managed by working memory, undergoes attentional interference. Indeed, crossmodal interactions between visual and auditory stimuli are possibly still effective in a continuous process, such as counting. Verbal counting is a well-learned ability; however, it can be difficult and error-prone, especially for large magnitudes: with an increasing number of objects to be counted it is easy to lose track of the running total.

An interesting result with a counting task in the crossmodal domain has been obtained by Carlyon and colleagues [98]. In this investigation, participants heard a sequence of tones and saw a sequence of visual stimuli; their task was to count, according to the request, either the number of visual or the number of auditory stimuli. The results showed a better performance when participants counted the auditory targets than in all the other conditions, confirming that the auditory dominance found in other temporal tasks applies to counting as well.

The counting process seems to be affected also by unimodal variables, for example, the so-called fluency, which is the timing regularity of the events to be counted. Carlson and Cassenti [99] examined specifically these effects: they asked participants to count events occurring at regular or irregular intervals and then to report their total count. In general, results showed that participants counted regular events more accurately. Stevenson and Carlson [100] investigated fluency as well: They asked participants to count asterisks as they appeared on the screen with regular (isochronous) or irregular Stimulus Onset Asynchronies (SOAs). Then, participants were asked to provide a final count and a confidence rating. Their results showed that participants were more accurate and produced greater confidence ratings with regular timings.

The superiority of regular sequences can be explained by the fact that temporal regularity between events can also greatly improve our perception as highlighted by the extensive work of Kia Nobre and Jennifer Coull [101]. Moreover, the Dynamic Attending Theory (DAT; [102–104]) suggests that attention is not allocated regularly in time but fluctuates periodically according to internal dynamic oscillators. These oscillators regulate an individual's attending rhythm and thus the creation of expectation over time. Accordingly, a study by Escoffier et al. [105], compared behavioral responses to visual stimuli presented in and out of synchrony with a musical rhythm and showed faster reaction time for picture presented with the musical rhythm and on-beat. Furthermore, it has been shown that crossmodal entrainment in rhythmic multisensory stimuli can impact behavioral outcomes of perception. In a recent study Lin, Yeh, and Shams [106] investigated the effect of congruence between visual and auditory stimuli on perceptual pleasure (positive emotional response to sensory stimuli), in the sense of the rhythmic temporal synchronization between auditory and visual stimuli. The authors found higher perceptual pleasure for temporal congruency compared with incongruency. These results indicate that rhythm entrains attention, even at a subliminal level [106] facilitating information processing in synchrony with the rhythm.

Lastly, another variable that may potentially affect counting is speed. The counting process may be considered as an ideomotor behavior. The ideomotor approach [107, 108], is influenced by the idea of a strong link between perception and action [109] (for crossmodal effects see also, [23, 110]). The idea is that the representation of actions and representation of stimuli

share the same format. However, some studies proved that the ideomotor concept could be much more than a motor mechanism. Indeed, Badets et al. ([111], Experiment 1) highlighted the role of an ideomotor mechanism in number processing without any movement [112]. Thanks to these studies it is possible to hypothesize that the activity of counting silently (covert counting), can be connected to an ideomotor simulation in which a process of subvocal numerical representation is activated, thanks to the articulatory process of the phonological loop [99]. In other words, even if it is not required to count aloud, counting may have limitations related to the internal articulation of numbers. If this is the case, we would expect to find significant effects connected to speed, as the articulatory process, even subvocally, is constrained to articulation limits (see for example [113, 114], regarding articulatory limitations related to the time required to pronounce some words in different languages). Moreover, considering the ideomotor account with respect to subvocalic repetition during a counting process, it is possible to hypothesize that when the items to be counted appear very quickly and their magnitude increases, performance may deteriorate, also because participants has not yet finished subvocalizing the current count that already appears the following stimulus to be counted.

Finally, counting processes can be affected by our preferred tempo. Indeed, “Spontaneous motor tempo”, namely the tempo of self-paced regular and repeated movements, that corresponds to the preferred and natural rhythm to execute isochronous motor actions (e.g., walking, or tapping a finger), which usually tends to cluster around 500–600 msec in adults [115–117] might modulate our performance in counting. To the extent of our knowledge, to date no one has explicitly investigated these aspects in counting, thus we find it important to include the variable of speed to study its effects in association with other variables that can influence the counting process. For example, the interplay between speed and fluency could yield interesting results. Since regularity is an important factor in a counting task [99, 100] it would be interesting if this could counterbalance the difficulty that might come with speed.

Counting can thus be a task prone to errors due to different crossmodal or unimodal features [74] as well as temporal variables such as regularity [99]. In this study, we focus on the crossmodal effect of simultaneously presented visual and auditory information on covert event counting (that is—to count internally without verbalizing the count). Our interest is to verify if task-irrelevant auditory information can attract exogenously our attention and disrupt the counting process. This is in line with the idea that crossmodal signals can affect sensory processing by directing attention [35].

In general, we also expect that the variables Speed (slow or fast presentation of stimuli to be counted) and Isochronicity (SOAs regularity or irregularity, or rhythmicity, see Shams et al. [35]) can affect participants’ performance. Specifically, the faster the visual events are presented, the more likely it will be to have errors in counting. Based on previous literature, slow presentation implies events (visual and/or auditory stimuli) that happen up to 95 bpm (e.g. 1.5 events every second); moderate presentation implies events that happen from 95 to 150 bpm (in line with the “Spontaneous motor tempo”; e.g. 2 events every second); instead, fast presentation implies events that happen above 150 bpm (e.g. 3 events every second [118, 119]).

In addition, SOAs regularity (isochronicity) should result in a better performance compared to irregular SOAs [99]. Finally, we expect better performance when visual and auditory stimuli are presented synchronously compared to when the presentation is unsynchronized [120].

To summarize, we have considered three different variables that could affect counting processes: 1. Speed—we hypothesize that increasing presentation speed will result generally in worse performance; 2. Isochronicity (regular or irregular time intervals between target stimuli)—we hypothesize that counting sequences featuring regular timing will be easier than counting events with irregular intervals [90, 91] 3. Synchrony—we hypothesize that the presence of

sounds in synchrony with visual stimuli to be counted could facilitate the performance; on the other hand, a condition where sounds are asynchronous with visual stimuli could result in an overcount or an undercount (e.g. [5]).

More specifically, regarding this last variable, we want to test two different hypotheses regarding the possible disruptive effects of Synchrony or the lack thereof. These two hypotheses are linked to the ability of sound to exogenously capture our attention, even if task-irrelevant [121–123]. The first hypothesis is that covert counting of visual stimuli can be distorted by the simultaneous presentation of asynchronous sounds that, by exogenously attracting our attention, can trigger the counting process and infiltrate the current count: This infiltration, where irrelevant asynchronous sounds are able to trigger our counting process more than necessary, would result in an overcount. We hypothesize that this could happen even if the amount of auditory and visual stimuli is the same since the variation in numerosity is given illusorily by asynchrony, in other words the fact that the stimuli are not perceptually aligned could elicit the perception of a different numerosity, just as in Shams' illusion [35]. The second hypothesis, on the other hand, postulates that the task-irrelevant sounds presented asynchronously with the visual stimuli will disrupt the counting process (e.g., [53, 124]): This disruption, by temporarily distracting the counting process, would result in missing some of the visual events and eventually in an undercount. Hence, the results can help us to glean a more complete picture on how asynchronous sounds can affect a visual counting task: if the participants tend to overcount, we can assume the first hypothesis (infiltration) to be true, while if they undercount, the second one (disruption) will be confirmed. To test these hypotheses, we conducted two experiments: We will now illustrate them and then interpret the results of both in the General Discussion.

Experiment 1

Method

All methods were carried out in accordance with declaration of Helsinki. The experimental protocol has been approved by the Ethics Review Board of the Università Europea di Roma.

Participants. Seventy-nine participants (65 females, mean age = 20.71 years; SD = 1.49 years; range = 19–25) took part in the experiment for course credit. The administrations were carried out between 20 January and 28 February 2022. All the participants reported normal or corrected-to-normal vision and audition. The participants were naïve as to the purpose of the study and gave informed written consent. Authors had access to information (Name and Surname, university registration number, gender, hours of sleep in the night before doing the experiment, age) that could identify individual participants during and after data collection.

Materials. Thirty-six video or audio-video sequences were created for the experiment. All sequences were generated at 12 fps (each frame lasted for approx. 84 msec) and featured a flashing white disc (visual angle: approx. 9° on a grey (50%) background; the measurement of the visual angle must be considered an approximation because the experiment was administered online, therefore allowing for different screen sizes and distances from the screen). Since the experiment was administered online, even though participant was instructed to sit approximately 60 cm from the screen it was not possible to control this variable accurately. However, the size of the visual stimulus was kept constant during the experiment, and it was irrelevant to the task. Even if the participants varied their distance from the screen to perform the task, we can assume this variation was random, thus not systematically affecting the results.

Each disc flash lasted for 2 frames (168 msec) in all sequences. The number of flashes in each sequence varied pseudo-randomly (average: 31, min: 28, max: 34). In half of the sequences, the flashes were isochronous (regular time interval between stimuli), while in the

other half they were irregular. Half of the sequences were "Moderate" (2 flashes per second on average) while the other half was "Fast" (3 flashes per second on average) and varied in length according to the total number of flashes (average length, in seconds, of "Moderate" sequences: 15.3, min: 13.6, max: 17.4; "Fast" sequences, average: 10.6, min: 9.2, max: 12.8). Sequences varied also according to sound presence and type: "No sound" sequences were only videos with no sounds; "Synchronous" sequences had sounds synchronized with the onset of disc flashes; "Asynchronous" sequences had the same amount of sound and flashes, but they were randomly scattered in the sequence, without any synchronization (i.e., no onset of the discs had a concurring sound onset). The sound used to create the sequences was a clearly audible "tick" lasting for 30 msec. The systematic crossing of these variables resulted in 12 different conditions (see Fig 1): 1. White disc flashing isochronously with no sound, moderate (2/s); 2. White disc flashing isochronously with no sound, fast (3/s); 3. White disc flashing isochronously with synchronous "tick" sounds, moderate (2/s); 4. White disc flashing isochronously with synchronous "tick" sounds, fast (3/s); 5. White disc flashing isochronously with asynchronous "tick" sounds, moderate (2/s); 6. White disc flashing isochronously with asynchronous "tick" sounds, fast (3/s); 7. White disc flashing non-isochronously with no sound, moderate (2/s); 8. White disc flashing non-isochronously with no sound, fast (3/s); 9. White disc flashing non-isochronously with synchronous "tick" sounds, moderate (2/s); 10. White disc flashing

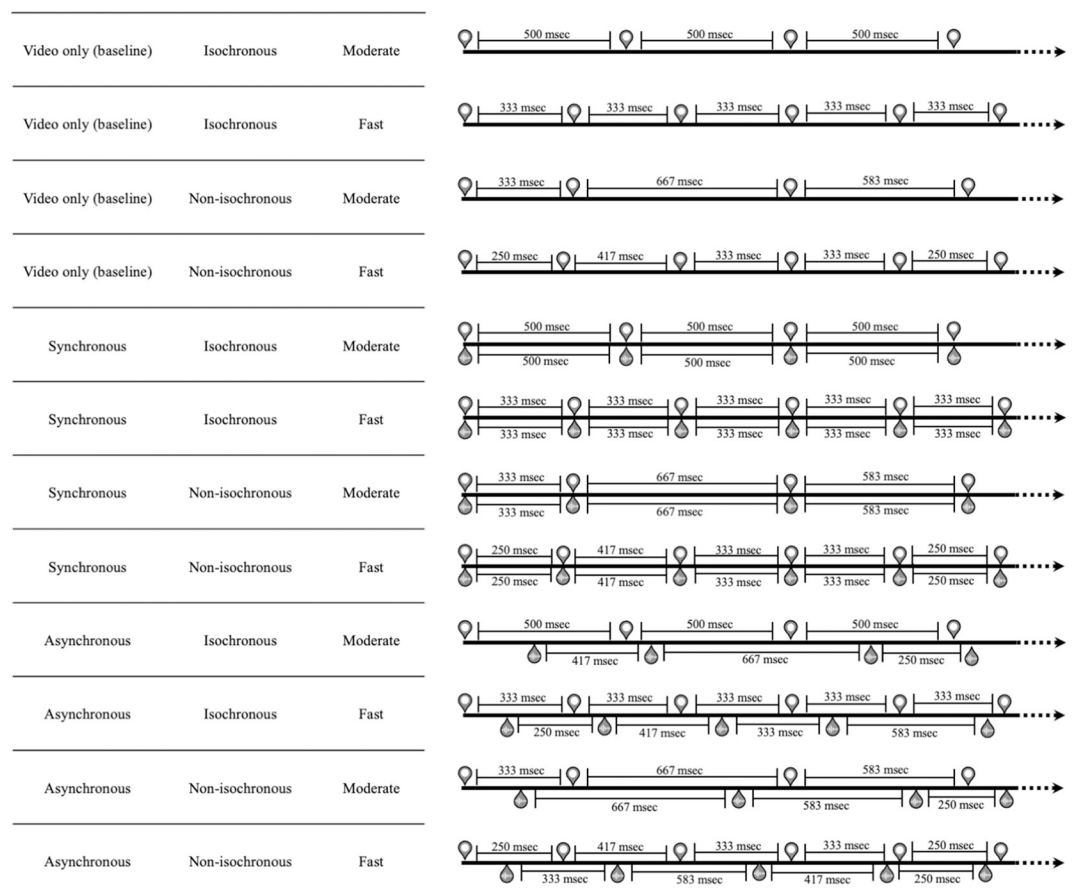


Fig 1. Examples of sequence types used in Experiment 1. All icons above the timelines represent visual stimuli (white discs on grey background), while all icons below the timelines represent "tick" sounds. The timings of all stimuli fit in the frames of the 12 fps structure (each frame is approximately 83 msec) used to create the video sequences.

<https://doi.org/10.1371/journal.pone.0294057.g001>

non-isochronously with synchronous “tick” sounds, fast (3/s); 11. White disc flashing non-isochronously with asynchronous “tick” sounds, moderate (2/s); 12. White disc flashing non-isochronously with asynchronous “tick” sounds, fast (3/s). In the bimodal conditions, the number of visual and auditory stimuli was equal. Each condition had 3 instances, for a total of 36 trials. The 36 trials containing the 12 different conditions were presented in a random order. Every sequence was followed, after a random interval of approx. 2 sec, by 1 out of 4 “control” sounds: gunshot, dog barking, cat meowing, glass shattering. The control sounds had a length varying from 250 msec to 600 msec. The volume of the control sounds was adjusted according to the volume of the auditory stimuli: specifically it was adjusted to a lower volume. In other words, the control sound was always at a lower volume in order to make it audible but to avoid increasing arousal levels.

Procedure. Each participant was presented with all the sequences. All sequences ended, a few seconds after the last flashing disc disappeared, with a control sound. The participants were invited to count the flashing white discs that appear in the sequences and to indicate, at the end of the presentation, their count, by clicking on an on-screen scale showing numbers ranging from 15 to 40 and to indicate what kind of sound they heard at the end. The task was the same throughout the whole experiment, which lasted approximately 20 min. After receiving instructions, the participants performed 5 training trials to get used to the task. This training involved both bimodal and unimodal (visual only) stimuli.

Since the experiment was administered online, several measures were taken to make sure they performed the experiment correctly: 1) The recognition of the control sound was included to make sure they had the volume turned on and at a hearable level while performing the task; 2) The interpolation of sequences without sounds was used both to record a unimodal baseline, but also to make sure that the participants would not risk to count sounds instead of flashing discs. No feedback about the correctness of the response was given.

Apparatus. Stimulus presentation, conditions, randomization, and the recording of responses were all controlled by a custom-made script in the PsychoPy (v 1.80; [125]) programming environment, with online administration using Pavlovia.org [125].

Results

To perform the analyses only those participants that clearly heard the sound for the entire experiment we included only the participants with accuracy at the control question above 90%: this led to the exclusion of two participants. Thus, the analysis was performed on 77 participants (10 males; mean age = 20.72 years; SD = 1.51; range 18–25 years). The average accuracy to the control question on this subsample of subjects was 99% (SD = 0.02%) indicating that the participants clearly heard the sounds for the entire duration of the experiment.

The next analyses were performed on Delta values, which were calculated by subtracting the correct count (CC) from the count given by the participant (PC; $PC - CC = \text{Delta value}$). In this way, we obtained a value that indicates the deviation of each participant’s count from the correct count: If the value is negative, it means that the participant underestimated the number of stimuli; if the value is positive, it means that the participant overestimated the count. We ran a three-way repeated-measure Bayesian analyses of variance (ANOVAs; [126]) with Isochrony (2 levels), Speed (2 levels) and Synchrony (3 levels) as independent within-subject factors, using Delta data as a dependent variable. The advantage of the Bayesian approach is that, while classical frequentist ANOVA allows only to accept or reject the presence of an effect, Bayesian ANOVA permits to recognize the best model describing the dependent variable by means of model comparisons (BF_M) and allows to estimate the strength of evidence in favor (BF_{10}) of each factor involved in the model according to the following interpretation: $(BF) < 1$:

Table 1. Analysis of the effects in the Bayesian ANOVA conducted on the Delta data in Experiment 1.

Effects	P(incl)	P(excl)	P(incl data)	P(excl data)	BF ₁₀
Isochrony	0.263	0.263	0.005	0.057	0.097
Synchrony	0.263	0.263	0.706	3.24x10 ¹⁴	2.18 x10 ¹³
Speed	0.263	0.263	0.052	2.53x10 ³⁷	2.041x10 ³⁵
Isochrony * Synchrony	0.263	0.263	0.117	0.803	0.146
Isochrony * Speed	0.263	0.263	0.914	0.006	149.184
Synchrony * Speed	0.263	0.263	0.171	0.805	0.213
Isochrony * Synchrony * Speed	0.053	0.053	0.023	0.017	1.328

P(incl) and P(excl) represent the priors of including and excluding the predictor before seeing the model. P(incl|data) and P(excl|data) represent the posteriors of including and excluding the predictor after seeing the model. BF₁₀ represents how much more likely the data are under the model including the predictor than under the model excluding the predictor.

<https://doi.org/10.1371/journal.pone.0294057.t001>

no evidence, 1–3: anecdotal evidence, 3–10: substantial evidence, 10–30: strong evidence, 30–100: very strong evidence, > 100: decisive evidence [127, 128]. The strength of evidence supporting the inclusion of each factor was evaluated by a comparison between the model that contains the effect of interest with the matched model deprived of the effect. For post hoc tests, correction for multiple comparisons was performed using the approach discussed in Westfall [129], which consists of the following steps (see also [130]): As a first step, Bayesian t-tests are computed for all pairwise comparisons, thus providing unadjusted Bayes factors; as a second step, the prior model odds are adjusted by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons; finally, the adjusted prior odds and the Bayes factor are used to calculate the adjusted posterior odds. All the statistical analyses were performed using JASP (Version 0.16.4, <https://jasp-stats.org/>).

Bayesian ANOVA revealed as best model, explaining the Delta data, the model containing the effect of Speed, Synchrony, and Isochrony, and the interaction between Isochrony and Speed (BF_M = 34.12, see also Table 1 and Fig 2 with all the experimental conditions). The evidence in favor of the inclusion of the main effect of Speed and Synchrony was decisive (all BF₁₀ > 100). On the contrary, there was no evidence for the main effect of Isochrony (BF₁₀ =

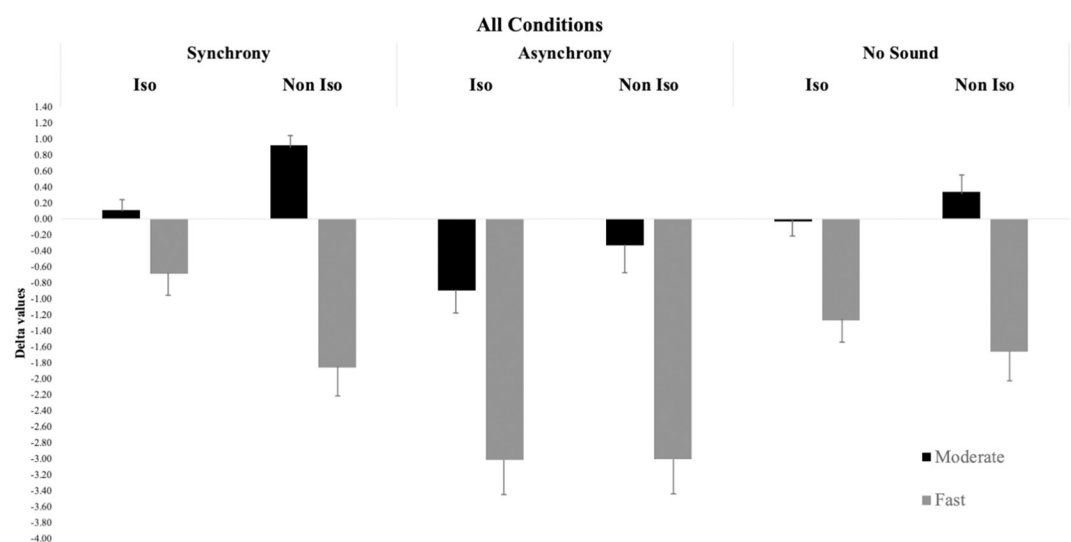


Fig 2. Bar graph displaying Delta data of all conditions of Experiment 1. The error bars represent the standard errors.

<https://doi.org/10.1371/journal.pone.0294057.g002>

0.10), but there was decisive evidence in favor of the inclusion of the interaction between Isochrony and Speed ($BF_{10} > 100$). As expected, there was no evidence in favor of the inclusion of all the other interactions (“Speed*Synchro”, $BF_{10} = 0.21$; “Isochrony*Synchro”, $BF_{10} = 0.15$; and the three-way interaction “Speed*Synchro*Isochrony” $BF_{10} = 1.33$).

Post-hoc analysis of the Speed effect revealed that participants were more accurate when the presentation of stimuli was at a moderate speed (Delta value: -0.08) as compared to the faster condition (Delta value: -2.11), ($BF_{10} > 100$, posterior odds = 1.79×10^{34}). When the presentation of stimuli was faster participants undercounted the stimuli, since the errors, in this condition, are all underestimation of the correct numerosity compared to the Moderate condition, where the Delta value is closer to zero. Moreover, post-hoc analysis of Synchro revealed the worst performance in the Asynchronous condition (Delta value: -1.98) compared to Synchronous (Delta value: -0.52) ($BF_{10} > 100$, posterior odds = 2.76×10^{11}) and to No sound (Delta value: -0.79) conditions ($BF_{10} > 100$, posterior odds = 1.54×10^9). There was no evidence for differences in the Synchronous and No Sound conditions ($BF_{10} = 0.65$, posterior odds = 0.38).

The post hoc analyses on the effect of Synchro show that participants were more accurate when auditory and visual stimuli were synchronized. The asynchronous presentation of auditory and visual stimuli makes participants lose more counts, so it appears that the auditory stimulus acts as an exogenous signal interfering with the counting process but not operating as a trigger and activating the next count.

As regarding the interaction “Isochrony * Speed” (see Fig 3), the post-hoc analysis revealed decisive evidence for a better performance in the Isochronous–Moderate condition (Delta value: -0.31) as compared to the Isochronous–Fast condition (Delta value: -1.77) ($BF_{10} > 100$, posterior odds = 235038.78); and a better performance in the Non-isochronous–Moderate condition (Delta value: 0.15) as compared to the Non-isochronous–Fast condition (Delta value: -2.46) ($BF_{10} > 100$, posterior odds = 7.44×10^{12}). Such effects showed in the first place that participants most underestimated the count when the stimuli presentation was faster than when it was slower (moderate) both in the case of Isochronous and Non-isochronous conditions, although the interaction revealed that such effect was greater in the Non-isochronous than the Isochronous conditions (see Fig 3). Secondly, with these two comparisons we can see different patterns of results. Indeed, when comparing Delta data of Moderate (-0.31) and Fast

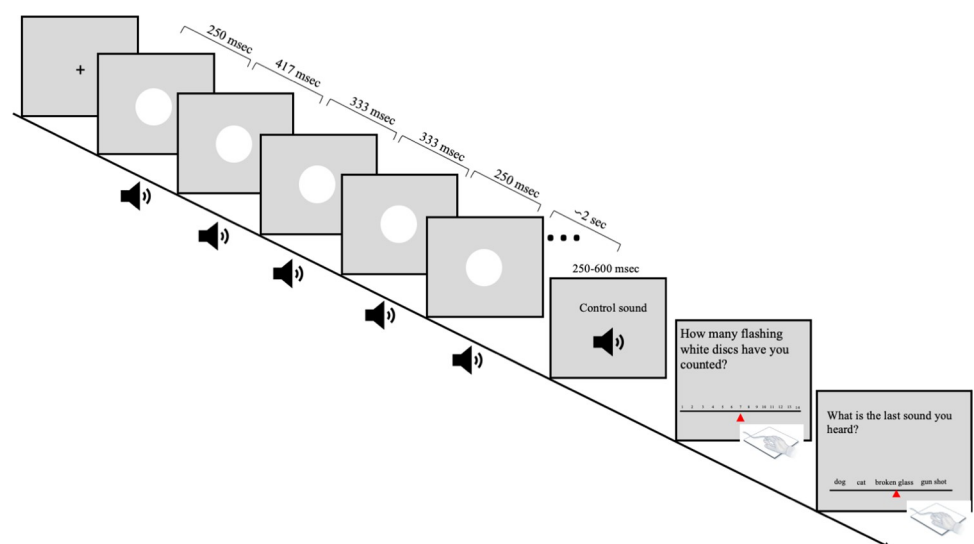


Fig 3. Graphical representation of one trial in the fast, non-isochronous, synchronous condition.

<https://doi.org/10.1371/journal.pone.0294057.g003>

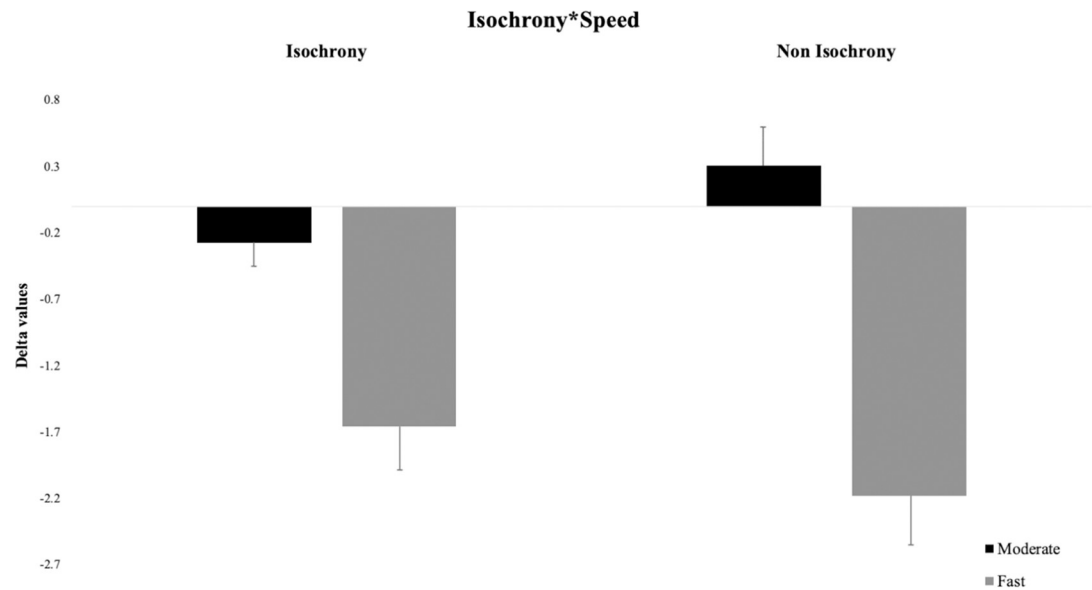


Fig 4. Bar graph displaying Delta data as a function of isochrony and speed in Experiment 1. The error bars represent the standard error.

<https://doi.org/10.1371/journal.pone.0294057.g004>

(-1.77) Isochronous conditions, it emerges that in each case these are underestimation errors. On the other hand, when comparing Delta data of Moderate (0.15) and Fast (-2.46) Non-isochronous conditions it emerges that when the presentation is moderate, the errors move towards overestimation, on the contrary, when the presentation is fast there are larger underestimation errors. Moreover, the interaction “Isochrony*Speed” revealed an increase of underestimation of faster stimuli when their presentation was not isochronous (Delta value: -2.46) compared with when it was isochronous (Delta value: -1.77), and an opposite pattern of results (increase of overestimation) with slower stimuli when their presentation was not isochronous (Delta value: 0.15) compared with when it was isochronous (Delta value: -0.31). In other words, the aspect that changes in the Fast vs. the Moderate conditions is the direction of the results. Indeed, in the Moderate Isochronous condition the errors made by the participants go in the direction of an underestimation while in the Non-isochronous Moderate condition the errors move more towards an overestimation. In summary, participants make more underestimation errors when they count stimuli presented slowly with regular intervals, and, on the other hand, they make more overestimation errors when stimuli are presented slowly with irregular intervals (see Fig 4). Therefore, in general, it seems that speed always produces underestimation, but when instead non-isochrony is analyzed specifically, speed and isochronicity have a different way of interacting: When intervals of presentation are regular, participants lose counts, while when they are irregular, participants seem to count more items than those presented.

Experiment 2

The results of Experiment 1 show that speed is a relevant factor in the counting process, leading participants to generate more undercount errors. Since there seems to be a linear effect of speed, we conducted a second experiment to further explore this hypothesis. In the Experiment 2 we added the "Slow" condition, in which there were 1.5 flashes per second on average.

Method

Participants. Twenty-six participants (5 males, mean age = 22.07 years; SD = 3.5 years; range = 19–30) took part in the experiment for course credit. All the participants reported normal or corrected-to-normal vision and audition. The participants were naïve as to the purpose of the study and gave informed written consent. Authors had access to information (Name and Surname, university registration number, gender, hours of sleep in the night before doing the experiment, age) that could identify individual participants during and after data collection.

Materials. Thirty-six video or audio-video sequences were created for the experiment. All sequences had the same fps rate, background, visual stimuli (flashing white discs), and auditory stimuli than the ones used in Exp. 1. Eighteen sequences were the same "Moderate" sequences used in Exp. 1, while the remaining 18 were slower ("Slow"). The number of flashes in each sequence varied pseudo-randomly (average: 31, min: 26, max: 34). In half the sequences the flashes were isochronous (with regular time interval), while in the other half they were irregular. In this case, half of the sequences was "Moderate" (2 flashes per second on average) while the other half was "Slow" (1.5 flashes per second on average) and varied in length according to the total number of flashes (average length, in seconds, of both "Moderate" and "Slow" sequences: 15.3, min: 13.6, max: 17.4). The sequences varied also according to sound presence and type, as in Exp. 1: "No sound", "Synchronous", and "Asynchronous" sequences had the same features described in Exp. 1. Similarly to Exp. 1, stimuli created 12 different conditions: 1. White disc flashing isochronously with no sound, moderate (2/s); 2. White disc flashing isochronously with no sound, slow (1.5/s); 3. White disc flashing isochronously with synchronous "tick" sounds, moderate (2/s); 4. White disc flashing isochronously with synchronous "tick" sounds, slow (1.5/s); 5. White disc flashing isochronously with asynchronous "tick" sounds, moderate (2/s); 6. White disc flashing isochronously with asynchronous "tick" sounds, slow (1.5/s); 7. White disc flashing non-isochronously with no sound, moderate (2/s); 8. White disc flashing non-isochronously with no sound, slow (1.5/s); 9. White disc flashing non-isochronously with synchronous "tick" sounds, moderate (2/s); 10. White disc flashing non-isochronously with synchronous "tick" sounds, slow (1.5/s); 11. White disc flashing non-isochronously with asynchronous "tick" sounds, moderate (2/s); 12. White disc flashing non-isochronously with asynchronous "tick" sounds, slow (1.5/s).

Each condition had 3 instances, for a total of 36 trials. The 36 trials containing the 12 different conditions were presented in a random order. Every sequence was followed, after a random interval of approx. 2 sec, by 1 out of 4 "control" sounds: gunshot, dog barking, cat meowing, glass shattering. The control sounds had a length varying from 250 msec to 600 msec and their volume was set to perceptually match that of the "tick" sound.

Procedure. Same procedure used in Exp. 1.

Apparatus. Same apparatus used in Exp. 1.

Results

As for Experiment 1, we include in the analyses only participants with accuracy at the control task above 90%, leading to the exclusion of three participants. Thus, analysis was performed on 23 participants (5 males; mean age = 22.3 years; SD = 3.7; range 19–30 years). The average accuracy of the control question on this subsample of subjects was 96% (SD = 0.02%) indicating that the participants clearly heard the sounds for the entire duration of the experiment.

The next analyses were performed on Delta values as in Experiment 1 (PC-CC). We ran a three-way repeated-measure Bayesian analyses of variance (ANOVAs; [126]) with Isochrony (2 levels), Speed (2 levels) and Synchrony (3 levels) as independent within-subject factors,

Table 2. Analysis of the effects in the Bayesian ANOVA conducted on the Delta data in Experiment 2.

Effects	P(incl)	P(excl)	P(incl data)	P(excl data)	BF ₁₀
Isochrony	0.263	0.263	0.192	0.737	0.260
Synchrony	0.263	0.263	0.652	0.04	173.21
Speed	0.263	0.263	0.636	0.001	428.31
Isochrony * Synchrony	0.263	0.263	0.025	0.237	0.104
Isochrony * Speed	0.263	0.263	0.051	0.212	0.239
Synchrony * Speed	0.263	0.263	0.329	0.666	0.494
Isochrony * Synchrony * Speed	0.053	0.053	2.28x10 ⁻⁴	0.001	0.166

P(incl) and P(excl) represent the priors of including and excluding the predictor before seeing the model. P(incl|data) and P(excl|data) represent the posteriors of including and excluding the predictor after seeing the model. BF₁₀ represents how much more likely the data are under the model including the predictor than under the model excluding the predictor.

<https://doi.org/10.1371/journal.pone.0294057.t002>

using Delta data as a dependent variable. As in the previous experiment, correction for multiple comparisons in the post hoc test was performed using the approach discussed in Westfall [129]. All the statistical analyses were performed using JASP (Version 0.16.4, <https://jasp-stats.org/>).

Bayesian ANOVA revealed as the best model, the model containing the effect of Speed, and Synchrony (BF_M = 17.7, see also Table 2, and Fig 5). The evidence in favor of the inclusion of the main effect of Speed and Synchrony was decisive (all BF₁₀ > 100). On the contrary, there was no evidence for the main effect of Isochrony (BF₁₀ = 0.260). Moreover, there was no evidence in favor of the inclusion of all the other interactions (“Isochrony*Speed”, BF₁₀ = 0.239; “Speed*Synchrony”, BF₁₀ = 0.494; “Isochrony*Synchrony”, BF₁₀ = 0.104; and the three-way interaction “Speed*Synchrony*Isochrony” BF₁₀ = 0.166).

Post-hoc analysis of the Speed effect revealed that participants were more accurate when the presentation of stimuli was slow (Delta value: 0.02) as compared to the Moderate condition (Delta value: -0.85), (BF₁₀ > 100, posterior odds = 879.30). When the presentation of stimuli

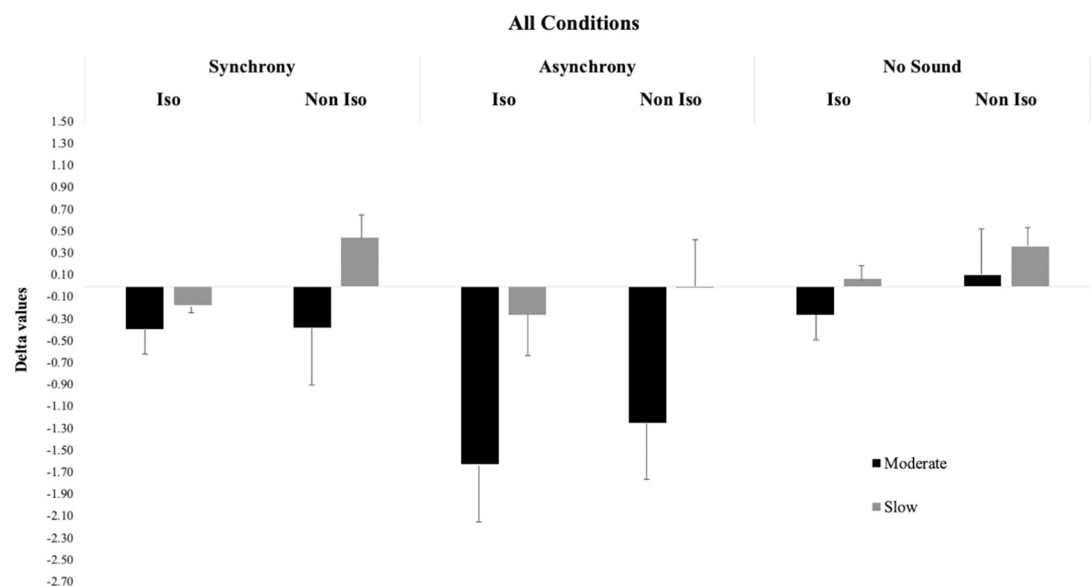


Fig 5. Bar graph displaying Delta data of all conditions of Experiment 2. The error bars represent the standard errors.

<https://doi.org/10.1371/journal.pone.0294057.g005>

was moderate participants undercounted the stimuli, as compared to the Slow condition, where the Delta value is closer to zero. Moreover, post-hoc analysis of Synchrony revealed the worst performance in the Asynchronous condition (Delta value: -1.04) compared to Synchronous (Delta value: -0.25) ($BF_{10} = 9.02$, posterior odds = 5.30) and to No sound (Delta value: 0.04) conditions ($BF_{10} > 100$, posterior odds = 104.04). There was no evidence for differences in the Synchrony and No Sound conditions ($BF_{10} = 0.375$, posterior odds = 0.220). The post hoc analyses on the effect of Synchrony show that participants were more accurate when auditory and visual stimuli were synchronized. Thus, the asynchronous presentation of auditory and visual stimuli makes participants lose more counts, confirming in this second Experiment as well that auditory stimuli interfered with the counting process without triggering additional counts.

General discussion

In this study, across two experiments, we tested two different hypotheses regarding the possible detrimental effect of a concurrent auditory stimulus on a visual event counting task. Our results support the hypothesis that auditory stimuli break into the counting process, disrupting it, as indicated in literature (e.g. [53]) acting as an exogenous stimulation that attracts the attention of the participants. With our results it is therefore possible to confirm previous studies in which when temporal judgment is required auditory processing dominates over visual processing [35, 40–48]. Moreover, our results highlight a further case of crossmodal interference between visual and auditory domains in a specific task such as counting [26, 66–74].

The interesting result is precisely represented by the direction that the results have, and not by a generic inaccuracy of the participants. In fact, according to our hypothesis, the positive or negative direction can offer insight into the precise disruption that auditory stimuli have on visual ones [26, 66–74]. The disruption we found generally resulted in an underestimation of numerosity, namely, participants seem to lose their count instead of counting more items. Based on these results, therefore, the hypothesis in which it was postulated that auditory stimuli could act as a trigger by entering the counting process and making participants count more stimuli, as in the illusions described by Shams and colleagues [35], does not seem to be validated in this case.

Our results show that speed is a relevant factor in the counting process. We manipulated the speed of presentation of the stimuli by creating different conditions, in Experiment 1 we compared a Moderate condition and a Fast condition while in Experiment 2 we graduated the speed of presentation by comparing the Moderate condition (same as Experiment 1) with a new Slow condition. Results in general show that as speed increases the participants tend to deviate more from the correct count in a negative way, generating more errors in defect. This is also interesting in relation to the results of Experiment 2: the Slow condition compared with the Moderate condition, suggests a linear trend of the speed factor. This can be explained by the fact that when the presentation speed increases, the next stimulus appears (therefore must be counted) before the processing of the previous stimulus is completed (or the count updated or consolidated), resulting in the damage of the counting process itself (similarly to what happens in the phenomenon of the Psychological Refractory Period, see [131]). This outcome can be explained by a counting process based on an internal motor simulation (e.g., articulating the numbers), affected by speed. Thus participants, to complete the task, covertly articulate the current count thanks to the phonological loop and when the presentation of stimuli to be counted is fast the ideomotor articulation of the current count is cut short by the appearance of the next target to be counted (as happens with subvocal articulation of words in different languages, see [113, 114]). This mechanism seems to determine several lost counts and a general tendency to underestimation.

Results of both Experiments did not reveal a facilitatory effect of synchronous stimuli presentation compared to the absence of sound, thus not replicating a crossmodal facilitation effect found in other crossmodal interactions. Diederich and Colonius [132] provided evidence of faster response times (RTs) to monochromatic lights when they were accompanied by simple tones presented simultaneously, while Sinnott et al. [32] found that participants were faster to detect visual targets when they were presented together with sounds. Moreover, in literature it is possible to find enhancement of the salience of visual stimuli by temporally congruent auditory stimuli [79, 133, 134]. However, the supposed crossmodal enhancement of the synchronous condition compared to the soundless condition seems not applicable to a continuous, longer task. It is possible to hypothesize that our task demand is not difficult enough to highlight the presence of a facilitation effect.

In the extant literature (see for example [24]), there are other cases of this lack of facilitation effect within the debate on the specific modality or supramodal nature of the attentional system. For example, a study by Li et al. [135], in the field of implicit statistical learning (ISL) investigated whether multimodal input sources are processed separately or are integrated across modalities. Results reported that there did not highlight neither a facilitative learning effect nor an interference effect. These findings confirm theories that suggest the existence of multiple modality-specific mechanisms rather than a single central one. Another study reported that neural responses in the primary visual cortex can be reduced by auditory inputs that are synchronously presented with visual inputs [136]. Moreover, it seems that the facilitation effect of multisensory integration is strong when each unimodal signal is weak whereas this facilitation effect seems to decrease with increasing intensity of unimodal stimuli [28, 132]. It would therefore be interesting in future research to manipulate the intensity of crossmodal stimulations in a counting task, to verify for the occurrence of crossmodal facilitation.

The Synchrony main effect in both Experiment 1 and 2 also presents notable results. The asynchronous sequences seem to produce a worse performance than when the presentation is synchronous. Specifically, when visual and auditory stimuli are non-synchronized, participants tend to lose count. This also happens when the speed of presentation of the stimuli is very low, as indeed can be seen from the results of Experiment 2 in which crossmodal interference also emerges in a situation in which the task is facilitated, as it happens for example in the Slow condition, generating a better performance than in the Moderate condition. This result is in line with crossmodal studies where interference is generated by a lack of correspondence between auditory and visual stimuli. Demonstrations of this phenomenon include crossmodal Stroop effects from a stream of auditory distractors on responses to a concurrent stream of visual targets (e.g., [137–144]). Also, McDonald et al. [78], demonstrated that the occurrence of an irrelevant sound improved the detectability of a subsequent light when the target light appeared in the same location as the sound, while performance decreased when there was a spatial mismatch between the irrelevant sound and the target light (see also [124] for similar results with complex visual scenes). In our results, a crossmodal interference produced by the desynchronization between visual and auditory stimuli deteriorates the performance, increasing underestimation errors.

Underestimation becomes even more relevant when the stimuli are presented at a faster pace. Results of Experiment 1 show a different pattern of effects in the interaction between Isochrony and Speed. Namely, participants tend to lose more count in the Non-isochronous condition when stimuli are presented faster compared to all other interactions. This result confirms previous studies on fluency [99, 100] while offering evidence about the interaction with speed, which in our study proves to modulate the effects of irregularity. While in Experiment 1 participants make roughly the same number of errors when the presentation speed is moderate in all sequences, the direction of the error changes according to Isochrony: in the

isochronous condition there are more underestimation errors while in the Non-isochronous condition the errors tend more towards overestimation. This pattern of results can be interpreted considering the role of attention. When the interval of presentation of the stimuli is irregular the participants must pay more attention because they are unable to create expectations about the appearance of the following stimuli. For this reason, in the Non-isochronous condition presented slowly, the sequence calls for an increased focus to detect the unexpected appearance of the next stimulus. The percentage of error, even if in excess, is still very low. On the other hand, in the condition in which the presentation of the stimuli is isochronous and the speed moderate, the process can rely on the formation of expectations and entrainment, which could result in a percentage of error, albeit minimal, towards an underestimation. This hypothesis could be based on studies on the "Dynamic attending theory" [102–104]. Since the amount of errors in our Moderate condition is very low, we can say that when the presentation is also isochronous, therefore making the task even easier, the participants tend to be less accurate and to make errors by underestimation, while when the non-regularity of the intervals does not allow an attentional entrainment, participants lose this attentional facilitation and tend to get more confused in counting, thus counting by excess, as if they had illusorily perceived extra stimuli. The lack of significant results for the Isochronous condition in Experiment 2, together with the absence of a significant interaction with the Speed factor (differently from the effects recorded in Experiment 1), gives us some interesting insights. We interpret this result, that seems not to benefit from the regularity of stimuli presentation in the Slow condition, as actually a confirmation of the results of the first experiment. Indeed, as the speed increases, it seems that the participants benefit from the regularity of the presentation intervals, thanks to the creation of expectations over time that regularity provides [102–104]. In a condition where the presentation interval is slower, this metacognitive cue to support the task, as intended by Stevenson and Carlson's [100] definition, seems to be no longer necessary. In other words, in the second experiment in which the stimuli presentation speed is moderate or slow, we can assume that the participants no longer need to rely on the creation of temporal expectations (given by the regularity of the intervals) and therefore this variable no longer emerges as significant, either as a main effect or as a function of speed.

To sum up, there appears to be a linear relationship between errors and speed. More interestingly, as demonstrated in Experiment 1, it seems that errors that come from the speed factor could be mitigated by synchrony and enhanced by non-isochrony. We can interpret these results in line with the idea that in event counting, each step involves the integration of the perception of a visual stimulus and a representation of the next number in a counting sequence, associated with a motor plan involving covert articulation of the current count [100]. Thus, following Stevenson and Carlson's interpretation [100], tasks that require regular updating of events, like covert counting, come with a cost, namely, that running counts could be unstable. This instability might be partially compensated with metacognitive monitoring, which could occur implicitly: Cues like temporal fluency (regularity) may be used to create expectancies useful to stabilize the running count. It is possible to summarize by saying that there are some factors that make the task of counting complex, but as long as it is possible to rely on a metacognitive cue such as regularity, we can compensate for the detrimental effects of Speed and Asynchrony.

In general, the only trace of facilitation we observed in Experiment 1, although not cross-modal, is with moderate and isochronous sequences, both synchronized with sounds or without any sound at all, partially confirming previous literature [76–79]. The effect of Non-isochronous intervals emerged especially in the "Fast" conditions, in which the performances seem to be the worst. These results primarily show us that task-irrelevant sounds can exogenously capture our attention [124] with the power to also disrupt a longer process such as counting, resulting in missing some task-relevant stimuli.

Finally, with this study it is possible to add evidence to the phenomenon of competition for attentional resources between the visual and the auditory modalities (see [85–87]). In our study, auditory stimuli that are not synchronized with the simultaneous presentation of visual stimuli negatively affected the performance in a counting task. This could happen because the attentional resources are automatically captured by the auditory modality, making them temporarily unavailable to the visual processes and consequently affecting the counting process itself. Thus, this study add evidence in support of the supramodal attentional model [85–87].

In conclusion, it should be noted that some limitations are present. Firstly, due to COVID-19 prevention measures, all administrations were conducted online. While it is true that online experiments may not provide the same level of accuracy as laboratory administrations, with slightly more variability in all measurements, several studies suggest that modern online administration platforms are reliable in terms of accuracy and precision of presentation and response times [145, 146]. In addition, Bridges, Pitiot, MacAskill & Peirce [147] found that PsychoPy is generally one of the best platforms, achieving results very close to millisecond accuracy on different browser/operating system combinations. A second limitation may be associated with the fact that since the number of visual and auditory stimuli was the same in the bimodal condition, we cannot completely rule out the possibility that some participants mistakenly counted sounds instead of visual stimuli. Despite this, the fact that this misunderstanding of the task could have had a significant impact on results can be excluded, since the participants should have had to: 1) completely ignore instructions and training; 2) switch their task continuously (as the stimuli were randomized), counting visual stimuli in unimodal sequences and then sounds in bimodal stimuli. The probability that several participants conducted the experiment in this way is extremely thin. Finally, a third limitation might be associated with the length of the sequences. Although a change in the degree of vigilance is unlikely with sequences as short as the ones used in this study, it is nevertheless possible that different sequence lengths had a different impact on performance.

Supporting information

S1 Data.

(XLSX)

S2 Data.

(XLSX)

Author Contributions

Conceptualization: Claudia Del Gatto, Riccardo Brunetti.

Data curation: Claudia Del Gatto, Allegra Indraccolo, Tiziana Pedale.

Formal analysis: Allegra Indraccolo, Tiziana Pedale, Riccardo Brunetti.

Investigation: Claudia Del Gatto.

Methodology: Claudia Del Gatto, Allegra Indraccolo, Riccardo Brunetti.

Project administration: Allegra Indraccolo.

Resources: Riccardo Brunetti.

Software: Claudia Del Gatto.

Supervision: Claudia Del Gatto, Riccardo Brunetti.

Validation: Claudia Del Gatto.

Visualization: Claudia Del Gatto.

Writing – original draft: Claudia Del Gatto.

Writing – review & editing: Claudia Del Gatto, Allegra Indraccolo, Tiziana Pedale, Riccardo Brunetti.

References

1. Logie RH & Baddeley AD. Cognitive processes in counting. *J Exp Psychol Learn Mem Cogn*. 1987; 13:310–26.
2. Hitch GJ. The role of short-term working memory in mental arithmetic. *Cogn Psychol*. 1978; 10:302–23.
3. Awh E, Vogel EK, Oh S H. Interactions between attention and working memory. *Neuroscience*. 2006; 139: 201–8. <https://doi.org/10.1016/j.neuroscience.2005.08.023> PMID: 16324792
4. Ortuno F, Ojeda N, Arbizu J, Lopez P, Marti-Climent JM, Peñuelas I, et al. Sustained attention in a counting task: normal performance and functional neuroanatomy. *Neuroimage*, 2002; 17:411–20. <https://doi.org/10.1006/nimg.2002.1168> PMID: 12482094
5. Spence C. Crossmodal correspondences: A tutorial review. *Attention, Perception, Psychophys*. 2011; 73:971–95. <https://doi.org/10.3758/s13414-010-0073-7> PMID: 21264748
6. Ben-Artzi E & Marks LE. Visual-auditory interaction in speeded classification: Role of stimulus difference. *Percept Psychophys*. 1995; 57:1151–62. <https://doi.org/10.3758/bf03208371> PMID: 8539090
7. Bernstein IH & Edelman BA. Effects of some variations in auditory input upon visual choice reaction time. *J Exp Psychol*. 1972; 87:241–47.
8. Evans KK & Treisman A. Natural cross-modal mappings between visual and auditory features. *J. Vis*. 2010; 10:6. <https://doi.org/10.1167/10.1.6> PMID: 20143899
9. Melara RD & O'Brien TP. Interaction between synesthetically corresponding dimensions. *J Exp Psychol Gen*. 1987; 116:323–36.
10. Patching GR & Quinlan PT. Garner and congruence effects in the speeded classification of bimodal signals. *J Exp Psychol Hum Percept Perform*. 2002; 28:755–75. PMID: 12190249
11. Rusconi E, Kwan B, Giordano B, Umiltà C & Butterworth B. Spatial representation of pitch height: The SMARC effect. *Cognition*. 2006; 99:113–29. <https://doi.org/10.1016/j.cognition.2005.01.004> PMID: 15925355
12. Ludwig VU, Adachi I & Matsuzawa T. Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans. *Proc Natl Acad Sci*. 2011; 108:20661–65. <https://doi.org/10.1073/pnas.1112605108> PMID: 22143791
13. Marks LE. On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *J Exp Psychol Hum Percept Perform*. 1987; 13:384–94. <https://doi.org/10.1037//0096-1523.13.3.384> PMID: 2958587
14. Hubbard TL. Synesthesia-like mappings of lightness, pitch, and melodic interval. *Am J Psychol*. 1996; 109:219–38. PMID: 8644886
15. Martino G & Marks LE. Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*. 1999; 28:903–23. <https://doi.org/10.1068/p2866> PMID: 10664781
16. Melara RD. Dimensional interaction between color and pitch. *J Exp Psychol Hum Percept Perform*. 1989; 15:69–79. <https://doi.org/10.1037//0096-1523.15.1.69> PMID: 2522534
17. Mondloch CJ & Maurer D. Do small white balls squeak? Pitch-object correspondences in young children. *Cogn Affect Behav Neurosci*. 2004; 4:133–6. <https://doi.org/10.3758/cabn.4.2.133> PMID: 15460920
18. Gallace A & Spence C. Multisensory synesthetic interactions in the speeded classification of visual size. *Percept Psychophys*. 2006; 68:1191–203. <https://doi.org/10.3758/bf03193720> PMID: 17355042
19. Clark HH & Brownell HH. Position, direction, and their perceptual integrality. *Percept Psychophys*. 1976; 19:328–34.
20. Arrighi R, Alais D, Burr D. Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *J Vision*. 2006; 6: 6–6. <https://doi.org/10.1167/6.3.6> PMID: 16643094
21. Bahrick LE, Lickliter R & Flom R. Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Curr Dir Psychol Sci*. 2004; 13:99–102.

22. Brunetti R, Indraccolo A, Del Gatto C, Spence C & Santangelo V. Are crossmodal correspondences relative or absolute? Sequential effects on speeded classification. *Atten Percept Psycho*. 2018; 80: 527–34. <https://doi.org/10.3758/s13414-017-1445-z> PMID: 29116614
23. Delogu F, Brunetti R, Inuggi A, Campus C, Del Gatto C, D'Ausilio A. That does not sound right: Sounds affect visual ERPs during a piano sight-reading task. *Behav Brain Res*. 2019; 367:1–9. <https://doi.org/10.1016/j.bbr.2019.03.037> PMID: 30922941
24. Del Gatto C, Brunetti R, Delogu F. Cross-modal and intra-modal binding between identity and location in spatial working memory: The identity of objects does not help recalling their locations. *Memory*. 2016; 24:603–15. <https://doi.org/10.1080/09658211.2015.1034137> PMID: 25920517
25. Robinson CW, Sloutsky VM. Effects of multimodal presentation and stimulus familiarity on auditory and visual processing. *J Exp Child Psychol*. 2010; 107:351–8. <https://doi.org/10.1016/j.jecp.2010.04.006> PMID: 20553691
26. Jones D, Morris N. Irrelevant speech and serial recall: Implications for theories of attention and working memory. *Scand J Psychol*. 1992; 33:212–29. <https://doi.org/10.1111/j.1467-9450.1992.tb00911.x> PMID: 1488646
27. Brunetti R, Indraccolo A, Mastroberardino S, Spence C, Santangelo V. The impact of cross-modal correspondences on working memory performance. *J Exp Psychol Hum Percept Perform*. 2017; 43:819–3. <https://doi.org/10.1037/xhp0000348> PMID: 28345948
28. Stein BE, Meredith MA. *The Merging of the Senses*. The MIT press; 1993.
29. Giard MH, Peronnet F. Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *J Cogn Neurosci*. 1999; 11:473–90. <https://doi.org/10.1162/089892999563544> PMID: 10511637
30. Bahrick LE, Lickliter R. Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev Psychol*. 2000; 36:190–201. <https://doi.org/10.1037/0012-1649.36.2.190> PMID: 10749076
31. Fort A, Delpuech C, Pernier J, Giard MH. Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cereb Cortex*. 2002; 12:1031–9. <https://doi.org/10.1093/cercor/12.10.1031> PMID: 12217966
32. Sinnott S, Soto-Faraco S, Spence C. The co-occurrence of multisensory competition and facilitation. *Acta Psychol*. 2008; 128:153–61. <https://doi.org/10.1016/j.actpsy.2007.12.002> PMID: 18207117
33. Colavita FB. Human sensory dominance. *Percept Psychophys*. 1974; 16:409–12.
34. McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature*. 1976; 264:746–8. <https://doi.org/10.1038/264746a0> PMID: 1012311
35. Shams L, Kamitani Y, Shimojo S. What you see is what you hear. *Nature*. 2000; 408:788.
36. Shams L, Kamitani Y, Shimojo S. Visual illusion induced by sound. *Cogn Brain Res*. 2002; 14:147–52.
37. Sloutsky VM, Napolitano AC. Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Dev*. 2003; 74:822–33. <https://doi.org/10.1111/1467-8624.00570> PMID: 12795392
38. Colavita FB, Tomko R, Weisberg D. Visual prepotency and eye orientation. *Bull Psychon Soc*. 1976; 8:25–6.
39. Posner MI, Nissen MJ, Klein RM. Visual dominance: An information-processing account of its origins and significance. *Psychol Rev*. 1976; 83:157–71. PMID: 769017
40. Burr D, Banks MS, Morrone MC. Auditory dominance over vision in the perception of interval duration. *Exp Brain Res*. 2009; 198(1):49–57. <https://doi.org/10.1007/s00221-009-1933-z> PMID: 19597804
41. Keil J. Double flash illusions: current findings and future directions. *Front Neurosci*. 2020; 14: 298. <https://doi.org/10.3389/fnins.2020.00298> PMID: 32317920
42. Shams L, Kim R. Crossmodal influences on visual perception. *Phys of life rev*. 2010; 7(3):269–84.
43. Senkowski D, Talsma D, Herrmann CS, Woldorff MG. Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. *Exp Brain Res*. 2005; 166:411–26 <https://doi.org/10.1007/s00221-005-2381-z> PMID: 16151775
44. Shams L, Kamitani Y, Thompson S, Shimojo S. Sound alters visual evoked potentials in humans. *Neuroreport*. 2001; 12:3849–52. <https://doi.org/10.1097/00001756-200112040-00049> PMID: 11726807
45. Keil J, Müller N, Hartmann T, Weisz N. Prestimulus beta power and phase synchrony influence the sound-induced flash illusion. *Cereb Cortex*. 2014; 24:1278–88. <https://doi.org/10.1093/cercor/bhs409> PMID: 23300109
46. Watkins S, Shams L, Josephs O, Rees G. Activity in human V1 follows multisensory perception. *Neuroimage*. 2007; 37:572–8. <https://doi.org/10.1016/j.neuroimage.2007.05.027> PMID: 17604652

47. Watkins S, Shams L, Tanaka S, Haynes JD, Rees G. Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage*. 2006; 31:1247–56. <https://doi.org/10.1016/j.neuroimage.2006.01.016> PMID: 16556505
48. Abadi RV, Murphy JS. Phenomenology of the sound-induced flash illusion. *Exp Brain Res*. 2014; 232:2207–20. <https://doi.org/10.1007/s00221-014-3912-2> PMID: 24691756
49. Gebhard JW, Mowbray GH. On discriminating the rate of visual flicker and auditory flutter. *Am J of Psychol*. 1959; 72: 521–9. PMID: 13827044
50. Regan D, Spekreijse H. Auditory—Visual Interactions and the Correspondence between Perceived Auditory Space and Perceived Visual Space. *Perception*. 1977; 6:133–8. <https://doi.org/10.1068/p060133> PMID: 876814
51. Shipley T. Auditory flutter-driving of visual flicker. *Science*. 1964; 145: 1328–30. <https://doi.org/10.1126/science.145.3638.1328> PMID: 14173429
52. Wada Y, Kitagawa N, Noguchi K. Audio–visual integration in temporal perception. *Int J Psychophysiol*. 2003; 50:117–24. [https://doi.org/10.1016/s0167-8760\(03\)00128-4](https://doi.org/10.1016/s0167-8760(03)00128-4) PMID: 14511840
53. Ljung R, Marsh JE, Sörqvist P. Distraction of counting by the meaning of background speech: Are spatial memory demands a prerequisite? *Appl Cogn Psychol*. 2015; 29:584–91.
54. Ngo MK, Cadieux ML, Sinnott S, Soto-Faraco S, Spence C. Reversing the Colavita visual dominance effect. *Exp brain res*. 2011; 214(4):607–18. <https://doi.org/10.1007/s00221-011-2859-9> PMID: 21909983
55. Goldstone S, Boardman WK, Lhamon W. T. Intersensory comparisons of temporal judgments. *J Exp Psychol*. 1959; 57(4):243. <https://doi.org/10.1037/h0040745> PMID: 13641599
56. Fendrich R, Corballis PM. The temporal cross-capture of audition and vision. *Percept Psychophys*. 2001; 63(4):719–25. <https://doi.org/10.3758/bf03194432> PMID: 11436740
57. Morein-Zamir S, Soto-Faraco S, Kingstone A. Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Res*. 2003; 17(1):154–63. [https://doi.org/10.1016/s0926-6410\(03\)00089-2](https://doi.org/10.1016/s0926-6410(03)00089-2) PMID: 12763201
58. Recanzone GH. Auditory influences on visual temporal rate perception. *J Neurophysiol*. 2003; 89(2):1078–93. <https://doi.org/10.1152/jn.00706.2002> PMID: 12574482
59. Walker JT, Scott KJ. Auditory–visual conflicts in the perceived duration of lights, tones, and gaps. *J Exp Psychol Human*. 1981; 7(6):1327. <https://doi.org/10.1037//0096-1523.7.6.1327> PMID: 6458656
60. Fraisse P II. Rythmes auditifs et rythmes visuels. *Ann Psychol*. 1948; 49(1):21–42.
61. Kolars PA, Brewster JM. Rhythms and responses. *J Exp Psychol Human*. 1985; 11(2):150. <https://doi.org/10.1037//0096-1523.11.2.150> PMID: 3159834
62. Repp BH, Penel A. Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *J Exp Psychol Human*. 2002; 28(5):1085. PMID: 12421057
63. Chen Y, Repp BH, Patel AD. Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions. *Hum Mov Sci*. 2002; 21(4):515–32. [https://doi.org/10.1016/s0167-9457\(02\)00138-0](https://doi.org/10.1016/s0167-9457(02)00138-0) PMID: 12450682
64. Repp BH, Penel A. Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychol Res*. 2004; 68(4):252–70. <https://doi.org/10.1007/s00426-003-0143-8> PMID: 12955504
65. Dunifon CM, Rivera S, Robinson CW. Auditory stimuli automatically grab attention: Evidence from eye tracking and attentional manipulations. *J Exp Psychol Human*. 2016; 42(12):1947. <https://doi.org/10.1037/xhp0000276> PMID: 27505224
66. Baddeley A, Salamé P. The unattended speech effect: Perception or memory? *J Exp Psychol Learn Mem Cogn*. 1986; 12:525–9. <https://doi.org/10.1037//0278-7393.12.4.525> PMID: 2945899
67. Buchner A, Irmen L, Erdfelder E. On the irrelevance of semantic information for the “irrelevant speech” effect. *Q J Exp Psychol A*. 1996; 49:765–79.
68. Hanley JR, Broadbent C. The effect of unattended speech on serial recall following auditory presentation. *Br J Psychol*. 1987; 78:287–97.
69. Jones DM, Macken WJ, Murray AC. Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Mem Cognit*. 1993; 21:318–28. <https://doi.org/10.3758/bf03208264> PMID: 8316094
70. LeCompte DC. Extending the irrelevant speech effect beyond serial recall. *J Exp Psychol Learn Mem Cogn*. 1994; 20:1396–408. <https://doi.org/10.1037//0278-7393.20.6.1396> PMID: 7983471
71. Lewis JL. Semantic processing with bisensory stimulation. *J Exp Psychol*. 1972; 96:455–7. <https://doi.org/10.1037/h0033653> PMID: 4645973

72. Sinnott S, Kingstone AA. preliminary investigation regarding the effect of tennis grunting: does white noise during a tennis shot have a negative impact on shot perception?. *PloS one*. 2010; 5(10): e13148.
73. Robinson CW, Parker JL. Effects of auditory input on a spatial serial response time task. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*; (eds. Papafragou, A., Grodner, D., Mirman, D. & Trueswell, J. C.) 2016; 2237–2242
74. Laughery D, Pesina N, Robinson CW. Tones disrupt visual fixations and responding on a visual-spatial task. *J Exp Psychol Human*. 2020; 46:1301–12. <https://doi.org/10.1037/xhp0000855> PMID: [32730069](https://pubmed.ncbi.nlm.nih.gov/32730069/)
75. Lo LY, Lai CC. Visual-auditory interactions on explicit and implicit information processing. *Cogn Process*. 2022; 23: 179–89. <https://doi.org/10.1007/s10339-022-01077-2> PMID: [35142948](https://pubmed.ncbi.nlm.nih.gov/35142948/)
76. Romei V, Murray MM, Cappe C, Thut G. The contributions of sensory dominance and attentional bias to cross-modal enhancement of visual cortex excitability. *J Cognitive Neurosci*. 2013; 25(7): 1122–35. https://doi.org/10.1162/jocn_a_00367 PMID: [23384192](https://pubmed.ncbi.nlm.nih.gov/23384192/)
77. Tivadar RI, Retsa C, Turoman N, Matusz PJ, Murray MM. Sounds enhance visual completion processes. *NeuroImage*. 2018; 179:480–8. <https://doi.org/10.1016/j.neuroimage.2018.06.070> PMID: [29959049](https://pubmed.ncbi.nlm.nih.gov/29959049/)
78. McDonald JJ, Teder-Sälejärvi WA, Hillyard SA. Involuntary orienting to sound improves visual perception. *Nature*. 2000; 407:906–8. <https://doi.org/10.1038/35038085> PMID: [11057669](https://pubmed.ncbi.nlm.nih.gov/11057669/)
79. Stein BE, London N, Wilkinson LK, Price DD. Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *J Cognitive Neurosci*. 1996; 8(6):497–506. <https://doi.org/10.1162/jocn.1996.8.6.497> PMID: [23961981](https://pubmed.ncbi.nlm.nih.gov/23961981/)
80. Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG. The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci*. 2010; 14(9):400–10. <https://doi.org/10.1016/j.tics.2010.06.008> PMID: [20675182](https://pubmed.ncbi.nlm.nih.gov/20675182/)
81. Van der Burg E, Olivers CN, Bronkhorst AW, Theeuwes J. Pip and pop: nonspatial auditory signals improve spatial visual search. *J Exp Psychol Human*. 2008; 34(5):1053. <https://doi.org/10.1037/0096-1523.34.5.1053> PMID: [18823194](https://pubmed.ncbi.nlm.nih.gov/18823194/)
82. Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Res*. 2002; 14(1):115–28. [https://doi.org/10.1016/s0926-6410\(02\)00066-6](https://doi.org/10.1016/s0926-6410(02)00066-6) PMID: [12063135](https://pubmed.ncbi.nlm.nih.gov/12063135/)
83. Alsius A, Navarra J, Campbell R, Soto-Faraco S. Audiovisual integration of speech falters under high attention demands. *Curr Biol*. 2005; 15(9):839–43. <https://doi.org/10.1016/j.cub.2005.03.046> PMID: [15886102](https://pubmed.ncbi.nlm.nih.gov/15886102/)
84. van Ee R. Stochastic variations in sensory awareness are driven by noisy neuronal adaptation: evidence from serial correlations in perceptual bistability. *JOSA A*. 2009; 26(12):2612–22. <https://doi.org/10.1364/JOSAA.26.002612> PMID: [19956332](https://pubmed.ncbi.nlm.nih.gov/19956332/)
85. Duncan J, Martens S, Ward R. Restricted attentional capacity within but not between sensory modalities. *Nature*. 1997; 387:808–10. <https://doi.org/10.1038/42947> PMID: [9194561](https://pubmed.ncbi.nlm.nih.gov/9194561/)
86. Eimer M, Driver J. An event-related brain potential study of cross-modal links in spatial attention between vision and touch. *Psychophysiology*. 2000; 37:697–705. PMID: [11037046](https://pubmed.ncbi.nlm.nih.gov/11037046/)
87. Eimer M, Van Velzen J. Crossmodal links in spatial attention are mediated by supramodal control processes: Evidence from event-related potentials. *Psychophysiology*. 2002; 39:437–49. PMID: [12212636](https://pubmed.ncbi.nlm.nih.gov/12212636/)
88. Martín-Signes M, Paz-Alonso PM, Chica AB. Connectivity of frontoparietal regions reveals executive attention and consciousness interactions. *Cerebr Cortex*. 2019; 29(11):4539–50. <https://doi.org/10.1093/cercor/bhy332> PMID: [30590403](https://pubmed.ncbi.nlm.nih.gov/30590403/)
89. Mullane JC, Lawrence MA, Corkum PV, Klein RM, McLaughlin EN. The development of and interaction among alerting, orienting, and executive attention in children. *Child Neuropsychol*. 2016; 22(2):155–76. <https://doi.org/10.1080/09297049.2014.981252> PMID: [25413609](https://pubmed.ncbi.nlm.nih.gov/25413609/)
90. Posner MI. Attentional networks and consciousness. *Front Psychol*. 2012; 3:64. <https://doi.org/10.3389/fpsyg.2012.00064> PMID: [22416239](https://pubmed.ncbi.nlm.nih.gov/22416239/)
91. Spagna A, Mackie MA, Fan J. Supramodal executive control of attention. *Front Psychol*. 2015; 6:65. <https://doi.org/10.3389/fpsyg.2015.00065> PMID: [25759674](https://pubmed.ncbi.nlm.nih.gov/25759674/)
92. Spagna A, Wu T, Kim K, Fan J. Supramodal executive control of attention: evidence from unimodal and crossmodal dual conflict effects. *Cortex*. 2020; 133:266–76. <https://doi.org/10.1016/j.cortex.2020.09.018> PMID: [33157346](https://pubmed.ncbi.nlm.nih.gov/33157346/)

93. Ricciardi E, Bonino D, Pellegrini S, Pietrini P. Mind the blind brain to understand the sighted one! Is there a supramodal cortical functional architecture? *Neurosci Biobehav R*. 2014; 41:64–77.
94. Roberts KL, Hall DA. Examining a supramodal network for conflict processing: a systematic review and novel functional magnetic resonance imaging data for related visual and auditory stroop tasks. *J Cognitive Neurosci*. 2008; 20(6):1063–78.
95. Allport DA, Antonis B, Reynolds P. On the division of attention: A disproof of the single channel hypothesis. *Q J Exp Psychol*. 1972; 24(2):225–35. <https://doi.org/10.1080/00335557243000102> PMID: 5043119
96. Arrighi R, Lunardi R, Burr D. Vision and audition do not share attentional resources in sustained tasks. *Front Psychol*. 2011; 2:56. <https://doi.org/10.3389/fpsyg.2011.00056> PMID: 21734893
97. Keitel C, Maess B, Schröger E, Müller M. M. Early visual and auditory processing rely on modality-specific attentional resources. *Neuroimage*. 2013; 70:240–9. <https://doi.org/10.1016/j.neuroimage.2012.12.046> PMID: 23287527
98. Carlyon RP, Plack CJ, Fantini DA, Cusack R. Cross-modal and non-sensory influences on auditory streaming. *Perception*. 2003; 32:1393–402. <https://doi.org/10.1068/p5035> PMID: 14959799
99. Carlson RA, Cassenti DN. Intentional control of event counting. *J Exp Psychol Learn Mem Cogn*. 2004; 30:1235–51. <https://doi.org/10.1037/0278-7393.30.6.1235> PMID: 15521801
100. Stevenson LM, Carlson RA. Consistency, not speed: Temporal regularity as a metacognitive cue. *Psychol Res*. 2020; 84: 88–98. <https://doi.org/10.1007/s00426-018-0973-z> PMID: 29330595
101. Nobre K, Coull JT. *Attention and time*. Oxford University Press. 2010.
102. Jones MR. Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychol Rev*. 1976; 83(5):323. PMID: 794904
103. Jones MR, Boltz M. Dynamic attending and responses to time. *Psychol Rev*. 1989; 96(3):459. <https://doi.org/10.1037/0033-295x.96.3.459> PMID: 2756068
104. Large EW, Jones MR. The dynamics of attending: How people track time-varying events. *Psychol Rev*. 1999; 106(1):119.
105. Escoffier N, Sheng DYJ, Schirmer A. Unattended musical beats enhance visual processing. *Acta Psychol*. 2010; 135(1):12–6. <https://doi.org/10.1016/j.actpsy.2010.04.005> PMID: 20451167
106. Lin C, Yeh M, Shams L. Subliminal audio-visual temporal congruency in music videos enhances perceptual pleasure. *Neurosci Lett*. 2022; 779: 136623. <https://doi.org/10.1016/j.neulet.2022.136623> PMID: 35398533
107. Nattkemper D, Ziessler M, Frensch PA. Binding in voluntary action control. *Neurosci Biobehav R*. 2010; 34(7):1092–1101. <https://doi.org/10.1016/j.neubiorev.2009.12.013> PMID: 20036685
108. Waszak F, Cardoso-Leite P, Hughes G. Action effect anticipation: neurophysiological basis and functional consequences. *Neurosci Biobehav R*. 2012; 36(2):943–59. <https://doi.org/10.1016/j.neubiorev.2011.11.004> PMID: 22108008
109. Prinz W, Aschersleben G, Koch I. Cognition and action. In: Prinz W, Aschersleben G, Koch I, editors. *Oxford handbook of human action*. Oxford University Press; 2009. p. 33–70.
110. D'Ausilio A, Brunetti R, Delogu F, Santonico C, Belardinelli M O. How and when auditory action effects impair motor performance. *Exp Brain Res*. 2010; 201(2):323–30. <https://doi.org/10.1007/s00221-009-2044-6> PMID: 19847407
111. Badets A, Koch I, Toussaint L. Role of an ideomotor mechanism in number processing. *Exp Psychol*. 2013; 60(1):34. <https://doi.org/10.1027/1618-3169/a000171> PMID: 22935328
112. Badets A, Pesenti M. Creating number semantics through finger movement perception. *Cognition*. 2020; 115(1):46–53.
113. Ellis N C, Hennesly RA. A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *Brit J Psychol*. 1980; 71(1):43–51.
114. Ellis N. Linguistic relativity revisited: The bilingual word-length effect in working memory during counting, remembering numbers, and mental calculation. In *Advances in psychology*. North-Holland; 1992. Vol. 83, p. 137–55.
115. Fraisse P. "Rhythm and Tempo", In Deutsch D. (Ed.), *The Psychology of Music*. Academic Press, New York, 1982; pp. 149–180.
116. Hammerschmidt D, Frieler K, Wöllner C. Spontaneous motor tempo: Investigating psychological, chronobiological, and demographic factors in a large-scale online tapping experiment. *Front Psychol*. 2021; 12: 677201. <https://doi.org/10.3389/fpsyg.2021.677201> PMID: 34248776
117. Moelants, D. Preferred tempo reconsidered. In *Proceedings of the 7th international conference on music perception and cognition*. 2002; pp. 1–4.

118. Madison G, Paulin J. Ratings of speed in real music as a function of both original and manipulated beat tempo. *J Acoust Soc Am*. 2010; 128: 3032–40. <https://doi.org/10.1121/1.3493462> PMID: [21110598](https://pubmed.ncbi.nlm.nih.gov/21110598/)
119. Pereira L A, Ramos D, & Bueno J L. The influence of different musical modes and tempi on time perception. *Acta Psychol*. 2022; 229:103701. <https://doi.org/10.1016/j.actpsy.2022.103701> PMID: [35952511](https://pubmed.ncbi.nlm.nih.gov/35952511/)
120. Radeau M, Bertelson P. Auditory-visual interaction and the timing of inputs. *Psychol Res*. 1987; 49:17–22.
121. Broadbent DE. Successive responses to simultaneous stimuli. *Q J Exp Psychol*. 1956; 8:145–52.
122. Simon JR, Craft JL. Effects of an irrelevant auditory stimulus on visual choice reaction time. *J Exp Psychol*. 1970; 86:272–4. <https://doi.org/10.1037/h0029961> PMID: [5496384](https://pubmed.ncbi.nlm.nih.gov/5496384/)
123. Spence C, Ranson J, Driver J. Cross-modal selective attention: On the difficulty of ignoring sounds at the locus of visual attention. *Percept Psychophys*. 2000; 62:410–24. <https://doi.org/10.3758/bf03205560> PMID: [10723219](https://pubmed.ncbi.nlm.nih.gov/10723219/)
124. Pedale T, Mastroberardino S, Capurso M, Bremner AJ, Spence C, Santangelo V. Crossmodal spatial distraction across the lifespan. *Cognition*. 2021; 210:104617. <https://doi.org/10.1016/j.cognition.2021.104617> PMID: [33556891](https://pubmed.ncbi.nlm.nih.gov/33556891/)
125. Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, et al. PsychoPy2: Experiments in behavior made easy. *Behav Res Methods*. 2019; 51:195–203. <https://doi.org/10.3758/s13428-018-01193-y> PMID: [30734206](https://pubmed.ncbi.nlm.nih.gov/30734206/)
126. Rouder JN, Morey RD, Speckman PL, Province JM. Default Bayes factors for ANOVA designs. *J Math Psychol*. 2012; 56:356–74.
127. Jeffreys H. *Theory of Probability*. Oxford University Press: Oxford, New York; 1961.
128. Keyzers C, Gazzola V, Wagenmakers EJ. Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nat Neurosci*. 2020; 23:788–99. <https://doi.org/10.1038/s41593-020-0660-4> PMID: [32601411](https://pubmed.ncbi.nlm.nih.gov/32601411/)
129. Westfall PH. Multiple testing of general contrasts using logical constraints and correlations. *J Am Stat Assoc*. 1997; 92:299–306.
130. van den Bergh D, Van Doorn J, Marsman M, Draws T, Van Kesteren EJ, Derks K, et al. A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. *Ann Psychol*. 2020; 120:73–96.
131. Welford AT. The 'psychological refractory period' and the timing of high-speed performance—a review and a theory. *Br J Psychol Gen Sect*. 1952; 43:2–19.
132. Diederich A, Colonius H. Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Percept Psychophys*. 2004; 66:1388–404. <https://doi.org/10.3758/bf03195006> PMID: [15813202](https://pubmed.ncbi.nlm.nih.gov/15813202/)
133. Lippert M, Logothetis NK, Kayser C. Improvement of visual contrast detection by a simultaneous sound. *Brain Res*. 2007; 1173:102–9. <https://doi.org/10.1016/j.brainres.2007.07.050> PMID: [17765208](https://pubmed.ncbi.nlm.nih.gov/17765208/)
134. Noesselt T, Bergmann D, Hake M, Heinze HJ, Fendrich R. Sound increases the saliency of visual events. *Brain Res*. 2008; 1220:157–63. <https://doi.org/10.1016/j.brainres.2007.12.060> PMID: [18234167](https://pubmed.ncbi.nlm.nih.gov/18234167/)
135. Li X, Zhao X, Shi W, Lu Y, Conway CM. Lack of cross-modal effects in dual-modality implicit statistical learning. *Front Psychol*. 2018; 9:146. <https://doi.org/10.3389/fpsyg.2018.00146> PMID: [29535653](https://pubmed.ncbi.nlm.nih.gov/29535653/)
136. Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, et al. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci*. 2007; 27(42):11431–41. <https://doi.org/10.1523/JNEUROSCI.2252-07.2007> PMID: [17942738](https://pubmed.ncbi.nlm.nih.gov/17942738/)
137. Cowan N. The reality of cross-modal Stroop effects. *Percept Psychophys*. 1989; 45:87–8. <https://doi.org/10.3758/bf03208039> PMID: [2913575](https://pubmed.ncbi.nlm.nih.gov/2913575/)
138. Cowan N, Barron A. Cross-modal, auditory-visual Stroop interference and possible implications for speech memory. *Percept Psychophys*. 1987; 41:393–401. <https://doi.org/10.3758/bf03203031> PMID: [3601620](https://pubmed.ncbi.nlm.nih.gov/3601620/)
139. Driver J, Baylis GC. Cross-modal negative priming and interference in selective attention. *Bull Psychon Soc*. 1993; 31:45–8.
140. Elliott EM, Cowan N, Valle-Inclan F. The nature of cross-modal color-word interference effects. *Percept Psychophys*. 1998; 60:761–7. <https://doi.org/10.3758/bf03206061> PMID: [9682602](https://pubmed.ncbi.nlm.nih.gov/9682602/)
141. Morton J. Interaction of information in word recognition. *Psychol Rev*. 1969; 76:165–78.

142. Shimada H. Effect of auditory presentation of words on color naming: The intermodal Stroop effect. *Percept Mot Skills*. 1990; 70:1155–61. <https://doi.org/10.2466/pms.1990.70.3c.1155> PMID: 2399091
143. Miles C, Madden C, Jones DM. Cross-modal, auditory-visual Stroop interference: A reply to Cowan and Barron (1987). *Percept Psychophys*. 1989; 45:77–81. <https://doi.org/10.3758/bf03208036> PMID: 2913574
144. Thackray RI, Jones KN. Level of arousal during Stroop performance: Effects of speed stress and “distraction”. *Psychon Sci*. 1971; 23:133–5.
145. Anwyl-Irvine A, Dalmaijer ES, Hodges N, Evershed JK. Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behav Res Methods*. 2021; 53:1407–25. <https://doi.org/10.3758/s13428-020-01501-5> PMID: 33140376
146. Sauter M, Draschkow D, Mack W. Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain sciences*. 2020; 10:251. <https://doi.org/10.3390/brainsci10040251> PMID: 32344671
147. Bridges D, Pitiot A, MacAskill MR, Peirce JW. The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*. 2020; 8: e9414. <https://doi.org/10.7717/peerj.9414> PMID: 33005482