Journal of Experimental Psychology: Human Perception and Performance

Volume, number, pages

Serial number or other reference

Splitting Time: Sound-Induced Illusory Visual Temporal Fission and Fusion.

W. Paul Boyce, Seb Whiteford, William Curran, Gary Freegard & Christoph T. Weidemann

Swansea University and Queen's University Belfast

Word count: 18,877

W. Paul Boyce, ISRC, Ulster University, Northland Rd, Londonderry BT48 7JL, UK.
Seb Whiteford, Gary Freegard & Christoph T. Weidemann, Department of Psychology, Swansea University, Sketty, Swansea SA2 8PQ, UK.
William Curran, School of Psychology, Queen's University Belfast, Belfast BT9 5BN, UK.
Correspondence concerning this article should be addressed to William Paul Boyce.
Email: cratlagh@gmail.com

Auditory stimuli have been shown to alter visual temporal perception. For example, illusory temporal order is perceived when an auditory tone cues one side of space prior to the onset of simultaneously presented visual stimuli. Competing accounts attempt to explain such effects. The spatial gradient account of attention suggests speeded processing of visual stimuli in the cued space, whereas the impletion account suggests a Gestalt-like process where an attempt is made to arrive at a 'realistic' representation of an event given ambiguous conditions. Temporal ventriloquism – where visual temporal order judgement performance is enhanced when a spatially uninformative tone is presented prior to, and after, visual stimuli onset – argues that the temporal relationship of the auditory stimuli to visual stimuli, as well as the number of auditory stimuli equalling the visual stimuli, drives the mechanisms underlying these and related effects. Results from a series of experiments highlight putative inconsistencies in both the spatial gradient account of attention and the classical temporal ventriloquism account. We present novel behavioural effects – illusory temporal order via spatially uninformative tones, and illusory simultaneity via a single tone prior to visual stimuli onset – that can be accounted for by an expanded version of the impletion account.

Public Significance Statement

The present study demonstrates novel audio-induced visual-temporal-order effects using spatially neutral tones, while replicating related classic audio-visual effects. We interpret these findings as evidence that audio-visual integration takes evidence from various processes, assigning different weightings to each process dependent upon relative spatial locations, temporal characteristics, relative number of stimuli, and featural characteristics. With this interpretation in mind we propose a unifying account of the observed effects. Additionally, we suggest the use of the paradigms within this manuscript (and the associated effects) should be considered as part of sensory testing when measuring typical audio-visual integration, such as in cases of cochlear implantation.

Building a unified and coherent percept of our environment requires the interaction of multiple modalities. These interactions are generally beneficial to our interpretation of spatial and temporal events that occur in our immediate proximity. However, on occasion one modality has greater influence than the other during these interactions and can result in a percept that does not reflect physical events.

The visual modality has traditionally been understood to be the ‘dominant’ one (in terms of having greater influence during integration across modalities) when auditory and visual stimuli interact. One such example is that of visual capture, in which illusory auditory motion is perceived in the same direction as actual visual motion (Mateeff, Hohnsbein, & Noack, 1985; Spence, 2015). With this effect participants perceive illusory auditory motion of a static auditory stimulus while viewing a stimulus moving at a constant velocity. Another example of visual stimuli ‘dominating’ auditory is that of spatial ventriloquism, in which an auditory stimulus appears to be shifted from its true source in space to the location of temporally synchronised visual motion. A prime example of this is the classic ventriloquist's dummy, in which the sound's perceived location is matched to the location of the dummy's mouth (Radeau & Bertelson, 1987).

In recent years it has been demonstrated that conditions exist in which auditory processing 'dominates' visual processing. For example, a sequence of auditory tones can induce perceptual flashing of a single visual stimulus (Shams, Kamitani, & Shimojo, 2002; Andersen, Tiippana, & Sams, 2004). Similarly, Hidaka et al. (2009), and others (Teramoto et al., 2010, 2012), have shown that a moving auditory stimulus can induce illusory visual motion of a static visual stimulus. Finally, several authors have reported that an auditory stimulus can alter the perceived temporal onset of a visual stimulus (Burr, Banks, & Morrone, 2009; Vroomen & de Gelder, 2004).

As has been demonstrated in the above research, when auditory and visual stimuli are integrated one modality often alters the final perception of another in quite a pronounced fashion. With that in mind, we will examine some classic audio-visual effects and explanations for them, with a view of highlighting differences, characteristics, and claims that, at face value, may not necessarily be compatible with any one explanation.

Here, we address three accounts of visual temporal perception that are altered by sound (and in the case of the gradient and impletion accounts, reference classic accounts where the cues were visual) – the spatial gradient of attention, impletion, and temporal ventriloquism (Table 1). The spatial gradient of attention account (where attention decreases from the focus of attention to unattended areas in the visual field) of speeded visual processing of cued space suggests that an auditory (or visual) cue can increase the speed of information processing from the cued space because of a shift in visual attention to that space. This results in early entry into the mechanism of motion detection of any stimulus presented to the cued side, relative to the uncued side. This, in turn, can result in illusory motion of a line presented all at once (the line motion illusion (LMI)), or in illusory sequential order of simultaneously presented circles (Hikosaka, Miyauchi, & Shimojo, 1993b, 1993a; Shimojo, Miyauchi, & Hikosaka, 1997). This account is consistent with the idea of prior entry, which postulates that a stimulus presented in the cued space enters the perceptual system first and therefore is perceived first in time (Spence, Shore, & Klein, 2001; Spence & Parise, 2010; Santangelo & Spence, 2008). The spatial gradient of attention account will be referred to from here onwards as the 'gradient account'. The LMI is not dependent upon the visual stimuli being presented along the horizontal axis. It is worth noting that Schmidt, Fisher, and Pylshyn (1998) demonstrated that, when using multiple cues, the LMI (where the line was presented at various angles to the cues) still persisted when the target line was presented in line with one of the cues, but not when presented between 2 cues; suggesting a capacity to attend to multiple locations when cued but not the entire scene as a whole without such direction of attention.

An alternative explanation to the gradient account of speeded visual processing is the 'impletion' account. The impletion account argues that the cued space is interpreted as the beginning of the target stimulus during the binding of salient information, rather than a shift of attention resulting in speeded processing. The impletion account suggests the LMI and illusory sequential order effects are a consequence of attempting to interpret the most likely real-world events from ambiguous and/or spatially congruent stimuli (Downing & Treisman, 1997; Eagleman & Sejnowski, 2003; Fuller & Carrasco, 2009). Downing and Treisman (1997) demonstrated that visual cues presented simultaneously at either end of the line resulted in a perception of 'inward' line motion, in which both ends of the line appeared to move away from the cues towards the centre of the display. Additionally, they demonstrated that when a second line is presented to the right side of the rightmost cue, simultaneously with the first line presentation, both lines are perceived as moving to the right (see Schmidt (2000) for a rebuttal of Downing and Treisman's (1997) experiment 3 regarding voluntary attention). Eagleman and Sejnowski (2003) went on to demonstrate that when a second cue is presented to the opposite end of the line to the first cue, *after* the target line offset, the direction of the illusory motion is reversed. Similar to Downing and Treisman (1997), Tse, Cavanagh, and Nakayama (1998) demonstrated that illusory line motion can be induced when both visual cues are presented simultaneously at either end of the target line. In contrast to Downing and Treisman's (1997) findings, when the line was touching a given cue the illusory motion was perceived to move away from that cue towards the other, and not perceived to move 'inwards' towards the centre of the display. This suggests, like Downing and Treisman (1997) and Eagleman and Sejnowski (2003), that an attentional shift is not a requisite for inducing the LMI. Of course, the fact that the LMI and illusory sequential order can be induced by auditory cues (i.e. non-visual cues) suggests that the gradient account may still have a role to play in these visual illusions, even if it is not the sole driver of the effects, since auditory stimuli are qualitatively different to visual stimuli and cannot be 'seen' as the physical starting point of the visual stimulus.

Fuller and Carrasco (2009) presented evidence for both the gradient account – where a single cue was used – and impletion – where distributed cues were used in order to diminish effects of focal attention. They posited that impletion is the larger driver of the LMI given that there were no discernible differences in the perceived LMI between the cue types used. Schmidt and Klein (1997) also provided evidence that the gradient account alone is not sufficient to explain illusions related to the LMI, and indeed proposed an 'extended' gradient account that posits that visual signals near a cue are transmitted for a longer period of time than visual signals more distant from the cue.

In contrast to the above effects, enhancement in performance accuracy in a visual temporal order judgement (TOJ) task using auditory tones (Morein-Zamir, Soto-Faraco, & Kingstone, 2003) does not rely on spatially relevant information. When two central tones are paired with two sequentially presented light emitting diodes (LEDs), with the first tone preceding the first LED and the second tone occurring after the onset of the second LED, participants tend to make more accurate TOJs (at small SOAs). This effect is referred to as ‘temporal ventriloquism’ and in the classic definition the timing of the auditory stimulus is the most important factor in this ‘auditory capture’ (Freeman & Driver, 2008; Morein-Zamir et al., 2003). It is argued that the auditory stimulus appears to ‘pull’ a visual stimulus towards it in temporal perception, thus making TOJs more accurate in terms of objective performance. When two central tones were presented temporally between sequential circle presentations participants error rates tended to increase. Again, this appears to suggest that the circles were ‘pulled’ towards the tones in time thus inducing a perceived shorter stimulus onset asynchrony (SOA) between the circles. Interestingly, Morein-Zamir et al. (2003) added a caveat that there must be equal numbers of auditory and visual stimuli in order to induce temporal effects. This is due to a single tone between sequential LEDs having no observable effect on performance. However, this could also be due to a lack of sensitivity in measurement techniques. For example, when participants were asked to report apparent motion, Getzmann (2007) found that one centrally presented click between sequential squares increased the perception of apparent motion compared to a no-click condition. This suggests that the perceived SOA was shortened, thus challenging Morein-Zamir et al.’s (2003) claim that equal numbers of auditory and visual stimuli are required to induce the above temporal effect.

Temporal ventriloquism’s definition can be expanded to include the notion that it is the binding of auditory and visual stimuli that are perceived to be related to each other *after* a process of featural discrimination; in this expansion, timing of the auditory stimuli is not the main factor in the phenomenon. Growing evidence that timing is not the only major factor in temporal ventriloquism effects has emerged in recent years where effects have been abolished by manipulating features of the auditory stimuli. For example, when presenting one sine wave tone and one white noise burst, any enhancement effects are no longer observed (Keetels, Stekelenburg, & Vroomen, 2007; Roseboom, Kawabe, & Nishida, 2013b). This suggests discrimination judgements are being made between auditory stimuli before any potential integration with visual stimuli. If the auditory stimuli are featurally similar they are deemed to belong to the same event and therefore both are combined with the succeeding visual stimuli. If the auditory stimuli are featurally distinct, only one, or neither, of the auditory stimuli are combined with a succeeding visual event. Similarly, the double flash illusion demonstrated by Shams et al. (2002), where one circle presentation was perceived as two when accompanied by two tones, was found to be abolished when the auditory stimuli used were featurally different (Roseboom, Kawabe, & Nishida, 2013a). This suggests that featural similarity is an important driver in audio-visual illusions and hints at an auditory discrimination stage prior to audio-visual integration. This view is consistent with a Gestalt-like process at the level of intramodal processes on the way, or prior, to crossmodal integration (Spence, Sanabria, & Soto-Faraco, 2007). However, it is worth noting findings by Klimova, Nishida, and Roseboom (2017) (where featural differences did not abolish the temporal ventriloquist effect) together with research by Kafaligonul and Stoner (2010, 2012) that support the notion that the degree of featural similarity between auditory (or cross-modal) flankers may not modulate a temporal influence on visual stimuli over short time scales. This hints at a potentially different mechanism at play than that observed when using stimuli over longer time scales (Roseboom et al., 2013b).

A Bayesian perspective on audio-visual integration, as outlined by Körding et al. (2007) in relation to a multi-sensory cue combination study, proposes a causal inference model, where an 'ideal-observer' makes estimates about the cues they are sensing. For example, the likelihood of a stimulus originating from a specific spatial location is estimated (where the source signal is cor-

Table 1
Main effects being examined in this research and associated literature.

Effect	Literature
Gradient Account of Attention/Illusory Temporal Order	Hikosaka et al. (1993b, 1993a); Shimojo et al. (1997)
Impletion	Downing and Triesman (1997); Eagleman and Sejnowski (2003); Fuller and Carrasco (2009)
Termporal Ventriloquism	Morein-Zamir et al. (2003); Keetels et al. (2007); Roseboom et al. (2013b)

rupted by noise) and prior experience of analogous scenarios inform the likelihood of two stimuli originating from the same source, or individual sources. This information is combined to reach an inferred estimate of whether both stimuli are from one causal event and also estimates the position of the stimuli in space. The model accurately predicts audio-visual integration in perception for two audio-visual localisation tasks: one where an auditory and visual stimulus were presented simultaneously and participants reported the perceived position of each stimulus; and one where participants reported whether there was a single cause, or separate causes, for auditory and visual stimuli. The model supports the idea that the spatial relationship between auditory and visual stimuli factor into the perception of where in space both stimuli are presented, and if they share a common cause. Inferences about the characteristics of one stimulus (e.g. visual) are reached based on its relationship to another (e.g. auditory), which lends credibility to the notion of impletion, as outlined previously.

Beierholm, Quartz, and Shams (2009) highlighted that the Bayes rule does not inherently imply that, in the face of significant changes in a given stimulus, priors remain constant. Employing an expanded version of the audio-visual localisation task used by Körding et al. (2007) (adding a second session with adjusted contrast for the visual stimuli) they provided evidence that priors are independent of likelihoods, suggesting they are processed independently, and are later bound on the way to perception. This is, again, consistent with the idea of impletion.

Sato, Toyoizumi, and Aihara (2007) modelled spatial ventriloquism from a Bayesian inference perspective. When taking into consideration the position and timing of audio-visual stimuli, and considering whether the stimuli should be bound at all, their model accounted for most of the effects they examined. This approach, including there being no automatic assumption that all audio-visual stimuli should be bound, is consistent with the impletion account. Additionally, Shams, Ma, and Beierholm (2005) modelled the double-flash illusion using an ‘ideal observer’ from a Bayesian perspective. Their modelling supported a Bayesian inference approach, in which evidence is weighted when processing audio-visual stimuli prior to perceptual integration. Shams et al. (2005) argued that the double-flash illusion itself is a by-product of a “statistically optimal computational strategy” (p. 1927).

Taken together, the above Bayesian modelling of audio-visual integration provides support for impletion in terms of taking all available evidence and arriving at the most likely outcome in perception. Evidence also exists at a neural level for these types of audio-visual integration processes (Ursino, Crisafulli, di Pellegrino, Magosso, & Cuppini, 2017; Rohe, Ehlis, & Noppeney, 2019).

The research discussed above highlights clear interactions between the auditory and visual modalities. The underlying mechanisms driving these interactions continue to be debated, though there is some overlap in the accounts offered. This is particularly apparent in the case of impletion and the expanded definition of temporal ventriloquism (where featural discrimination appears to occur prior to audio-visual binding). Both give an account of the perceptual process where potential relationships between disparate stimuli are weighted and an attempt is made to arrive at an ecologically plausible representation in perception (this is distinct from the classic ventriloquism account that relies on SOA characteristics alone to describe and account for the observed 'pulling' effects). This suggests that perhaps there are common factors in the accounts outlined. The following research further examines the role of auditory stimuli when TOJs and a simultaneity judgement (SJ) were combined in a ternary response visual task. By doing so, the gradient account, impletion, and the original temporal ventriloquism account (where featural differences between auditory stimuli were not taken into account) described above were explored. In addressing the gradient account we used cues that coincided in space with visual target stimuli in order to induce illusory sequential order, referred to from here onwards as temporal fission. Note that the term 'temporal fission' should not to be confused with the fission effect reported by Shams et al. (2002), which 'split' a single visual stimulus in perception and increased the perceived number of stimuli, rather than temporal fission, which 'splits' a perceived temporal event in perception into two separate temporal events. We also used cues that were presented in neutral space (space that did not match that of the visual stimuli). We found that temporal fission could be induced by both cue conditions – i.e cues that were presented at the same spatial location as the target stimuli or at a neutral location. We will argue that this supports a role for the impletion account, and challenges the gradient account.

We addressed both the impletion and the original temporal ventriloquism accounts by presenting a single auditory cue to neutral space (space where no visual targets were presented) prior to sequential visual stimuli onset. This was done to test whether an auditory stimulus would 'pull' a visual stimulus towards it in perceptual time. Additionally, this also tested Morein-Zamir et al.'s (2003) claim that the number of auditory stimuli should match the number of visual stimuli in order to induce these types of audio-visual effects. We found that it was not necessary that the number of auditory and visual stimuli must be matched. We also found that illusory simultaneity (from here onwards referred to as temporal fusion) was achieved when the auditory stimulus was presented prior to sequential visual onset, which cannot be easily explained by the 'pulling' mechanism outlined in the original account of temporal ventriloquism. Note that the term 'temporal fusion' should not be confused with the fusion effect reported by Andersen et al. (2004) that 'fused' multiple stimuli in perception, thus reducing the perceived number of stimuli, rather than temporal fusion, which 'fuses' separate temporal events in perception into a single temporal event. We will also show that while a single tone presented prior to visual onset induces illusory temporal fusion, it also trends towards increased simultaneity report bias of simultaneous presentation of the visual stimuli. This suggests that there may be a relationship between the number of auditory and visual stimuli, and the relative spatial location of the auditory and visual stimuli, in terms of what type of illusion might be expected to be perceived.

Finally, we make a case that providing an SJ response allowed for a more sensitive measurement of perception, as detailed further in the discussion section of Experiment 2.

Experiment 1

In this experiment we used 2 tones presented to the left and right ears (via headphones), each approximately matching one of the visual target presentation locations, when attempting to induce temporal fission. The classic paradigm only uses 1 tone (Shimojo et al., 1997). We chose 2 tones in order to compare the 'strength' of temporal fission of spatially congruent tones (tones presented to analogous space to that of the visual stimuli – namely left and right ears/space) with the 'strength' of temporal fission of 2 tones in neutral space ('central' space – approximating the fixation cross in a given trial). As seen in Appendix C, we present data showing that 2 spatially congruent sequential tones, one each presented to analogous space to that of the respective visual stimuli, did in fact induce a stronger perception of temporal fission in our paradigm. Using 2 tones in opposing space (i.e. different locations on the horizontal axis either side of fixation - the left and right ear) as we did, allowed for a more straight-forward design in Experiment 2.

Methods

Participants. Twenty-seven participants, 8 male and 19 female (mean age 22.2 yrs; SD=4.45), with normal or corrected-to-normal vision and self-reported normal hearing, participated. All were students from Swansea University and were naïve to the purposes of the study. Ethical approval was received from the Department of Psychology Ethics Committee for this research.

An a priori power analysis was applied using data collected in a pilot study conducted prior to the experiments reported here. An identical condition to that used in this design displayed an effect size of $d = 4.39$ when comparing differences in the means of report bias corresponding to the actual presentation order of the visual stimuli between collapsed spatially opposing tones and baseline (no tones) in the simultaneous visual condition ($t(11) = 10.75$, $p < .001$, $SE = .06$, where (Bayes Factors) $BF = 4.147e + 04$, which provides extreme evidence indicating the presence of temporal fission – see the Results section in Experiment 1 for notes on how the BF was computed). This condition exists explicitly to detect whether temporal fission via prior entry was present, and is therefore one of the most important effects under consideration. Using GPower (Faul, Erdfelder, Lang, & Buchner, 2007) with 95% power and $\alpha = .001$ (consistent with the reported p-value from the pilot study) in a difference between two dependent means (matched pairs) power analysis, the recommended sample size was 8 for an actual power estimate of 97.59%. The sample size used here was deliberately larger due to concerns about baseline performance. For example, in the pilot study cited, only 12 participants remained from 27 in the analysis after the application of the exclusion criteria detailed below. Based on this concern, a strict time window for data collection, and potential for novel effects with unknown effect sizes, we set a stopping rule of 30, with a minimum of 25 participants in experiments 1-3.

Apparatus. Visual stimuli were presented using OpenSesame experimental software with PsychoPy backend on a 18" CRT LG monitor (resolution 1280x1024) with a 100Hz refresh rate, using a Windows XP PC. The monitor was 58cm from a chin rest. Auditory stimuli were presented via Sony Stereo Headphones. A photo-diode attached to the monitor triggered onset of auditory stimuli by activating a circuit switch which sampled a continuous tone from a Cello DVD player, amplified by a Technics Stereo Integrated Amplifier. Responses were made using a custom built three-button response box.

Stimuli and Procedure. Participants were instructed to choose one report from three options: both circles were presented simultaneously; the circle left of fixation was presented first; or the circle right of fixation was presented first. Each circle was 3.95 degrees of visual angle in diameter, and the distance from the centre of fixation to the centre of each flanking circle (one left and one right of fixation) was 15.16 degrees of visual angle.

Before the beginning of the experiment participants completed a staircase procedure of the task where visual stimuli only were presented and feedback was provided after every trial (a ‘thumbs up’ corresponded to a report that aligned with the actual presentation of the visual stimuli, and a ‘thumbs down’ corresponded to a report that differed to the actual presentation of the visual stimuli). This ensured that the task was not too easy or too difficult and catered for each individual’s

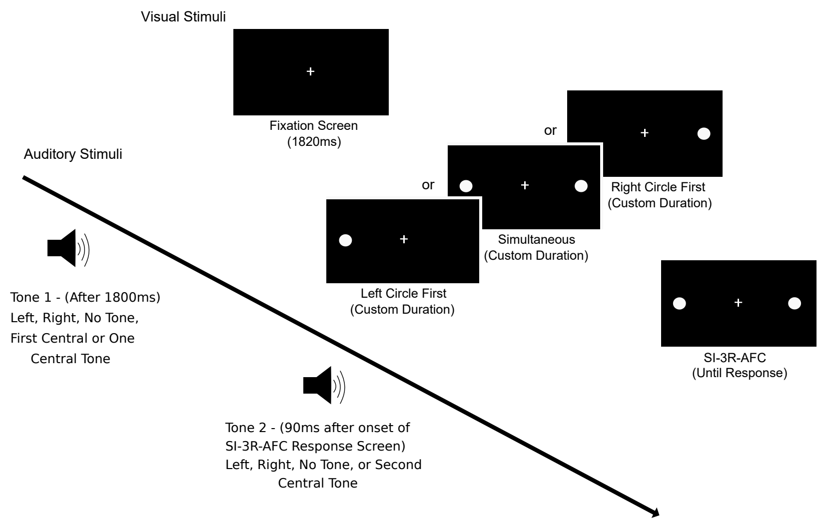

Figure 1. Trial sequence and timings for Experiment 1. The arrow shows the order of events from top to bottom of visual and auditory stimuli with the associated presentation times. Custom Duration reflects the presentation time acquired from the staircase phase of the task that was used for the duration of the first visual stimulus/stimuli. This varied across participants and was fixed for each individual experiment. Tone 1 was presented 1800ms into the fixation screen and 20ms later the first visual stimulus was presented and displayed on the monitor for the Custom Duration (ms) and consisted of one of three possibilities: a left circle, a right circle or both circles simultaneously. Tone 2 was presented 90ms after the second visual stimulus onset - the SI-3R-AFC screen (single interval, 3 response alternative forced choice) - which always consisted of both circles. The speaker icons list the possible tones that were presented to participants' left or right ears, or when presented 'centrally' via simultaneous binaural presentation, at the stated passage of time.

perceptual ability. The intended baseline for reports that corresponded to the actual presentation of the visual stimuli was approximately 75%. The procedure consisted of 6 blocks, each visual condition (both sequential presentation visual conditions, and the simultaneous presentation visual condition) appearing in 4 trials per block. All visual conditions taken together resulted in 12 trials in total per block. The starting default duration of the first visual stimulus in a sequential visual condition trial (either a circle left of fixation, or a circle right of fixation) was 40ms (this was the starting duration which was then adjusted as a participant undertook the staircase) followed by the onset of the second visual stimulus (presented to the opposite side of fixation). If, in a given block, a participant's report corresponded to the actual presentation of the visual stimuli less than 75% of the time, the following block's duration of the first presented visual stimulus increased by 10ms. If a participant's report corresponded to the actual presentation of the visual stimuli greater than 75% of the time the same duration was decreased by 10ms. If a participant's report corresponded to the actual presentation of the visual stimuli 75% of the time there was no change to the duration. The use of a staircase helped avoid ceiling and floor effects and ensured that participants' reports aligned with the actual presentation of visual stimuli ~75% of the time in the control conditions (where no tones were presented). It also helped address concerns raised by Van der Burg, Olivers, Bronkhorst, and Theeuwes (2008) and Schneider and Bavelier (2003) regarding the use of a ternary-response task. Namely, a large variability may exist among participants in terms of what criteria they set in order to make a simultaneous report. Since the point of subjective simultaneity (PSS) was not being examined explicitly, the staircase approach helped to ensure each participant could differentiate between sequentially and simultaneously presented visual stimuli consistently.

Once the staircase was completed, participants were asked to wear headphones that would present 7ms tones at a frequency of 3500Hz (at ~70 dB across conditions) to one ear followed by the other. They were instructed to ignore these tones as they did not provide any useful information regarding the visual task. It was stressed to participants that the aim of the task was to report what they actually perceived. Participants received feedback after every trial with the view to test whether any observed effects were resistant to feedback.

The experiment consisted of 3 visual conditions X 5 auditory conditions. The three visual conditions were; 1 circle left of fixation followed by 1 circle right of fixation (referred to as a 'sequential visual condition'); 1 circle right of fixation followed by 1 circle left of fixation (referred to as a 'sequential visual condition'); and both circles (one either side of fixation) presented simultaneously (referred to as the 'simultaneous visual condition'). The five auditory conditions were; 1

tone presented to the left ear followed by 1 tone presented to the right ear (referred to as 'spatially opposing tones'); 1 tone presented to the right ear followed by 1 tone presented to the left ear (referred to as 'spatially opposing tones'); 2 tones presented to analogous central space (achieved via 1 tone presented to both ears simultaneously twice); 1 tone presented to analogous central space (achieved via 1 tone presented to both ears simultaneously once); and a control condition where there was no auditory stimulus. For clarity, in reporting tones presented to 'central space' we will report the number of tones perceived rather than the number of tones actually presented, e.g. 1 tone presented to both ears simultaneously will be reported as 1 tone centrally. All visual conditions were matched with all auditory conditions for a completely balanced design. Hatched plots (Figures 2, 5, and 8) highlight all conditions and report options for all experiments. Tables 2 and 3 list the visual and auditory conditions by presentation category and spatial category respectively. These categories will be referenced often below, in the results and discussion sections. NOTE: when 1 tone was presented simultaneously to both ears to achieve analogous central presentation, volume was not adjusted compared to conditions where 1 tone was presented to 1 ear at a time.

Table 2
Visual stimuli arranged by presentation category.

Sequential Visual Conditions	Simultaneous Visual Condition
1 circle right of fixation followed by 1 circle left of fixation	Both circles simultaneously
1 circle left of fixation followed by 1 circle right of fixation	

When two auditory stimuli were presented sequentially in a condition, the first tone was always presented 20ms before visual onset and the second tone was always presented 90ms after the custom duration established in the staircase. When only one auditory stimulus was used, it was always presented 20ms before the first visual stimulus onset. Visual stimuli always remained displayed until report. Below, we explain the rationale for each condition, beyond having a balanced design that helped avoid any strategies that participants might employ.

Simultaneous visual condition and the various auditory conditions: When 2 tones accompanied the simultaneous visual condition we were measuring whether participants perceived temporal fission, when compared to the equivalent no tone control condition. When the 2 tones were presented to analogous space (via headphones) to that of the circles this was a variation of the classic temporal fission effect (Shimojo et al., 1997). When the 2 tones were presented to analogous central space (neutral space – i.e., the tones' location did not match the location of the visual stimuli) we were measuring whether spatially uninformative tones could also induce temporal fission, which would challenge the gradient account of the effect. One tone presented to analogous central space was measuring whether there was any effect on report bias from the perspective of claims made by Morein-Zamir et al. (2003), i.e., the claim that the number of tones must match the number of visual stimuli to induce an effect.

Table 3
Auditory stimuli arranged by spatial category.

Spatially Neutral Tone/s	Spatially Opposing Tones	Control
2 tones presented to analogous central space	1 tone in left ear followed by 1 tone in right ear	No tones
1 tone presented to analogous central space	1 tone in right ear followed by 1 tone in left ear	

Sequential visual conditions and the various auditory conditions: When tones were presented to analogous space to that of the visual stimuli, we were measuring whether there was an increase in report bias in line with the actual presentation order of the visual stimuli when the presentation order of the tones matched the presentation order of the circles (supporting the gradient account). When the presentation order of the tones was the inverse of the presentation order of the circles, we were measuring whether prior entry was present (which would support the gradient account) via decreased bias in report in line with that of the actual presentation order of the visual stimuli. When 2 tones were presented to analogous central space, we were measuring if there was an increase in bias of report in line with the actual presentation order of visual stimuli, which would be consistent with the classic temporal ventriloquism effect. One tone presented to analogous central space was measuring whether there was any effect on report bias from the perspective of claims made by Morein-Zamir et al. (2003).

Results

Participants whose reports did not correspond to the actual presentation of visual stimuli at least 34% (which equates to 17 trials out of 50. We rounded up from 33.33% of 50 trials due to it equating to 16.66 trials) of the time in any of the control (no tone) conditions were removed from subsequent analyses. This resulted in no removal of participants from Experiment 1. Similarly, observations/trials with response times <250 or >2500 ms were removed on the grounds that these observations were unlikely to have arisen from the decision processes of interest. This resulted in the exclusion of 505 observations (2.49% of trials).

Prior to analysis, the data were transformed using the arcsine of the square root of the proportion of trials where report bias corresponded to the actual presentation of visual stimuli in order to normalise the distribution for the data used in null hypothesis significance testing. The transformed data were used in calculating Bayes Factors (*BF*) using the ttestBF function, from the BayesFactor package in R Statistical Software (R Development Core Team, 2008), which performs

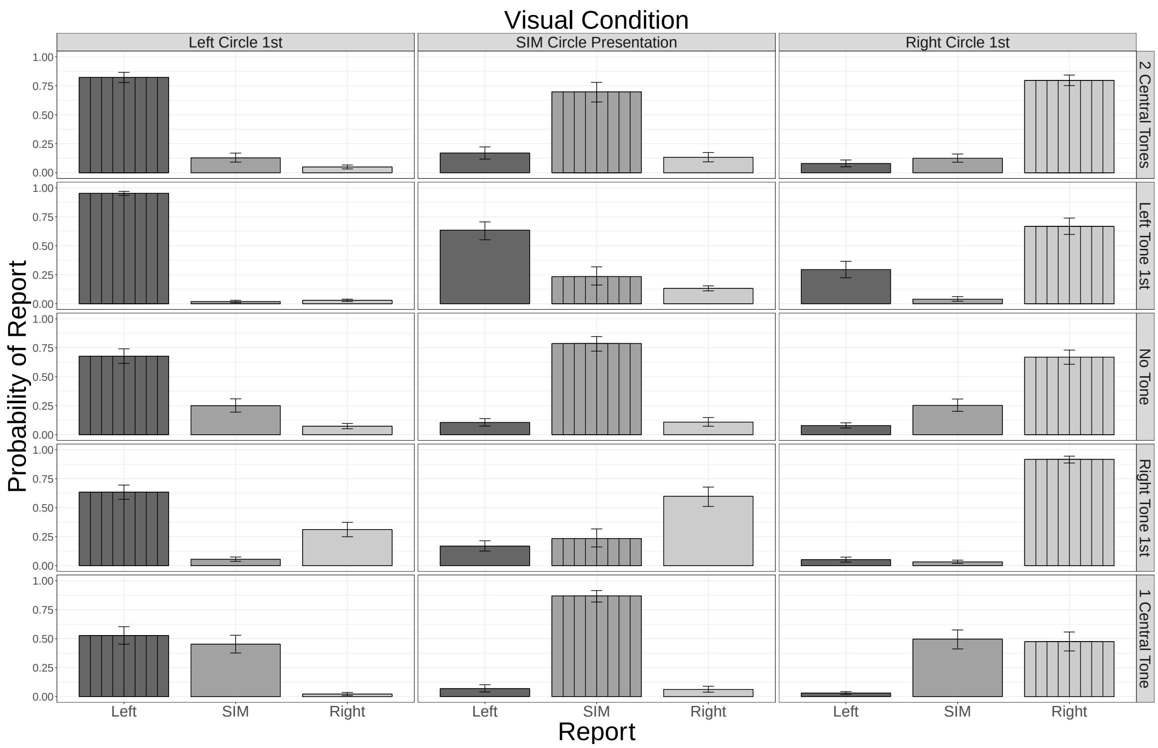

Figure 2. Experiment 1 report probability: The 3 visual conditions are labelled at the top of the grid horizontally. The leftmost column denotes sequential presentation of circles, where the first circle was presented to the left of fixation. The rightmost column denotes sequential presentation of circles, where the first circle was presented to the right of fixation. The central column denotes simultaneous presentation of circles, where a circle was presented to both left and right of fixation simultaneously. The 5 auditory conditions are labelled vertically on the rightmost edge of the grid, denoting (from top-to-bottom) the presentation of: 2 tones in analogous central space; a tone presented to the left ear followed by a tone presented to the right ear; no tones; a tone presented to the right ear followed by a tone presented to the left ear; 1 tone in analogous central space respectively. Error bars are bootstrapped within-subject 95% confidence intervals. Reports are labelled on the x-axis with reports corresponding to the actual presentation of visual stimuli highlighted with vertical hatching.

a ‘JZS’ t-test as described by Rouder, Speckman, Sun, Morey, and Iverson (2009). The default priors scale $r = \sqrt{2}/2$ was used, unless otherwise stated (for example, when prior evidence was available). Labelling used for interpretation of the BF values are based on those suggested by Jeffreys (1961) and adapted by Lee and Wagenmakers (2013). All statistical analyses, data shaping, and graphs of results contained herein were undertaken using RStudio (R Development Core Team, 2008; RStudio Team, 2015) and the package ggplot2 was used for plot generation (Wickham, 2009). Custom hatching patterns were accomplished using the EggHatch function developed by Boyce (2018).

Analysis of Report Bias Corresponding to Actual Presentation of Visual Stimuli. All *t*-tests below have been adjusted for multiple comparisons (including those only reported in figures) using the false discovery rate (FDR) correction (via the *p.adjust* function in R). Response probability t-tests were subject to a separate FDR correction due to examining the probability of reporting 1 of 3 potential reports rather than explicitly examining reports that corresponded to the actual presentation of visual as in the other t-tests.

Due to the relatively complex design (that was balanced in terms of conditions so as to avoid adoption of response strategies by participants) we conducted factorial analyses of sub-groups of conditions with the aim of establishing the presence of classic effects (temporal ventriloquism, and temporal fission). We also aimed to establish whether certain conditions must be met in order to induce said effects (shared space of auditory and visual stimuli for temporal fission, and the number of auditory stimuli matching the number of visual stimuli in order to induce illusory effects). These ANOVAs and t-tests helped provide support for and/or against prior entry and/or impletion.

We tested whether the classic temporal fission effect was replicated, which would support the gradient account. We conducted a 1 (visual condition: simultaneous visual condition) × 2 (auditory condition: collapsed spatially opposing tones vs. no tone) repeated measures ANOVA on report bias corresponding to the actual presentation of the visual stimuli.

There was a significant main effect of auditory condition $F(1, 26) = 200.47$, $MSE = 0.028$, $p < .001$, $\eta^2_G = .664$.

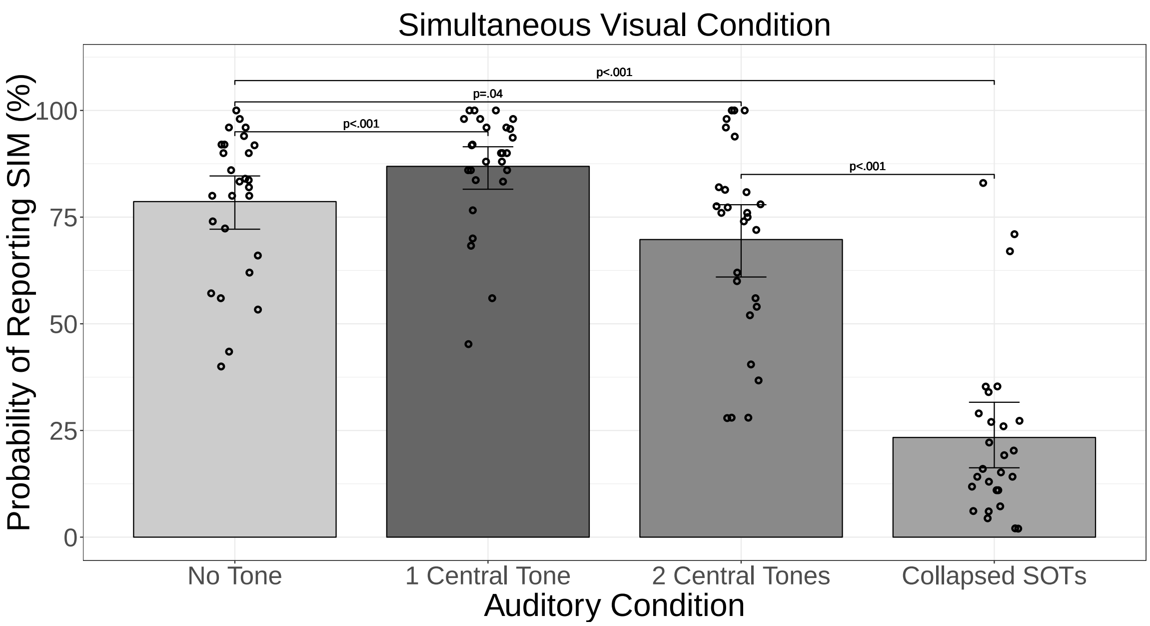

Figure 3. The probability of reporting simultaneous presentation when visual stimuli were presented simultaneously in Experiment 1: The probability (%) of reporting that visual stimuli were presented simultaneously is plotted on the y-axis and the auditory stimuli are labelled on the x-axis where the SOTs (spatially opposing tones) conditions have been collapsed. The reported *p*-values were obtained via null-hypothesis t-tests. Error bars are bootstrapped within-subject 95% confidence intervals.

Figure 3 shows that there was a reduction in report bias corresponding to the actual presentation of visual stimuli (reporting simultaneity) in the collapsed spatially opposing tones condition compared to baseline (no tone). The $BF = 3.09e + 12$ (adjusted using the "evidence updating" method – with the pilot data for the same condition – outlined by Ly, Etz, Marsman, and Wagenmakers (2017)), which provides extreme evidence indicating the presence of the temporal fission illusion in the collapsed spatially opposing tones condition. This replicates the classic temporal fission illusion which supports the gradient account from the view that the tones share the same space as the visual stimuli.

We tested whether there was further evidence for the gradient account via a 1 (visual condition: collapsed sequential visual conditions) × 2 (auditory condition: collapsed spatially opposing tones vs. no tone) repeated measures ANOVA on report bias corresponding to the actual presentation order of the visual stimuli.

There was a significant main effect of auditory condition $F(1,26) = 82.84$, $MSE = .005$, $p < .001$, $\eta^2_G = .251$, which, as can be seen in Figure 4, shows an increase in report bias corresponding to the actual presentation order of visual stimuli overall compared to baseline (no tone). NOTE: The probability of report data in Figure 4 show reports that corresponded to the actual sequential presentation order of a given sequential visual presentation. Report biases of sequential order opposite to the actual order are not included in the plot. For example, if the sequential visual condition was 'left circle first' we only included reports of 'left circle first'. The equivalent was true for the 'right circle first' visual condition.

The $BF = 9.42e{+}07$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides extreme evidence indicating the presence of increased report bias corresponding to the actual presentation order of visual stimuli in the collapsed spatially opposing tones condition, which in turn supports the gradient account. A note on the collapsed data here: conditions where a) the first tone cueing the analogous space the first circle was presented to, and b) the first tone cueing the analogous space of the second circle was presented to were collapsed (collapsed spatially opposing tones), and as a result some nuance is lost. Figure 2 shows increased report bias corresponding to the actual presentation order of visual stimuli when the first tone cues the same analogous space as the first circle presented in sequence, but conversely shows a reduction in report bias corresponding to the actual presentation order of the visual stimuli when the first tone presented cues the analogous space the second circle is presented to in sequence, consistent with the gradient account.

We tested whether classic temporal ventriloquism-like effects (in this instance reflected as an increase in report bias corresponding to the actual presentation order of visual stimuli – see Figure 2

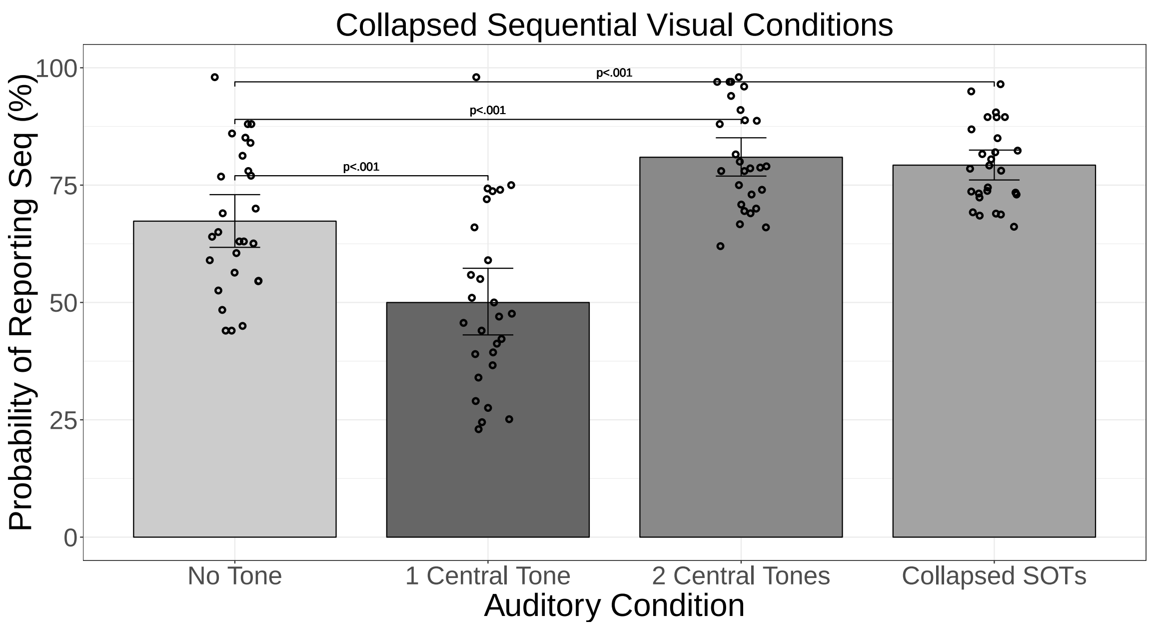

Figure 4. The probability of reporting sequential presentation (left circle first reports and right circle first reports that corresponded to the actual presentation order of visual stimuli collapsed) when visual stimuli were presented sequentially (left circle first and right circle first conditions collapsed) in Experiment 1: The probability (%) of reporting sequential order of visual stimuli corresponding to the actual presentation order of visual stimuli is plotted on the y-axis and the auditory stimuli are labelled on the x-axis where the SOTs (spatially opposing tones) conditions have been collapsed. The reported *p*-values were obtained via null-hypothesis t-tests. Error bars are bootstrapped within-subject 95% confidence intervals.

for illustration of this increase in probability of report bias corresponding to the presentation order of visual stimuli) in collapsed sequential visual conditions via 2 spatially neutral tones was replicated. We conducted a 1 (visual condition: collapsed sequential visual conditions) × 2 (auditory condition: 2 tones presented to analogous central space vs. no tone) repeated measures ANOVA on report bias corresponding with the actual presentation order of visual stimuli.

There was a significant main effect of auditory condition $F(1, 26) = 48.63$, $MSE = .008$, $p < .001$, $\eta^2_G = .201$. The relevant report bias data is contained in Figure 4.

The $BF = 75564.29$, which provides extreme evidence indicating the presence of increased report bias corresponding to the actual presentation order of visual stimuli in the 2 central tones condition, which is consistent with the classic temporal ventriloquism effect.

We conducted a repeated measures ANOVA on report bias corresponding to the actual presentation of the visual stimuli to determine if the spatial location of tones, relative to the visual stimuli, had an effect on temporal fission in Experiment 1. The ANOVA was a 1 (visual condition: simultaneous visual condition) × 3 (auditory presentation location: spatially opposing tones presented to analogous space to that of the visual stimuli vs. two tones presented to neutral space (analogous central space in this instance) vs. no tone presented to any space) design.

There was a significant main effect of auditory presentation location $F(2, 52) = 103$, $MSE = 0.033$, $p < .001$, $\eta^2_G = .561$.

As expected, spatial location is important when inducing visual temporal fission via auditory tones. However, the above ANOVA does not make clear if it is a requisite that auditory tones be presented to the same space as the visual stimuli in order to induce temporal fission (as would be the case if the gradient account was the sole driver for the effect). We performed t-tests below, and calculated Bayes Factors, with the view to clarifying this. Figure 3 contains the relevant plots for the data used in the means comparisons.

There was a reduction in report bias corresponding to the actual presentation of visual stimuli (reporting simultaneity) in the simultaneous visual condition when 2 tones were presented to analogous central space compared to baseline (no tone) $t(26) = 2.17$, $p = .039$, $d = 0.59$, $SE = 0.04$. The $BF = 1.5$, which provides anecdotal evidence indicating the presence of the temporal fission illusion in the 2 central tones condition, which tenuously supports an impletion account of temporal fission where tones are not required to share the same space as the visual stimuli.

Spatially opposing tones were significantly more likely to result in report bias that did not correspond to the actual presentation of visual stimuli in the simultaneous visual condition when compared to 2 tones presented in analogous central space, $t(26) = 9.61$, $p < .001$, $d = 2.62$, $SE = 0.06$. The $BF = 2.25e+07$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides extreme evidence indicating the presence of a stronger temporal fission illusion in the collapsed spatially opposing tones condition, which supports both impletion and the gradient account as elaborated on in the discussion below.

We conducted a repeated measures ANOVA on report bias corresponding to the actual presentation of visual stimuli to determine if the number of tones, relative to visual stimuli (which always consisted of 2 circles, although they differed in presentation: sequential vs. simultaneous), had an effect on report bias in Experiment 1. The ANOVA was a 2 (visual condition: simultaneous visual condition vs. collapsed sequential visual conditions) × 3 (number of tones: 1 tone presented to analogous central space; 2 tones presented to analogous central space; and no tones presented to any space) design. The auditory conditions used in this analysis were chosen due to their contrasting number of presentations, while all auditory stimuli shared the same analogous presentation space (analogous central space which was neutral relative to the visual stimuli locations).

There was a significant main effect of visual condition $F(1, 26) = 11.93$, $MSE = 0.09$, $p = .002$, $\eta^2_G = .123$. There was a significant main effect of the number of tones $F(2, 52) = 9.63$, $MSE = 0.01$, $p < .001$, $\eta^2_G = .019$. There was a significant interaction of visual condition and number of tones $F(2, 52) = 49.25$, $MSE = 0.02$, $p < .001$, $\eta^2_G = .225$.

The ANOVA above shows that the number of tones presented is important when inducing visual temporal effects. However, it does not make clear if it is a requisite that the number of auditory tones should match the number of visual stimuli in order to induce said temporal effects (as would be the case if Morein-Zamir et al.'s (2003) account is accurate). We performed t-tests below, and calculated Bayes Factors, with the view of clarifying this. Figures 3 and 4 contain most of the plots for the data used in the means comparisons. More nuanced increase in report bias corresponding to the actual presentation order of the visual stimuli data in sequential visual conditions with collapsed spatially opposing tones is contained in Figure 2.

One central tone accompanying collapsed sequential visual conditions reduced report bias corresponding to the actual presentation order of visual stimuli when compared to baseline (no tone), $t(26) = 6.08$, $p < .001$, $d = 1.66$, $SE = 0.03$. The $BF = 9393.87$, which provides extreme evidence indicating the presence of a temporal fusion illusion in the 1 central tone condition, which is consistent with Getzmann's (2007) finding that 1 tone was sufficient to induce temporal ventriloquism-like effects in report bias.

One tone resulted in an increase in report bias matching the actual presentation of visual stimuli when compared to baseline (no tone) in the simultaneous visual condition $t(26) = 5.83$, $p < .001$, $d = 1.59$, $SE = 0.02$. The $BF = 5231.26$, which provides extreme evidence indicating the presence of increased report bias corresponding to the actual presentation of visual stimuli in the 1 central tone condition, again supporting Getzmann's (2007) findings.

Left or Right Circle First Report Probability Analyses. Due to the use of a ternary-response task, this allowed us to examine with greater resolution whether tones could induce responses consistent with the gradient account, and indeed examine whether auditory cues to either ear resulted in a left or right circle first report bias when sequential presentation of stimuli was reported. Figure 2 shows each report category in all conditions which should be referenced for analyses below.

We conducted a repeated measures ANOVA on left or right circle first report probability to determine if the first tone in the spatially opposing tones conditions had an effect on probability of report in Experiment 1. The ANOVA was a 2 (first tone presentation location: a left tone first (followed by a right tone) vs. a right tone first (followed by a left tone)) × 3 (visual condition: 1 circle left of fixation followed by 1 circle right of fixation vs. 1 circle right of fixation followed by 1 circle left of fixation vs. both circles simultaneously) × 2 (response made: left circle first vs. right circle first) design.

Mauchly's test for Sphericity failed for visual condition $W = .126$, $p < .001$, and for the interaction of visual condition and response type $W = .343$, $p < .001$. Therefore, the degrees of freedom were corrected using Greenhouse-Geisser Estimate ϵ (Greenhouse & Geisser, 1959).

There was no significant main effect of first tone presentation location, $F(1, 26) = 3.08$, $MSE = 0.002$, $p = .09$, $\eta^2_G < .001$. There was a significant main effect of visual condition, $F(1.07, 27.76) = 11.49$, $MSE = 0.041$, $p < .001$, $\eta^2_G = .043$. There was no significant main effect of response type made, $F(1, 26) = .26$, $MSE = 0.076$, $p = .616$, $\eta^2_G = .002$. There was no significant interaction between first tone presentation location and visual condition, $F(2, 52) = .32$, $MSE = 0.002$, $p = .725$, $\eta^2_G < .001$. There was a significant interaction between first tone presentation location and response made, $F(1, 26) = 180.16$, $MSE = 0.089$, $p < .001$, $\eta^2_G = .599$. There was a significant interaction between visual condition and response made, $F(1.21, 31.38) = 149.97$, $MSE = 0.114$, $p < .001$, $\eta^2_G = .459$. There was a significant interaction between first tone presentation location, visual condition, and response made, $F(2, 52) = 9.77$, $MSE = 0.019$, $p < .001$, $\eta^2_G = .034$.

The above ANOVA demonstrates that the first tone presentation location had an effect on probability of report made when interacting with visual condition. The following t-tests examine if there was a bias in report in the temporal fission illusion specifically, in line with what would be expected for the gradient account.

When a left tone occurred before a right tone in the simultaneous visual condition, participants made more left-first reports than right-first reports, $t(26) = 11.33$, $p < .001$, $d = 3.08$, $SE = 0.05$. The $BF = 2.14e + 09$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides extreme evidence that report bias favoured the side of space the first tone was presented to, which in turn supports the gradient account.

When a right tone occurred before a left tone in the simultaneous visual condition, participants made more right-first reports than left-first reports, $t(26) = 7.31$, $p < .001$, $d = 1.99$, $SE = 0.07$. The $BF = 1.62e + 06$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides extreme evidence that response bias favoured the side of space the first tone was presented to, which in turn supports the gradient account.

Discussion

Figure 3 shows the probability of reporting the presentation of both circle stimuli as being simultaneous in the simultaneous visual condition. Collapsed spatially opposing tones induced a temporal fission illusion. This effect is pronounced, which appears to support the gradient account in that attentional focus was drawn to one side of space before visual onset, thus the corresponding circle was processed first and the second tone drew attention quickly to the next circle in turn, which served to process it second.

Interestingly, 2 central tones (1 before visual onset and 1 after) often induced temporal fission when visual stimuli were simultaneously presented, although the evidence supporting this statistically is relatively weak. Presenting two central tones in the simultaneous visual condition appeared to 'pull' the visual stimuli apart in temporal perception. This, arguably, directly contradicts the findings of Getzmann (2007) in similar conditions. For example, when Getzmann (2007) presented 2 clicks (1 before simultaneous visual onset and 1 after) it did not increase reporting of apparent motion, which might be expected if simultaneously presented visual stimuli were teased apart in temporal perception. However, this is difficult to state with any certainty in the absence of 'successive' presentation and 'broken motion' reports for this auditory condition, especially as the report of 'successive' presentation of the squares would be a closer match to this TOJ finding. Bear in mind, participants were not making TOJs in Getzmann's (2007) research and the results focused on the reported presence/absence of apparent motion, disregarding other reports. However, Getzmann (2007) drew analogies between apparent motion findings and Morein-Zamir et al.'s (2003) research.

Two central tones inducing temporal fission has, to the best of our knowledge, not been demonstrated before. The gradient account does not easily explain this finding, since neither of the 2 tones corresponded in analogous space to that of the visual stimuli, yet sequential order was often perceived. The classic account of temporal ventriloquism is somewhat supported in that the circles appear to have been 'pulled' in time towards the tones, thus inducing an increased SOA perceptually. The presence of temporal fission induced by 2 static tones casts doubt on any suspicion that spatially opposing tones, by merely being directional in-and-of-themselves (due to their presentation to the left and right ears, or vice versa), may bias participants to make a directional response.

Impletion, and the expanded account of temporal ventriloquism (where featural characteristics of auditory stimuli are taken into account on the way to integration), also lends explanatory power to this finding; namely that the auditory and visual stimuli may have been deemed related and the fact that the auditory tones were clearly sequential may have influenced visual perception at the audio-visual integration stage.

Another point of interest here is the fact that 1 tone presented to analogous central space before sequential visual conditions onset often resulted in temporal fusion, as shown in Figure 4. According to Morein-Zamir et al.'s (2003) and the classic temporal ventriloquism account, this should not happen. A tone presented before a circle should ‘pull’ that circle in time towards the tone. This should result in report bias towards the actual sequential order of visual stimuli but, as reported, quite the opposite was found. However, this temporal fusion effect may have been present in the classic Morein-Zamir et al. (2003) experiment but the binary response approach may not have been sensitive enough a measure to detect it. Since participants could only respond ‘top’ or ‘bottom’, the effect may not have been strong enough to reverse the perception of the sequential order. It may have been strong enough, as was found here, to introduce sufficient ambiguity that the difference between the TOJ corresponding to the actual presentation order and an SJ were reduced to the point of non-discrimination. It is worth noting here also that Getzmann (2007) demonstrated that a single tone presented between visual stimuli presentation in time tended to induce a stronger perception of apparent motion, which suggests again a ‘pulling’ in time process that Morein-Zamir et al. (2003) discounted as being possible.

However, a striking difference between Getzmann’s (2007) findings and those here was the temporal placement of the single tone. Getzmann (2007) presented the single tone between the onsets of both visual stimuli (after the first, and before the second visual stimulus), whereas we presented the single tone prior to any visual onset. The results presented here suggest that the single tone prior to visual onset did not ‘pull’ either of the visual stimuli towards it in perceptual time as there was no observed increase in report bias towards the actual sequential order of visual stimuli, as would be expected. Instead, the placement of the tone prior to visual stimuli onset introduced sufficient ambiguity so as to render little difference in the likelihood of perceiving sequential presentation that corresponded to the actual presentation order of visual stimuli or simultaneous presentation of visual stimuli. Conceivably, it may be possible that the second circle was ‘pulled’ further in perceptual time towards the tone than the first circle, but it is difficult to explain why this would be the case.

The reported temporal fusion effect is not consistent with the fundamental claims made by Morein-Zamir et al. (2003) in terms of temporal ventriloquism. Taken with the findings of Getzmann (2007) this suggests that response type, and options, may play a role in how sensitive a measure is at capturing the influence of auditory stimuli on visual events succinctly. Indeed, Getzmann (2007) also demonstrated that there was no reversal of perceived apparent motion when a single tone was presented to analogous central space thus suggesting the effect is not strong enough to reverse the perceived order of sequential presentation.

In addition to this point on measurement sensitivity, had an SJ not been included as an option, the temporal fission effect found with two centrally presented tones would have gone undetected due to there inherently being no left or right spatial bias in report (as shown in Figure 2).

It is worth noting that in this experiment, and the following two experiments, due to the customised timings acquired in the staircase, it would be expected that observed effects would vary between participants with shorter SOAs between stimuli than those with longer. This in turn renders the individual data points contained in the reported figures of limited use.

Conditions analogous to those used in Morein-Zamir et al.'s (2003) research have yet to be examined in this paradigm: namely, vertical presentation of visual stimuli where a top circle is followed by a bottom circle, or a bottom circle is followed by a top circle. In Experiment 2, these visual conditions were replicated with the inclusion of an SJ response option for simultaneously presented top and bottom circles. By adopting a full orthogonal approach similar to that of Spence et al. (2001), this helped rule out any bias in response that may have been induced via auditory stimuli cueing the analogous space where the visual stimuli were presented to. This approach completely removes any spatially congruent audio-visual information.

Experiment 2

Methods

Participants. Twenty-five participants, 10 male and 15 female (mean age 21.96, SD=3.24), with normal or corrected-to-normal vision, and self-reported normal hearing participated. All were students from Swansea University. All participants were naïve to the purposes of the study. Ethical approval was received from the Department of Psychology Ethics Committee for this research.

An a priori power analysis was applied using the data collected in Experiment 1. An identical condition to that used in this design (with the exception of vertical presentation of visual stimuli as detailed below) displayed an effect size of $d = 1.66$ when comparing differences in the means of report bias corresponding to the actual presentation order of visual stimuli between 1 central tone and baseline (no tones) in the collapsed sequential visual conditions ($t(26) = 6.08$, $p < .001$, $d = 1.66$, $SE = 0.03$, where $BF = 9393.87$ which provides extreme evidence indicating the presence of a temporal fusion illusion – – see the Results section in Experiment 1 for notes on how the BF was computed). This condition was first used in Experiment 1 and existed explicitly to detect whether temporal fusion was present (an effect not previously detected in this type of paradigm to the best of our knowledge), and is therefore one of the most important effects under consideration. Using GPower (Faul et al., 2007) with 95% power and $\alpha = .001$ (consistent with the reported p-value from Experiment 1) in a difference between two dependent means (matched pairs) power analysis, the recommended sample size was 22 for an actual power estimate of 95.12%. The sample size used here was deliberately larger due to concerns about baseline performance as outlined in Experiment 1.

Apparatus. The apparatus were the same as Experiment 1. The CRT and response box were rotated 90° and text instructions, as well as feedback, etc. were rotated also. This was to ensure identical temporal accuracy as Experiment 1.

Table 4
Visual stimuli arranged by presentation category.

Sequential Visual Conditions	Simultaneous Visual Condition
1 circle above fixation followed by 1 circle below fixation	Both circles simultaneously
1 circle below fixation followed by 1 circle above fixation	

Stimuli and Procedure. The auditory and visual stimuli, and procedure, were identical to Experiment 1, with the exception that instead of allowing a left circle first, right circle first and an SJ, participants were asked to report if they perceived; a top circle being presented first; a bottom circle being presented first; or if both were presented simultaneously. Tables 4 and 5 list the visual and auditory conditions by presentation category, and spatial category respectively. These categories will be referenced often below, in the results and discussion sections.

The same exclusion criteria used in Experiment 1 resulted in the removal of three participants. These participants summed with the removal of trials that were below or above the RT criteria saw the total removal of 2526 observations (13.47% of trials) from Experiment 2.

The same transformation was applied to the data for null hypothesis testing as was used in Experiment 1. The same approach was used when calculating the *BF* as Experiment 1. We also created subgroups of the data in a similar fashion to those in Experiment 1 for purposes of analysis.

Table 5
Auditory stimuli arranged by spatial category.

Centrally Presented Neutral Tone/s	Neutral Spatially Opposing Tones	Control
2 tones presented to analogous central space	1 tone in left space followed by 1 tone in right space	No tones
1 tone presented to analogous central space	1 tone in right space followed by 1 tone in left space	

Results

Analysis of Report Bias Corresponding to Actual Presentation of Visual Stimuli. We tested whether the temporal fission effects reported in Experiment 1 were present here despite tones never being presented to the same space as the visual stimuli. We conducted a 1 (visual condition: simultaneous visual condition) $\times$ 3 (auditory condition: 2 tones presented to analogous central space vs. collapsed neutral spatially opposing tones vs. no tone) repeated measures ANOVA on report bias corresponding to the actual presentation of visual stimuli.

Mauchly's test indicated that the assumption of sphericity had been violated for the auditory conditions, $W = .730$, $p = .043$. Therefore, the degrees of freedom were corrected using Greenhouse-Geisser Estimate ϵ (Greenhouse & Geisser, 1959).

There was a significant main effect of auditory condition, $F(1.58, 33.08) = 36.16$, $MSE = 0.034$, $p < .001$, $\eta^2_G = .376$. A series of t-tests were run to establish the direction of the effects.

Figure 6 shows that there was a reduction in report bias corresponding to the actual presentation of visual stimuli (reporting simultaneity) in the simultaneous visual condition in the collapsed neutral spatially opposing tones condition compared to baseline (no tone), $t(21) = 7.90$, $p < .001$, $d = 2.38$, $SE = 0.06$. The $BF = 4.98e + 09$ (adjusted using the "evidence updating" method (Ly et

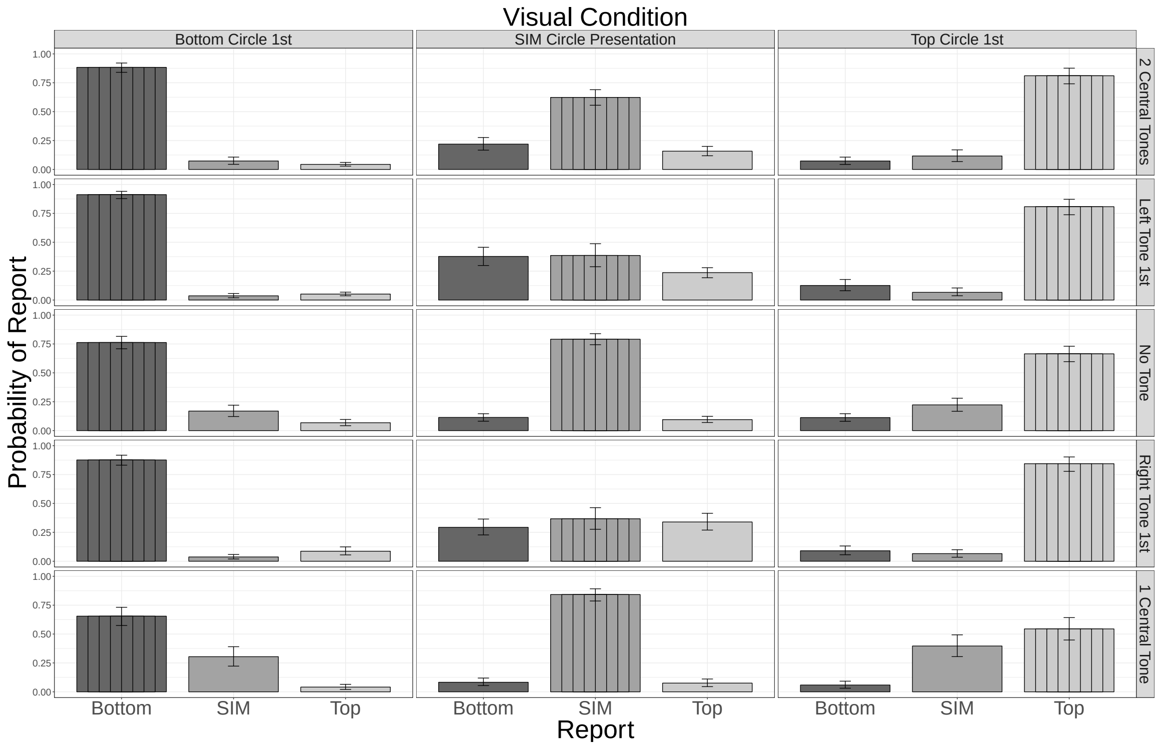

Figure 5. Experiment 2 report probability: The 3 visual conditions are labelled at the top of the grid horizontally. The leftmost column denotes sequential presentation of circles, where the first circle was presented below fixation (bottom space). The rightmost column denotes sequential presentation of circles, where the first circle was presented above fixation (top space). The central column denotes simultaneous presentation of circles, where a circle was presented to above and below fixation (top and bottom space) simultaneously. The 5 auditory conditions are labelled vertically on the rightmost edge of the grid, denoting (from top-to-bottom) the presentation of: 2 tones in analogous central space; a tone presented to the left ear followed by a tone presented to the right ear; no tones; a tone presented to the right ear followed by a tone presented to the left ear; 1 tone in analogous central space respectively. Error bars are bootstrapped within-subject 95% confidence intervals. Reports are labelled on the x-axis with reports corresponding to the actual presentation of visual stimuli highlighted with vertical hatching.

al., 2017)), which provides extreme evidence indicating the presence of the temporal fission illusion in the collapsed neutral spatially opposing tones condition. This replicates the temporal fission illusion despite the neutral spatially opposing tones not being presented to the same space as the visual stimuli. This provides strong evidence for an impletion account of temporal fission.

There was a reduction in report bias corresponding to the actual presentation of visual stimuli

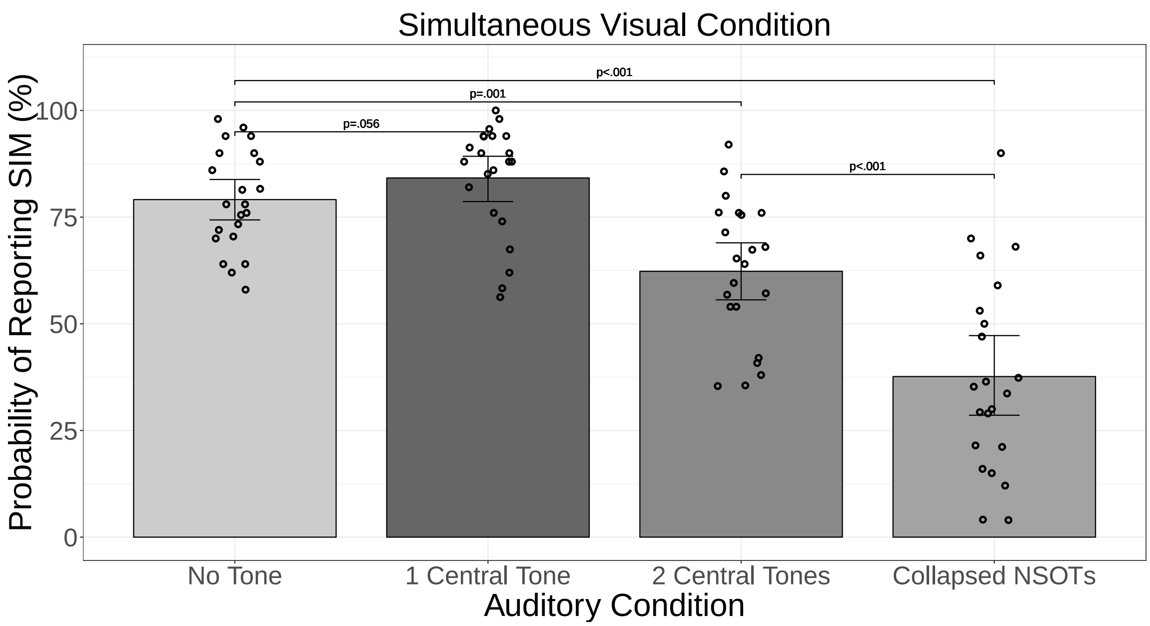

Figure 6. The probability of reporting simultaneous presentation when visual stimuli were presented simultaneously in Experiment 2: The probability (%) of reporting that visual stimuli were presented simultaneously is plotted on the y-axis and the auditory stimuli are labelled on the x-axis where the NSOTs (neutral spatially opposing tones) conditions have been collapsed. Error bars are bootstrapped within-subject 95% confidence intervals.

(reporting simultaneity) in the simultaneous visual condition when 2 tones were presented to analogous central space compared to baseline (no tone), $t(21) = 4.07$, $p = .001$, $d = 1.23$, $SE = 0.05$. The $BF = 1.80e + 02$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides extreme evidence indicating the presence of the temporal fission illusion in the 2 central tones condition, which replicates the findings in Experiment 1, providing further evidence of an impletion account of temporal fission.

Neutral spatially opposing tones were significantly more likely to result in report bias that did not correspond to the actual presentation of visual stimuli in the simultaneous visual condition when compared to 2 tones presented to analogous central space, $t(21) = 7.36$, $p < .001$, $d = 2.22$, $SE = 0.04$. The $BF = 5.43e + 03$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides extreme evidence indicating the presence of a stronger temporal fission illusion in

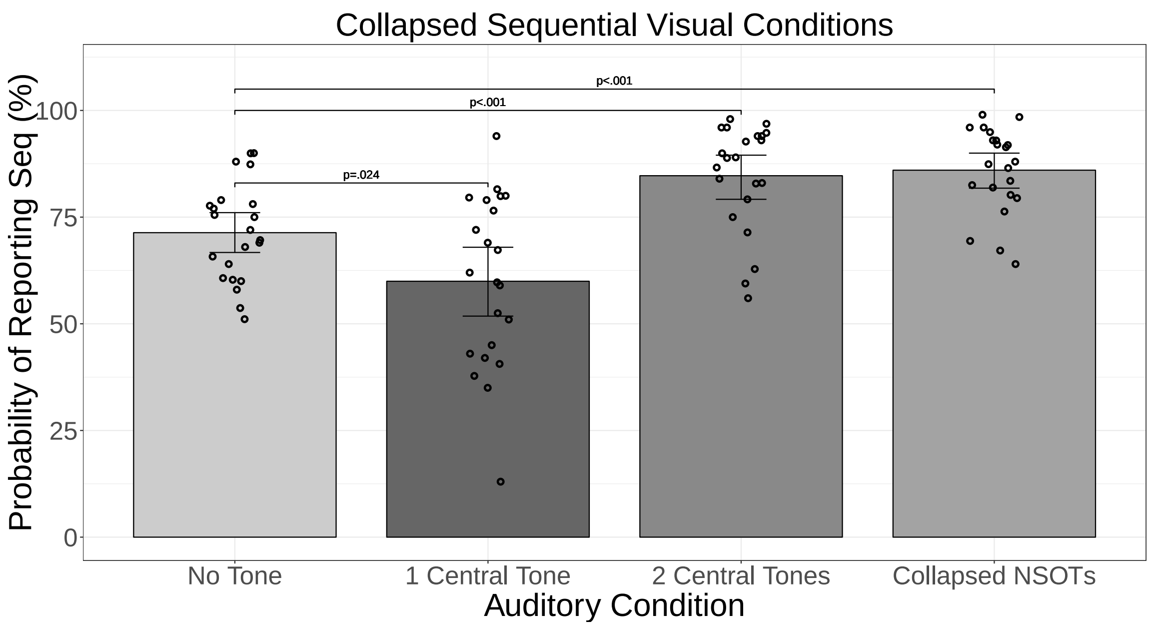

Figure 7. The probability of reporting sequential presentation (bottom circle first reports and top circle first reports that corresponded to the actual presentation order of visual stimuli collapsed) when visual stimuli were presented sequentially (bottom circle first and top circle first conditions collapsed) in Experiment 2: The probability (%) of reporting sequential order of visual stimuli corresponding to the actual presentation order of visual stimuli is plotted on the y-axis and the auditory stimuli are labelled on the x-axis where the NSOTs (neutral spatially opposing tones) conditions have been collapsed. The reported *p*-values were obtained via null-hypothesis t-tests. Error bars are bootstrapped within-subject 95% confidence intervals.

the collapsed neutral spatially opposing tones conditions.

Secondly, we tested whether classic temporal ventriloquism-like effects (specifically increased report bias corresponding to the actual presentation order of visual stimuli in collapsed sequential visual conditions via 2 tones) was replicated. We conducted a 1 (visual condition: collapsed sequential visual conditions) $\times$ 3 (auditory condition: 2 tones presented to analogous central space vs. collapsed neutral spatially opposing tones vs. no tone) repeated measures ANOVA on report bias corresponding to the actual presentation order of visual stimuli..

NOTE: The probability of report data in Figure 7 show reports corresponding to the actual sequential presentation order of a given sequential visual presentation. Report biases of sequential order opposite to the actual order were not included in the plot. For example, if the sequential visual condition was ‘bottom circle first’ we only included reports of ’bottom circle first’. The equivalent was true for the ‘top circle first’ visual condition.

Mauchly’s test indicated that the assumption of sphericity had been violated for auditory condition, $W = .418$, $p < .001$. Therefore, the degrees of freedom were corrected using Greenhouse-Geisser Estimate ϵ (Greenhouse & Geisser, 1959).

There was a significant main effect of auditory condition $F(1.26, 26.54) = 17.36$, $MSE = 0.016$, $p < .001$, $\eta^2_G = .168$. A series of t-tests was run to establish the direction of the effects. The relevant report bias data is contained in Figure 7.

When 2 tones were presented in analogous central space during collapsed sequential visual conditions, there was increase in report bias corresponding to the actual presentation order of visual stimuli (see Figure 5 for illustration of this increase in probability of report bias corresponding to the presentation order of visual stimuli) observed when compared to baseline (no tone), $t(21) = 4.63$, $p < .001$, $d = 1.40$, $SE = 0.04$. The $BF = 6.42e + 02$ (adjusted using the “evidence updating” method (Ly et al., 2017)), which provides extreme evidence indicating the presence of increased report bias that corresponded to the actual presentation order of visual stimuli in the 2 central tones condition, which is consistent with the classic temporal ventriloquism effect.

When neutral spatially opposing tones were presented with collapsed sequential visual conditions, an increase in report bias corresponding to the actual presentation order of visual stimuli (see Figure 5 for illustration of this increase in probability of report bias corresponding to the presentation order of visual stimuli) was observed overall when compared to baseline (no tone), $t(21) = 6.86$, $p < .001$, $d = 2.07$, $SE = 0.03$. The $BF = 3.03e + 08$ (adjusted using the “evidence updating” method (Ly et al., 2017)), which provides extreme evidence indicating increased report bias that corresponded to the actual presentation order of visual stimuli in the collapsed neutral spatially opposing tones conditions, which is consistent the classic temporal ventriloquism effect.

We investigated if the number of auditory stimuli was required to match the number of visual stimuli in order to induce audio-visual effects by conducting a 2 (visual condition: simultaneous visual condition vs. collapsed sequential visual conditions) × 3 (number of tones: 1 tone presented to analogous central space; 2 tones presented to analogous central space; and no tones presented to any space) repeated measures ANOVA on report bias corresponding to the actual presentation order of visual stimuli. The auditory conditions used in this analysis were chosen due to their contrasting number of presentations, while all auditory stimuli shared the same presentation space (analogous central space which was neutral relative to the visual stimuli locations and also always the same space, unlike neutral spatially opposing tones).

Mauchly's test indicated that the assumption of sphericity had been violated for the interaction between visual condition and number of tones, $W = .666$, $p = .017$. Therefore, the degrees of freedom were corrected using Greenhouse-Geisser Estimate ϵ (Greenhouse & Geisser, 1959).

There was no significant main effect of visual condition, $F(1, 21) = 1.22$, $MSE = 0.033$, $p = .282$, $\eta^2_G = .010$. There was no significant main effect of the number of tones, $F(2, 42) = 0.92$, $MSE = 0.007$, $p = .404$, $\eta^2_G = .003$. There was a significant interaction of visual condition and number of tones, $F(1.5, 31.5) = 15.64$, $MSE = 0.062$, $p < .001$, $\eta^2_G = .245$.

The ANOVA showed no main effect of the number of tones presented, however, there was a significant interaction with the visual conditions, consistent with arguments made in the discussion section of Experiment 1. However, it does not make clear if it is a requisite that the number of auditory tones should match the number of visual stimuli in order to induce temporal effects (as would be the case if Morein-Zamir et al.'s (2003) account is accurate). We performed t-tests below, and calculated Bayes Factors, with the view to clarifying this. Figures 6 and 7 contain the plots for the data used in the following means comparisons.

Despite a slight increase in report bias towards the actual presentation order of visual stimuli, there was no statistical difference in the 1 central tone condition in reporting simultaneity relative to the baseline (no tone), $t(21) = 2.15$, $p = .056$, $d = 0.65$, $SE = 0.04$. The $BF = 6.56e - 01$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides anecdotal evidence for the null hypothesis in the single tone condition.

When 1 central tone was presented during collapsed sequential presentation of circles, there was a significant difference in report bias that corresponded to the actual presentation order of visual stimuli when compared to baseline (no tone), $t(21) = 2.62$, $p = .024$, $d = 0.79$, $SE = 0.05$. The $BF = 3.76e + 00$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides moderate evidence for a temporal fusion effect in the 1 central tone condition.

Bottom or Top Circle First Report Probability Analyses. No ANOVA was conducted here since no audio and visual stimuli shared space (unlike Experiment 1). However, a visual inspection of the report probability data and relevant confidence intervals (see Figure 5) warranted an examination of the reports made in the simultaneous visual condition for any statistical indication of bottom or top circle first report bias.

When a left tone occurred before a right tone with simultaneous circle presentations, participants made slightly more bottom-first reports than top-first reports, $t(21) = 3.08$, $p = .011$, $d = 0.93$, $SE = 0.05$. The $BF = 8.05e + 00$, which provides moderate evidence indicating a bias in report favouring bottom circle first in the left tone first condition.

When a right tone occurred before a left tone with simultaneous circle presentations, there was no statistically significant difference in the proportion of bottom first or top first reports

$t(21) = .84$, $p = .411$, $d = 0.25$, $SE = 0.06$. The $BF = 3.05e - 01$, which provides moderate evidence for no report bias in the right tone first condition.

Discussion. The main results from Experiment 1 were replicated here.

A note on the apparent bias in response when a left or right tone was presented first in the simultaneous visual condition: When the first tone was in left analogous space participants tended to report perceiving the bottom circle first more often than the top circle first. This initially appears to be a counter-intuitive finding as the auditory and visual stimuli are not in matching analogous space. However, if one considers the “orthogonal Simon effect”, where participants tend to have lower response times when a left key is matched to a lower location in space and a right key is matched to higher location in space (Lu & Proctor, 1995), we are arguably seeing an analogous effect here. When a tone was presented first in left analogous space, participants were more inclined to choose the circle in lower space as being presented first, and when a tone was presented first in right analogous space, participants were slightly more inclined to choose the circle in the higher location as being presented first (however not at statistically significant levels in that case).

As can be seen in Figure 6, temporal fission was strongest in Experiment 2 when tones were presented in neutral opposing space on the x-axis. This is difficult to explain via the gradient account of speeded visual processing as the auditory tones are never presented in the same analogous space as the visual stimuli and therefore attention is never drawn to them; instead, attention is shifted away from both visual stimuli.

Effects consistent with those described by Morein-Zamir et al. (2003) regarding enhancement in TOJs were observed with two central tones, as well as tones in neutral opposing space on the x-axis (as shown in Figure 5), where there was an increase in report bias towards the actual presentation order of visual stimuli. However, again a single tone presented before sequential presentation of circles often resulted in temporal fusion which does not fit with the classic temporal ventriloquism account.

It might be supposed that neutral spatially opposing tones on the x-axis should have the same effect as two tones presented to analogous central space as neither condition provides any spatially relevant information about the visual stimuli. However, this is not the case, and neutral spatially opposing tones induce a stronger temporal fission illusion.

The explanation for this is not easily provided by the temporal ventriloquism, gradient, or impletion accounts. There may be a more general role of attention here. When participants' attentional focus is drawn onto and shifted across the x-axis while the visual task is presented on the y-axis, this may result in reduced temporal salience of visual stimuli. However, reduced temporal salience of visual stimuli via auditory stimuli is not a blanket explanation for all observed effects in these studies when the effects that show an increase in report bias towards the actual presentation order of visual stimuli are taken into consideration. It is also possible that two centrally presented tones may induce a small habituation effect due to rapidly repeated stimulation of the same neurons, but when tones are presented in opposing space separate neurons are activated, avoiding habituation. If this explanation was accepted, it could conceivably fit with the impletion account best as the observed effects arguably are the result of weighted evidence. Additionally, motion processing may play an important role here. The tones in opposing space are likely perceived as apparent motion stimuli, whereas the centrally presented tones are 'static' in terms of spatial location. This would potentially align with the gradient account in terms of apparent auditory motion activating shared audio-visual motion processing, thus increasing the likelihood of perceiving visual motion.

The effect size for the main effect of auditory condition in the report bias consistent with temporal ventriloquism was slightly larger for Experiment 1 than Experiment 2, however this difference is arguably negligible which would reflect the reality that the auditory stimuli examined in both experiments never shared analogous space with visual stimuli. The effect size for the interaction of visual stimuli and the number of auditory stimuli were similar and the small difference between Experiment 1 (which had a marginally smaller effect size) and 2 was negligible, again consistent with spatially neutral auditory stimuli. The BFs for temporal fission via spatially opposing tones for Experiments 1 and 2 both provided extreme evidence for the effect, thus supporting the notion that spatial congruency between auditory and visual stimuli is not necessary to induce temporal fission. Although, it is worth noting that the BF for Experiment 1 is considerably larger than that of Experiment 2. This is likely due to spatially opposing tones in Experiment 1 sharing analogous space with the visual stimuli (which was not the case in Experiment 2) and therefore prior entry, and/or spatial report bias, likely bolstered the effect.

Experiment 3 was conducted to address a non-orthogonal concern present in Experiments 1 and 2. Namely, that the three button responses were oriented consistently with regard to the centrally presented tone auditory condition that induced temporal fusion. The report for simultaneous circle presentation was always the middle button regardless of visual axis orientation. There was a small chance that a centrally presented tone may increase the chance of choosing a centrally positioned button, thus producing an effect not based on visual perception. The ternary-response task used in Experiments 1 and 2 was replaced with a simultaneity-judgment task, where participants either reported ‘sequential’ presentation of circles, or ‘simultaneous’ presentation of circles. The assigned value for each button response was counterbalanced. This approach also addressed further concerns surrounding the use of a ternary-response paradigm and varying criteria for simultaneity. Results for Experiment 3 can be found in Appendix A. Relevant effects reported in Experiment 1 and replicated in Experiment 2 were again replicated in Experiment 3.

General Discussion

The three behavioural experiments described demonstrate that auditory stimuli can influence temporal perception of visual events. The observed temporal order judgment (TOJ) report biases that favoured, and those that opposed, the actual presentation order of visual stimuli were consistent with effects reported in the literature. This supports a version of temporal ventriloquism that suggests auditory stimuli must match visual stimuli in quantity to induce such effects (Morein-Zamir et al., 2003). However, the finding of temporal fusion when a single tone is presented before sequential circles forces a more nuanced definition of temporal ventriloquism-like effects. The classic account of temporal ventriloquism may not have used a sensitive enough measure to detect the influence of a single tone between sequential circles. Had a simultaneity judgment (SJ) response been included in Morein-Zamir et al.'s (2003) experiments, it may have revealed the temporal fusion found in our experiments. The effect may not have been strong enough to reverse the direction of perceived sequential order, but including an SJ response afforded participants the opportunity to report their perception beyond a forced sequential order task. In this instance, it would appear that a single tone can induce a temporal fusion illusion that otherwise would have gone unreported. This finding is consistent with Getzmann's (2007) research which demonstrated that the number of auditory and visual stimuli need not be equal to facilitate temporal ventriloquism-like effects. Additionally, the temporal placement of the tone relative to the visual stimuli defies the classic temporal ventriloquism notion of auditory stimuli 'pulling' visual stimuli towards them in temporal perception.

The gradient account of speeded visual processing, akin to prior entry, is somewhat supported in the findings here (additional evidence of prior entry via auditory stimuli cueing incongruent space to that of the sequence of visual presentation is included in Appendix B). However, this account falls short when considering that 2 tones presented to analogous central (spatially neutral) space can induce the temporal fission illusion; as can neutral spatially opposing tones, suggesting a potential role of auditory apparent-motion, as demonstrated in Experiment 2. Neither of these auditory conditions inherently draw attention to a circle and yet the illusion persists.

The expanded account of temporal ventriloquism (Keetels et al., 2007; Roseboom et al., 2013b) suggests that stimuli generally have to be featurally similar in order to induce effects associated with temporal ventriloquism (at least at the times scales used in the research presented here). We conducted a pilot experiment that abolished temporal fission (and temporal ventriloquism-like effects) when the 2 auditory stimuli presented centrally were not featurally similar to each other, i.e. a sinewave tone and a white noise burst, as shown in Figure 13 in Appendix D. However, in the same experiment, temporal fission was preserved when the sinewave and white noise burst tones were presented in congruent space to that of the visual stimuli. Taking this into account, it seems reasonable to presume that featural similarity does have a role to play in audio-driven visual temporal perception, especially in the absence of other spatially congruent information.

Impletion, as described by Downing and Treisman (1997), taken in conjunction with elements of the gradient and temporal ventriloquism accounts, appears to be the most reasonable explanation for the effects described here and elsewhere. While the asynchronous auditory and visual stimuli are undoubtedly important factors in manipulating report bias, and there does appear to be cue-induced speeded processing a la the gradient account/prior entry, we suggest that these elements are taken together to create the most likely real-world representation of events in perception. This would explain why 2 tones can induce the illusion of 2 temporally sequential events when circles are presented simultaneously. This also helps explain why 1 tone can induce the illusion of a single temporal event when circles are sequentially presented. However, when auditory stimuli are spatially congruent to visual stimuli, prior entry evidence appears to carry greater weight than spatially incongruent stimuli when arriving at a perception of temporal events (as evidenced by spatial report biases in Experiment 1, and a considerably larger *BF* in Experiment 1 than 2 for spatially opposing tones induced temporal fission). When one considers the temporal fission illusion induced via 2 tones presented to analogous central (neutral) space, it becomes clear that, in an attempt to integrate audio-visual events, an average of sorts is arrived at. In the absence of spatially congruent audio-visual information, evidence is approximately equal for each circle being presented first; hence no spatial bias in response is observed in this temporal fission effect.

It would appear that audio-visual temporal perception uses a process that combines various sources of information such as relative spatial positioning, and indeed how many individual stimuli exist in a given time window. In support of this view are the results presented in Figure 12 in Appendix C. Specifically, when comparing a single auditory tone before visual stimuli onset to a single auditory tone after visual onset in the simultaneous visual presentation condition, an interesting pattern emerges. The former condition results in the prior entry spatial report bias associated with the gradient account; however, despite no statistical difference in report bias favouring the actual presentation order of visual stimuli, the latter condition shows no *spatial* bias in reports opposing the actual presentation order of the visual stimuli. What this tells us is that the auditory tone presented after visual onset (aside from clearly not being a prior entry effect) provides equal amounts of evidence for either circle being presented first (when illusory order is perceived). The tone is always perceived as being associated with the onset of the circle it shares analogous space with, but that perceived visual onset could conceivably be the circle being presented first, or second, in the visual stimuli sequence. In each scenario, the tone is perceived as the onset of the circle sharing the analogous space, but there is no evidence (or put another way, there is equal evidence) afforded to the perceiver as to which order the circles are presented in the (illusory) sequence. This results in no spatial bias in the illusory temporal order being observed.

As discussed previously, Körding et al. (2007) made a compelling case for Bayesian causal inference in multisensory perception. The model supports the idea that an 'ideal-observer' arrives at an inferred estimate of a given scenario (in their example whether auditory and visual stimuli originate from the same causal event) via the combination of the likelihood of a stimulus originating from a specific spatial location, and prior knowledge of similar scenarios informing what the real-world likelihood is of both stimuli (in an audio-visual localisation task) originating from the same source, or different sources. Viewing the reported findings here through the lens of causal inference may help explain the observed effects (Körding et al., 2007). For example, temporal fission was strongest when the auditory and visual stimuli shared analogous space (the spatially opposing tones temporal fission condition in Experiment 1). This form of temporal fission persisted even when the auditory stimuli were featurally distinct from each other (see Experiment 4(ii)). As Körding et al. (2007) demonstrated, the spatial relationships between auditory and visual stimuli factor into the perception of where in space the stimuli are presented. It is conceivable that the spatial, temporal, and featural relationship between auditory and visual stimuli are factored in (via causal inference) when arriving at the perceptions reported here (Wallace et al., 2004; Shams et al., 2005; Shams & Beierholm, 2010; Sato et al., 2007; Roseboom et al., 2013b, 2013a). The sequential nature of the auditory stimuli in the temporal fission conditions likely provided increased likelihood, informed by prior knowledge, that the visual stimuli were also sequentially presented. Therefore, so long as participants perceived that auditory and visual stimuli shared analogous space – and in turn, increased likelihood of sharing the same source – we might assume a strong temporal fission illusion, regardless of how featurally similar the auditory stimuli were to each other. However, the weaker form of temporal fission (where the auditory and visual stimuli are not perceived to have shared analogous space) is consistent with the idea that there was reduced likelihood of the auditory and visual stimuli originating from the same source. Indeed, when the auditory stimuli were featurally distinct and did not share analogous space with the visual stimuli, temporal fission (and temporal ventriloquism-like effects) was completely abolished (see Figure 13 and associated statistics in Appendix D). This is consistent with the idea that not only was there weak evidence, or a low likelihood, of the auditory and visual stimuli sharing the same location but the featurally distinct auditory stimuli decreased the likelihood of them originating from the same source (Roseboom et al., 2013b, 2013a). Similarly, when a single tone prior to visual onset increased report of simultaneous visual presentation of circles, causal inference could explain this as sufficiently ambiguous, or noisy, temporal evidence being introduced in combination with prior knowledge of single visual temporal events corresponding to 1 auditory stimulus. This would increase the likelihood that the visual stimuli were 1 temporal event (Rohe et al., 2019). Additionally, while not at statistically significant levels, simultaneity report bias increased consistent with only 1 tone being integrated with the visual stimuli when the temporal ventriloquism-like effects were abolished via 2 centrally presented featurally distinct auditory stimuli (see Figure 13), which is consistent with this Bayesian causal inference perspective.

There is one final consideration when examining the reported results from the perspective of causal inference. The strong neutral spatially opposing tones temporal fission observed in the orthogonal design in Experiment 2 suggests a further step or process may be a factor in causal inference. Namely, strong temporal fission is preserved despite the neutral spatially opposing auditory stimuli not sharing analogous space with the visual stimuli (although the fission illusion in Experiment 1 is stronger still as evidenced by differences in respective BFs, as mentioned previously).

The reason behind the preservation of the strength of the illusion compared to the weaker form of temporal fission may have something to do with auditory apparent motion. There is evidence to suggest that the perception of auditory and visual motion share, or partially share, neural substrates (Berger & Ehrsson, 2016). It is conceivable that the presentation of a tone to one ear followed by the other would induce similar processes to that of auditory apparent motion. Should this be the case and shared visual motion neurons are activated it would provide more evidence of visual motion than via static auditory tones. This could explain why the strength of the temporal fission illusion persists despite not sharing analogous space with the visual stimuli.

In conclusion, we propose an expanded, unifying account of impletion consistent with Bayesian causal inference. This account acts like an umbrella for the gradient account, temporal ventriloquism, and impletion, with an emphasis placed on the most likely real-world events. It considers each factor with varying weightings given to various processes (e.g. where attention is focused/drawn, the number of stimuli in each modality, or where these stimuli are relative to each other in space), and builds an approximate perceptual representation of visual temporal events. For example, if auditory and visual stimuli are perceived to have originated from the same location and subsequently the same source (which is likely the case in the strongest temporal fission illusion in Experiment 1) greater weight is given to this spatial relationship than when auditory stimuli are less likely to be deemed as originating from the same location and the same source as the visual stimuli (which is likely the case in the weaker temporal fission illusion) (Rohe & Noppeney, 2015). Additionally, featural similarity of auditory stimuli is especially important in the absence of spatial congruency. As demonstrated in Experiment 4(ii) spatial congruency trumps the distinct featural differences in the spatially opposing tones temporal fission illusion. However, when spatial congruency with visual stimuli is absent the temporal fission illusion is abolished via featurally distinct tones. Taken together this suggests that the spatial relationship between the auditory and visual stimuli carries more weight than auditory featural similarity when the stimuli share space. However, when the spatial relationship between auditory and visual stimuli is more ambiguous, featural similarity of auditory stimuli is given greater weighting. Similarly to Roseboom et al. (2013a) we suggest that these processes are in line with Bayesian causal inference, where prior knowledge about the world influences integration and segregation, and that featural similarity of sitmuli plays an important role.

Future research should consider the relative weights spatially congruent and featurally similar stimuli have (as well as examining what role motion and apparent motion play) in the visual temporal perception discussed here. In addition to this, the effects described and the proposed expansion of the impletion account would benefit from investigation through the lens of Bayesian inference. Finally, we suggest that variations of the paradigms (and associated effects) reported here be considered for utilisation as part of sensory testing when measuring typical audio-visual integration, such as in cases of cochlear implantation.

References

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

Beierholm, U. R., Quartz, S. R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of vision*, *9*(5), 23–23.

Berger, C. C., & Ehrsson, H. H. (2016). Auditory motion elicits a visual motion aftereffect. *Frontiers in Neuroscience*, *10*, 559.

Boyce, W. P. (2018). *Egghatch — a function for easy pattern addition in ggplot2 bar plots.* Retrieved 2018-11-11, from https://hubrishubrishubris.wordpress.com/2018/01/12/egghatch-a-function-for-easy-pattern-addition-in-ggplot2-bar-plots/

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, *198*, 49–57.

Downing, P. E., & Treisman, A. M. (1997). The line–motion illusion: Attention or impletion? *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 768–779.

Eagleman, D. M., & Sejnowski, T. J. (2003). The line-motion illusion can be reversed by motion signals after the line disappears. *Perception*, *32*, 963–968.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007, May 01). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. Retrieved from https://doi.org/10.3758/BF03193146 doi: 10.3758/BF03193146

Freeman, E., & Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Current biology*, *18*, 1262–1266.

Fuller, S., & Carrasco, M. (2009). Perceptual consequences of visual performance fields: The case of the line motion illusion. *Journal of vision*, *9*(4), 13–13.

Getzmann, S. (2007). The effect of brief auditory stimuli on visual apparent motion. *Perception*, *36*, 1089.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112.

Hidaka, S., Manaka, Y., Teramoto, W., Sugita, Y., Miyauchi, R., Gyoba, J., … Iwaya, Y. (2009). The alternation of sound location induces visual motion perception of a static object. *PLoS One*, *4: e8188*, doi: 10.1371/journal.pone.0008188.

Hikosaka, O., Miyauchi, S., & Shimojo, S. (1993a). Focal visual attention produces illusory temporal order and motion sensation. *Vision research*, *33*, 1219–1240.

Hikosaka, O., Miyauchi, S., & Shimojo, S. (1993b). Voluntary and stimulus-induced attention detected as motion sensation. *Perception*, *22*, 517–526.

Jeffreys, H. (1961). *Theory of probability*. Oxford: Clarendon Press.

Kafaligonul, H., & Stoner, G. R. (2010). Auditory modulation of visual apparent motion with short spatial and temporal intervals. *Journal of vision*, *10*(12), 31–31.

Kafaligonul, H., & Stoner, G. R. (2012). Static sound timing alters sensitivity to low-level visual motion. *Journal of vision*, *12*(11), 2–2.

Keetels, M., Stekelenburg, J., & Vroomen, J. (2007). Auditory grouping occurs prior to intersensory pairing: evidence from temporal ventriloquism. *Experimental Brain Research*, *180*, 449-456.

Klimova, M., Nishida, S., & Roseboom, W. (2017). Grouping by feature of cross-modal flankers in temporal ventriloquism. *Scientific reports*, *7*(1), 7615.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS one*, *2*(9), e943.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lu, C.-H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the simon and spatial stroop effects. *Psychonomic Bulletin and Review*, *2*, 174-174.

Ly, A., Etz, A., Marsman, M., & Wagenmakers, E.-J. (2017). Replication bayes factors from evidence updating. *Behavior research methods*, 1–11.

Mateeff, S., Hohnsbein, J., & Noack, T. (1985). Dynamic visual capture: apparent auditory motion induced by a moving visual target. *Perception*, *14*, 721–727.

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

R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org` (ISBN 3-900051-07-0)

Radeau, M., & Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs. *Psychological Research*, *49*, 17–22.

Rohe, T., Ehlis, A.-C., & Noppeney, U. (2019). The neural dynamics of hierarchical bayesian causal inference in multisensory perception. *Nature communications*, *10*(1), 1907.

Rohe, T., & Noppeney, U. (2015). Cortical hierarchies perform bayesian causal inference in multisensory perception. *PLoS Biology*, *13*(2), e1002073.

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

Roseboom, W., Kawabe, T., & Nishida, S. Y. (2013b). Direction of visual apparent motion driven by perceptual organization of cross-modal signals. *Journal of Vision*, *13*, 1–13.

Rouder, J., Speckman, P., Sun, D., Morey, R., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin and review*, *16(2)*, 225-237.

RStudio Team. (2015). Rstudio: Integrated development environment for r [Computer software manual]. Boston, MA. Retrieved from `http://www.rstudio.com/`

Santangelo, V., & Spence, C. (2008). Crossmodal attentional capture in an unspeeded simultaneity judgement task. *Visual Cognition*, *16*(2-3), 155–165.

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural computation*, *19*(12), 3335–3355.

Schmidt, W. C. (2000). Endogenous attention and illusory line motion reexamined. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(3), 980.

Schmidt, W. C., Fisher, B. D., & Pylshyn, Z. W. (1998). Multiple-location access in vision: Evidence from illusory line motion. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 505.

Schmidt, W. C., & Klein, R. M. (1997). A spatial gradient of acceleration and temporal extension underlies three illusions of motion. *Perception*, *26*(7), 857–874.

Schneider, K. A., & Bavelier, D. (2003). Components of visual prior entry. *Cognitive psychology*, *47(4)*, 333-366.

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in cognitive sciences*, *14*(9), 425–432.

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

Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, *16*(17), 1923–1927.

Shimojo, S., Miyauchi, S., & Hikosaka, O. (1997). Visual motion sensation yielded by non-visually driven attention. *Vision Research*, *37*, 1575–1580.

Spence, C. (2015). Cross-modal perceptual organization. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 649–664).

Spence, C., & Parise, C. (2010). Prior-entry: A review. *Consciousness and cognition*, *19(1)*, 364-379.

Spence, C., Sanabria, D., & Soto-Faraco, S. (2007). Intersensory Gestalten and crossmodal scene perception. In K. Noguchi (Ed.), *Psychology of beauty and kansei: New horizons of gestalt perception* (pp. 519–579). Fuzanbo International, Tokyo.

Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology*, *130(4)*, 799-831.

Teramoto, W., Hidaka, S., Sugita, Y., Sakamoto, S., Gyoba, J., Iwaya, Y., & Suzuki, Y. (2012). Sounds can alter the perceived direction of a moving visual object. *Journal of Vision*, *12*, 1–12.

Teramoto, W., Manaka, Y., Hidaka, S., Sugita, Y., Miyauchi, R., Sakamoto, S., … Suzuki, Y. (2010). Visual motion perception induced by sounds in vertical plane. *Neuroscience Letters*, *479*(3), 221–225.

Tse, P., Cavanagh, P., & Nakayama, K. (1998). The role of parsing in high-level motion processing. *High-level motion processing: Computational, neurobiological, and psychophysical perspectives*, 249–266.

Ursino, M., Crisafulli, A., di Pellegrino, G., Magosso, E., & Cuppini, C. (2017). Development of a bayesian estimator for audio-visual integration: A neurocomputational study. *Frontiers in computational neuroscience*, *11*, 89.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Audiovisual events capture attention: Evidence from temporal order judgments. *Journal of Vision*, *8(5)*, 1-10.

Vroomen, J., & de Gelder, B. (2004). Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology*, *30*, 513–518.

Wallace, M. T., Roberson, G., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, *158*(2), 252–258.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from `http://ggplot2.org`

Acknowledgements

The authors would like to thank Annabel Field, Nadine Garland, Nikolaos Kostis, and Bonita Fung for their help with data collection. The authors would also like to thank Davide Ricco for his assistance with CRT measurements. Finally, we would like to thank Lord Dominique Makowski and Arkady Zgonnikov for feedback on the text, and Nadine Garland for proofreading the manuscript.

Appendix A

Experiment 3

Methods

Participants. Twenty-five participants, 10 male and 15 female (mean age 22.2yrs, SD=2.84), with normal or corrected-to-normal vision, and self-reported normal hearing participated. All were students from Swansea University. All participants were naïve to the purposes of the study. Ethical approval was received from the Department of Psychology Ethics Committee for this research.

An a priori power analysis was applied using the data collected in Experiment 1. An identical condition to that used in this design displayed an effect size of $d = 1.66$ when comparing differences in the means of report bias corresponding to the actual presentation order of visual stimuli between 1 central tone and baseline (no tones) in the collapsed sequential visual conditions ($t(26) = 6.08$, $p < .001$, $d = 1.66$, $SE = 0.03$, where $BF = 9393.87$ which provides extreme evidence indicating the presence of a temporal fusion illusion – – see the Results section in Experiment 1 for notes on how the BF was computed). This condition was first used in Experiment 1 and existed explicitly to detect whether temporal fusion was present, and is therefore one of the most important effects under consideration. Using GPower (Faul et al., 2007) with 95% power and α = .001 (consistent with the reported p-value from Experiment 1) in a difference between two dependent means (matched pairs) power analysis, the recommended sample size was 22 for an actual power estimate of 95.12%. The sample size used here was deliberately larger due to concerns about baseline performance as outlined in Experiment 1.

Apparatus. The apparatus used was the same as Experiment 1.

Stimuli and Procedure. The stimuli and procedure was the same as Experiment 1 with the exception that there were only two response options in an SJ design; sequential presentation of circles (either left or right circle first); and simultaneous presentation of circles. The buttons on the response box were oriented vertically to remove axis congruency with all audio and all visual stimuli and the buttons representing each choice were counter-balanced. The lack of a central button ensured that there was no mapping to central space in the centrally presented auditory conditions.

Results

Application of consistent exclusion criteria (50% threshold instead of 34% threshold here for report bias of interest due to binary response options) resulted in the removal of nine participants. These participants summed with the removal of trials that were below or above the RT criteria saw the total removal of 7018 observations (37.43% of trials) from Experiment 3. The stimuli categories are the same as those listed in Tables 2 and 3.

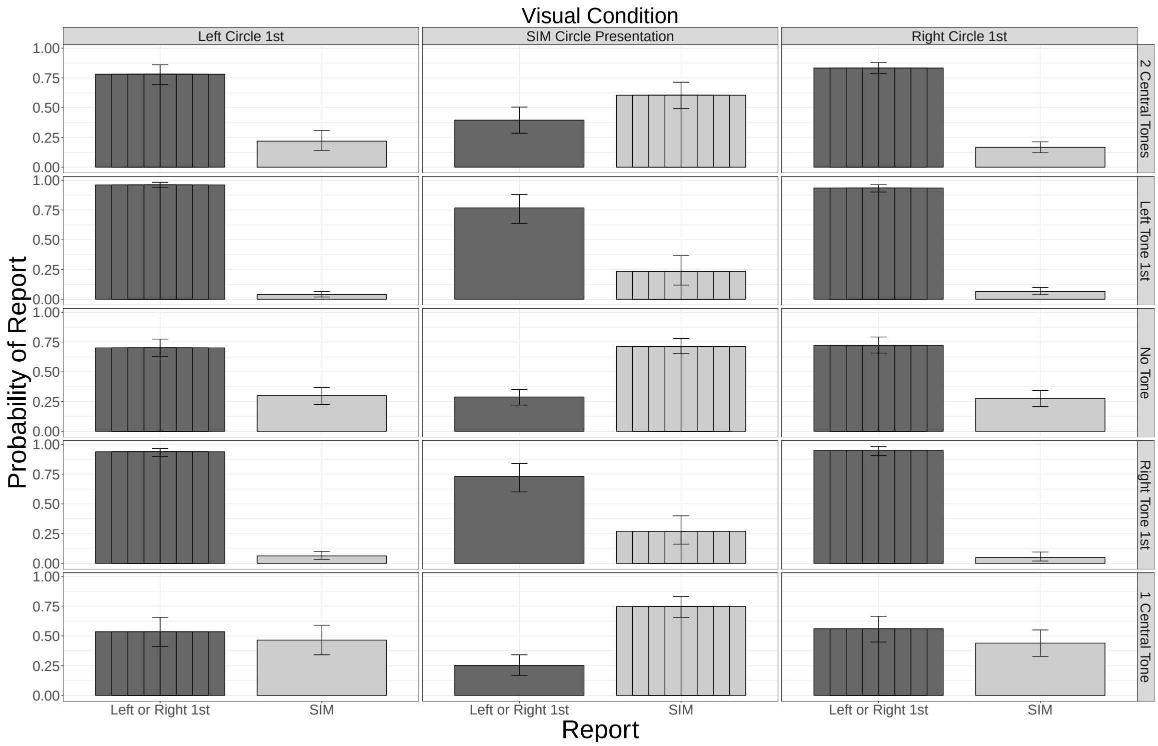

Figure 8. Experiment 3 report probability: The 3 visual conditions are labelled at the top of the grid horizontally. The leftmost column denotes sequential presentation of circles, where the first circle was presented to the left of fixation. The rightmost column denotes sequential presentation of circles, where the first circle was presented to the right of fixation. The central column denotes simultaneous presentation of circles, where a circle was presented to both the left and the right of fixation simultaneously. The 5 auditory conditions are labelled vertically on the rightmost edge of the grid, denoting (from top-to-bottom) the presentation of: 2 tones in analogous central space; a tone presented to the left ear followed by a tone presented to the right ear; no tones; a tone presented to the right ear followed by a tone presented to the left ear; 1 tone in analogous central space respectively. Error bars are bootstrapped within-subject 95% confidence intervals. Reports are labelled on the x-axis with reports corresponding to the presentation category of visual stimuli (simultaneous vs sequential presentation) highlighted with vertical hatching.

The same transformation was applied to the data for null hypothesis testing as used in Experiment 1 and 2. The same approach was used when calculating the *BF* as Experiments 1 and 2. We also created subgroups of the data in a similar fashion to those in Experiment 1 for purposes of analysis.

Analysis of Report Bias Corresponding to Presentation Category of Visual Stimuli. We examined whether there was report bias indicative of the classic temporal fission effect. We conducted a 1 (visual condition: simultaneous visual condition) × 2 (auditory condition: collapsed spatially opposing tones vs. no tone) repeated measures ANOVA on simultaneity report bias.

There was a significant main effect of auditory condition, $F(1, 15) = 62.33$, $MSE = 0.038$, $p < .001$, $\eta^2_G = .527$. A series of t-tests were run to establish the direction of the effects.

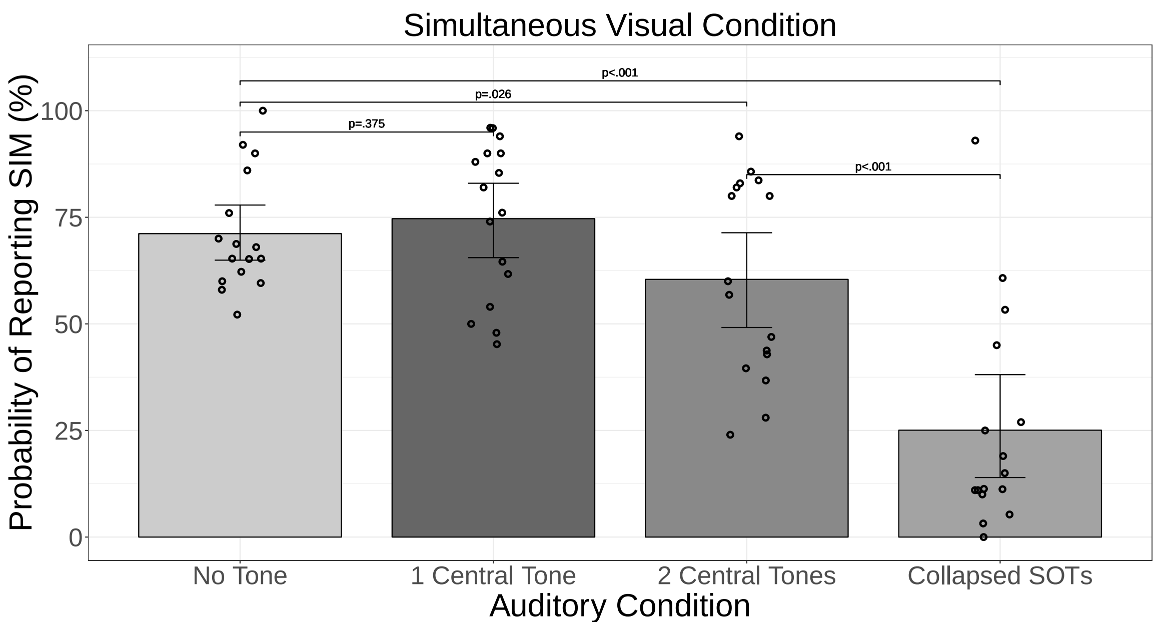

Figure 9. The probability of reporting simultaneous presentation when visual stimuli were presented simultaneously in Experiment 3: The probability (%) of reporting that visual stimuli were presented simultaneously is plotted on the y-axis and the auditory stimuli are labelled on the x-axis where the SOTs (spatially opposing tones) have been collapsed. Error bars are bootstrapped within-subject 95% confidence intervals.

Figure 9 shows that there was a reduction in report bias corresponding to the simultaneous visual presentation condition (reporting simultaneity) in the collapsed spatially opposing tones condition compared to baseline (no tone). The $BF = 5.27e{+}09$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides extreme evidence indicating the presence of the temporal fission illusion in the collapsed spatially opposing tones condition. This replicates the classic temporal fission illusion, which supports the gradient account from the view that the tones share the same analogous space as the visual stimuli.

We tested whether there was further evidence for the gradient account via a 1 (visual condition: collapsed sequential visual conditions) × 2 (auditory condition: collapsed spatially opposing tones vs. no tone) repeated measures ANOVA on sequential report bias.

There was a significant main effect of auditory condition, $F(1, 15) = 63.71$, $MSE = .016$, $p < .001$, $\eta^2_G = .589$, which, as can be seen in Figure 10, shows an increase in sequential report bias corresponding to sequential presentation of visual stimuli overall compared to baseline (no tone).

The $BF = 1.05e + 06$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides extreme evidence indicating the presence of increased sequential report bias corresponding to sequential presentation of visual stimuli in the collapsed spatially opposing tones condition, which in turn is consistent the gradient account. Figure 8 shows an increase in sequential report bias corresponding to sequential presentation of visual stimuli regardless of whether the first tone cues the same space as the first circle presented in sequence due to the responses afforded the participants (i.e. prior entry reversing the direction of perceived sequential presentation was not detected due to 'sequential presentation' and 'simultaneous presentation' being the only responses available to participants).

We tested whether there was an indication that the classic temporal ventriloquism effect (in this instance reflected as an increase in sequential report bias corresponding to sequential visual presentation – see Figure 8 for illustration of this increase in probability of sequential report bias corresponding to sequential presentation of visual stimuli) *may* have been present. We conducted a 1 (visual condition: collapsed sequential visual conditions) × 2 (auditory condition: 2 tones presented to analogous central space vs. no tone) repeated measures ANOVA on sequential report bias.

There was a significant main effect of auditory condition, $F(1, 15) = 5.47$, $MSE = .022$, $p = .034$, $\eta^2_G = .115$.

Figure 10 shows that there was an increase in sequential report bias corresponding to sequential presentation of visual stimuli when 2 tones were presented to analogous central space during collapsed sequential visual conditions compared to baseline (no tone). The $BF = 1.75e + 05$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides extreme evidence that there was an increase in sequential report bias corresponding to sequential presentation of visual stimuli in the two centrally presented tones condition. Without left or right circle first report options we cannot be certain that there was an increase in report bias similar to Experiments 1 and 2. It is entirely possible there was an increased likelihood that participants perceived sequential order in general that may not have corresponded to the actual order of visual stimuli presentation. However, this seems unlikely given previous results.

We conducted an ANOVA to determine if the spatial location of tones, relative to the visual stimuli, had an effect on simultaneity report bias in Experiment 3. A 1 (visual condition: simultaneous visual condition) × 3 (auditory presentation location: spatially opposing tones presented to analogous space to that of the visual stimuli vs. two tones presented to neutral space (analogous central space in this instance) vs. no tone presented to any space) repeated measures ANOVA on simultaneity report bias was conducted.

Mauchly's test for Sphericity failed for auditory presentation location, $W = .611$, $p = .032$, $p < .001$. Therefore, the degrees of freedom were corrected using Greenhouse-Geisser Estimate ϵ (Greenhouse & Geisser, 1959).

There was a significant main effect of auditory presentation location, $F(1.44, 21.6) = 26.05$, $MSE = .050$, $p < .001$, $\eta^2_G = .328$.

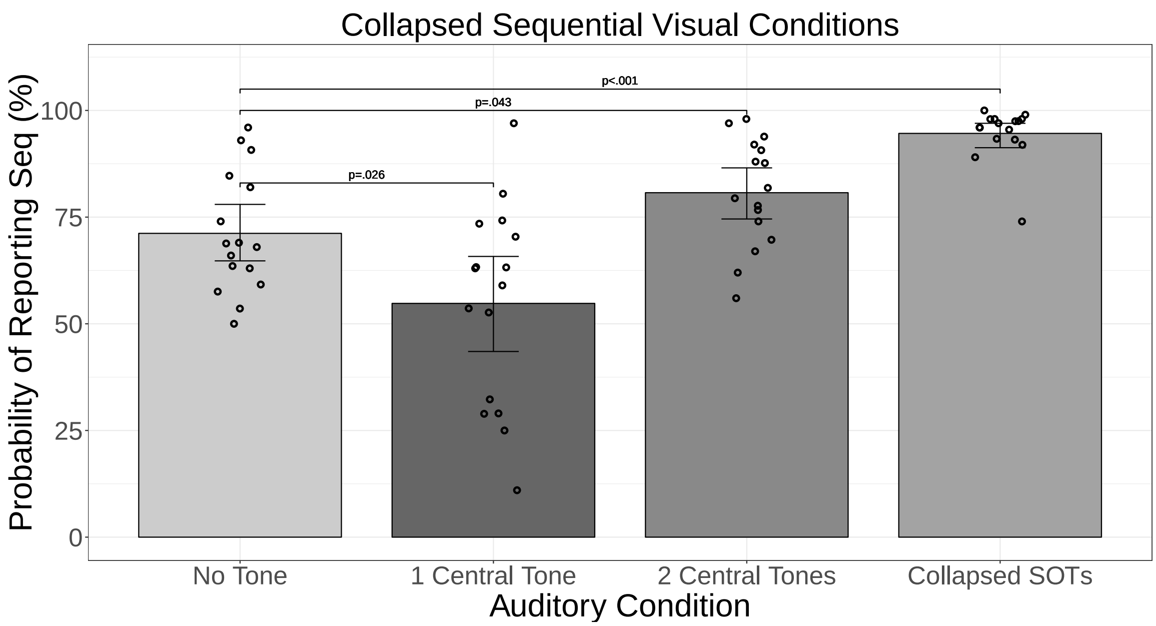

Figure 10. The probability of reporting sequential presentation (note: one response option 'sequential' was afforded for both left circle first and right circle first perceptions) when visual stimuli were presented sequentially in Experiment 3: The probability (%) of reporting sequential presentation is plotted on the y-axis and the auditory stimuli are labelled on the x-axis where SOTs (spatially opposing tones) were collapsed. Error bars are bootstrapped within-subject 95% confidence intervals.

The above ANOVA replicates Experiments 1 and 2 by demonstrating spatial location is important when inducing visual temporal fission via auditory tones. However, the above ANOVA does not make clear if it is a requisite that auditory tones be presented to the same space as the visual stimuli in order to induce temporal fission (as would be the case if the gradient account was the sole driver for the effect). We performed t-tests below, and calculated Bayes Factors, with the view to clarifying this. Figure 9 contains the relevant plots for the data used in the means comparisons.

There was a reduction in report bias corresponding to presentation of the visual stimuli (reporting simultaneity) in the simultaneous visual condition when 2 tones were presented to analogous central space compared to baseline (no tone), $t(15) = 2.78$, $p = .026$, $d = 0.98$, $SE = 0.04$. The $BF = 27.41$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides strong evidence indicating the presence of the temporal fission illusion in the 2 central tones condition, which supports an impletion account of temporal fission where tones are not required to share the same space as the visual stimuli.

Spatially opposing tones were significantly more likely to result in report bias in opposition to the actual presentation of visual stimuli (simultaneous visual condition) when compared to 2 tones presented in analogous central space, $t(15) = 5.08$, $p < .001$, $d = 1.80$, $SE = 0.08$. The $BF = 5.18e + 10$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides extreme evidence indicating the presence of a stronger temporal fission illusion in the collapsed spatially opposing tones condition, which supports both impletion and the gradient account, as elaborated on in the discussion for Experiment 1.

We conducted an ANOVA to determine if the number of tones, relative to visual stimuli (which always consisted of 2 circles, although they differed in presentation: sequential vs. simultaneous), had an effect on report bias corresponding to simultaneous or sequential presentation of visual stimuli in Experiment 3. A 2 (visual condition: simultaneous visual condition vs. collapsed sequential visual conditions) × 3 (number of tones: 1 tone presented to analogous central space; 2 tones presented to analogous central space; and no tones presented to any space) repeated measures ANOVA on report bias corresponding to the actual presentation type of visual stimuli. The auditory conditions used in this analysis were chosen due to their contrasting number of presentations, while all auditory stimuli shared the same presentation space (analogous central space which was neutral relative to the visual stimuli locations).

There was no significant main effect of visual condition, $F(1, 15) = 2e - 04$, $MSE = 0.103$, $p = .989$, $\eta^2_G =< .001$. There was a significant main effect of the number of tones, $F(2, 30) = 8.70$, $MSE = .006$, $p = .001$, $\eta^2_G = .024$. There was a significant interaction of visual condition and number of tones, $F(2, 30) = 12.37$, $MSE = .036$, $p < .001$, $\eta^2_G = .170$.

The ANOVA above shows that the number of tones presented is important when inducing visual temporal effects. However, it does not make clear if it is a requisite that the number of auditory tones should match the number of visual stimuli in order to induce said temporal effects (as would be the case if Morein-Zamir et al.'s (2003) account is accurate). We performed t-tests below, and calculated Bayes Factors, with the view of clarifying this. Figures 9 and 10 contain the plots for the data used in the means comparisons.

One central tone accompanying collapsed sequential visual conditions reduced sequential report bias corresponding to sequential presentation of visual stimuli compared to baseline (no tone), $t(15) = 2.77$, $p = .026$, $d = 0.98$, $SE = 0.07$. The $BF = 1.34e + 05$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides extreme evidence indicating the presence of a temporal fusion illusion in the 1 central tone condition, which is consistent with Getzmann's (2007) finding that 1 tone was sufficient induce temporal ventriloquism-like effects. The relatively small effect size (in the null hypothesis t-test), compared to Experiments 1 and 2, for this condition may be due to the sample size being smaller (after applying exclusion criteria) than the 22 that was recommended in the reported power analysis.

Despite a slight increase in report bias corresponding to the actual presentation of visual stimuli, there was no statistical difference when 1 tone was presented to analogous central space when compared to baseline (no tone) in the simultaneous visual condition, $t(15) = .92$, $p = .375$, $d = 0.32$, $SE = 0.05$. The $BF = 3.29e - 02$ (adjusted using the "evidence updating" method (Ly et al., 2017)), which provides very strong evidence for the null hypothesis in the 1 central tone condition.

Discussion. The main findings of Experiment 1 were largely replicated here (some cannot be confirmed due to the reduced resolution of a SJ-judgment task). Temporal fission was induced via 2 tones presented to analogous central space and also via 2 spatially opposing tones.

Temporal fusion was also induced via 1 centrally presented tone before visual onset. This finding is of particular importance here as it helps rule out the possibility of button arrangement influencing responses, and addresses concerns surrounding the use of a ternary response task in the previous experiments.

Appendix B

All t-tests included in this section were included in the relevant FDR corrections in the main body of the manuscript.

Supplementary t-tests for Experiment 1

When the spatially opposing tones from left-to-right ears were presented with sequential circle presentation from right-to-left, there was a reduction in report bias corresponding to the actual presentation order of visual stimuli when compared to spatially congruent audio stimuli, $t(26) = 9.62$, $p < .001$, $d = 2.62$, $SE = 0.04$. The $BF = 2.28e + 07$ which provides extreme evidence indicating the presence of prior entry.

When the spatially opposing tones from right-to-left ears were presented with sequential circle presentation from left-to-right, there was a reduction in report bias corresponding to the actual presentation order of visual stimuli when compared to spatially congruent audio stimuli, $t(26) = 11.97$, $p < .001$, $d = 3.26$, $SE = 0.04$. The $BF = 1.89e + 09$, which provides extreme evidence indicating the presence of prior entry.

The above BFs show that, while collapsed spatially opposing tones in the main results showed enhancement, there was detriment in performance when the audio stimuli cued the space that the second circle was presented to and then cued the space the first circle was presented to. As can be seen in Figure 2, the effect of prior entry was often so strong that it reversed the direction of presentation in perception.

Supplementary t-tests for Experiment 3

When the spatially opposing tones from left-to-right ears were presented with sequential circle presentation from right-to-left, there was no reduction in sequential report bias corresponding to sequential presentation of visual stimuli when compared to spatially congruent audio stimuli, $t(15) = 2$, $p = .072$, $d = 0.71$, $SE = 0.03$. The $BF = 1.24$, which provides anecdotal evidence indicating the presence of prior entry.

When the spatially opposing tones from right-to-left ears were presented with sequential circle presentation from left-to-right, there was a reduction in sequential report bias corresponding to sequential presentation of visual stimuli when compared to spatially congruent audio stimuli, $t(15) = 2.60$, $p = .03$, $d = 0.92$, $SE = 0.02$. The $BF = 3.09$, which provides moderate evidence indicating the presence of prior entry.

The above BFs show that, while collapsed spatially opposing tones in the main results showed increased sequential report bias, there was a reduction in sequential report bias when the audio stimuli cued the space that the second circle was presented to and then cued the space the first circle was presented to. This can be seen in Figure 8, where small variations in sequential report bias corresponding to sequential presentation of visual stimuli are shown. Due to the task being an SJ-judgment, it was not possible to ascertain if the prior entry was strong enough to reverse the perceived direction of visual presentation, as was the case in Experiment 1 above.

Appendix C

Below is sample data taken from a pilot with various featural and spatial manipulations of auditory stimuli. The methods and basic design were the same as in Experiment 1. Ten participants, 3 female, and 7 male (mean age = 23.8, SD = 4.44), 9 of whom were naïve to the purpose of the experiment and 1 of whom was the experimenter, participated in the experiment. Using the same exclusion criteria as all previous experiments resulted in 7 participants being included in the analysis below.

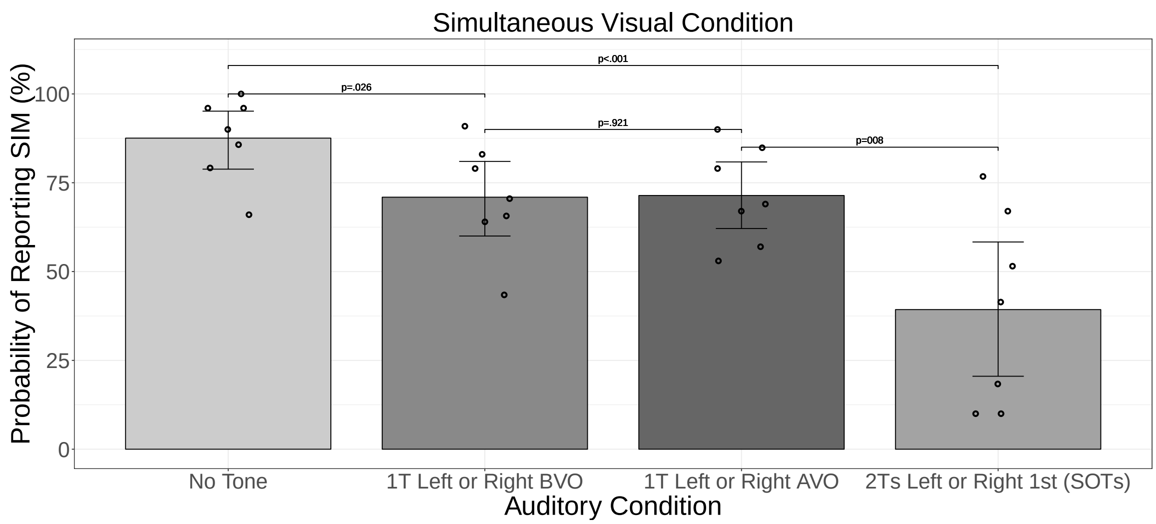

Figure 11. The probability of reporting simultaneous visual stimuli presentation in Experiment 4(i): The y-axis represents the probability (%) of reporting simultaneous presentation of visual stimuli in the simultaneous visual presentation condition. The x-axis represents the various auditory conditions where the SOTs (spatially collapsed tones) conditions have been collapsed, '1T' equates to '1 tone', BVO equates to 'before visual onset', and AVO equates to 'after visual onset'. Error bars are bootstrapped within-subject 95% confidence intervals.

Table 6

FDR corrected t-tests for the simultaneous visual condition Exp 4.

Condition 1	Condition 2						
Auditory	Auditory	t	$p - value$	df	d	SE	BF
No Tone	1T Left or Right BVO	3.09	p = .026	6	1.65	.08	3.762311e+00
No Tone	1T Left or Right AVO	3.48	p = .02	6	1.86	.07	5.462537e+00
No Tone	2Ts Left or Right 1st (SOTs)	6.86	p < .001	6	3.67	.09	7.475399e+01
2Ts Left or Right 1st (SOTs)	1T Left or Right BVO	4.22	p = .011	6	2.25	.08	1.055174e+01
2Ts Left or Right 1st (SOTs)	1T Left or Right AVO	4.91	p = .008	6	2.62	.07	1.869991e+01
1T Left or Right BVO	1T Left or Right AVO	0.10	p = .921	6	0.06	.04	3.547582e-01

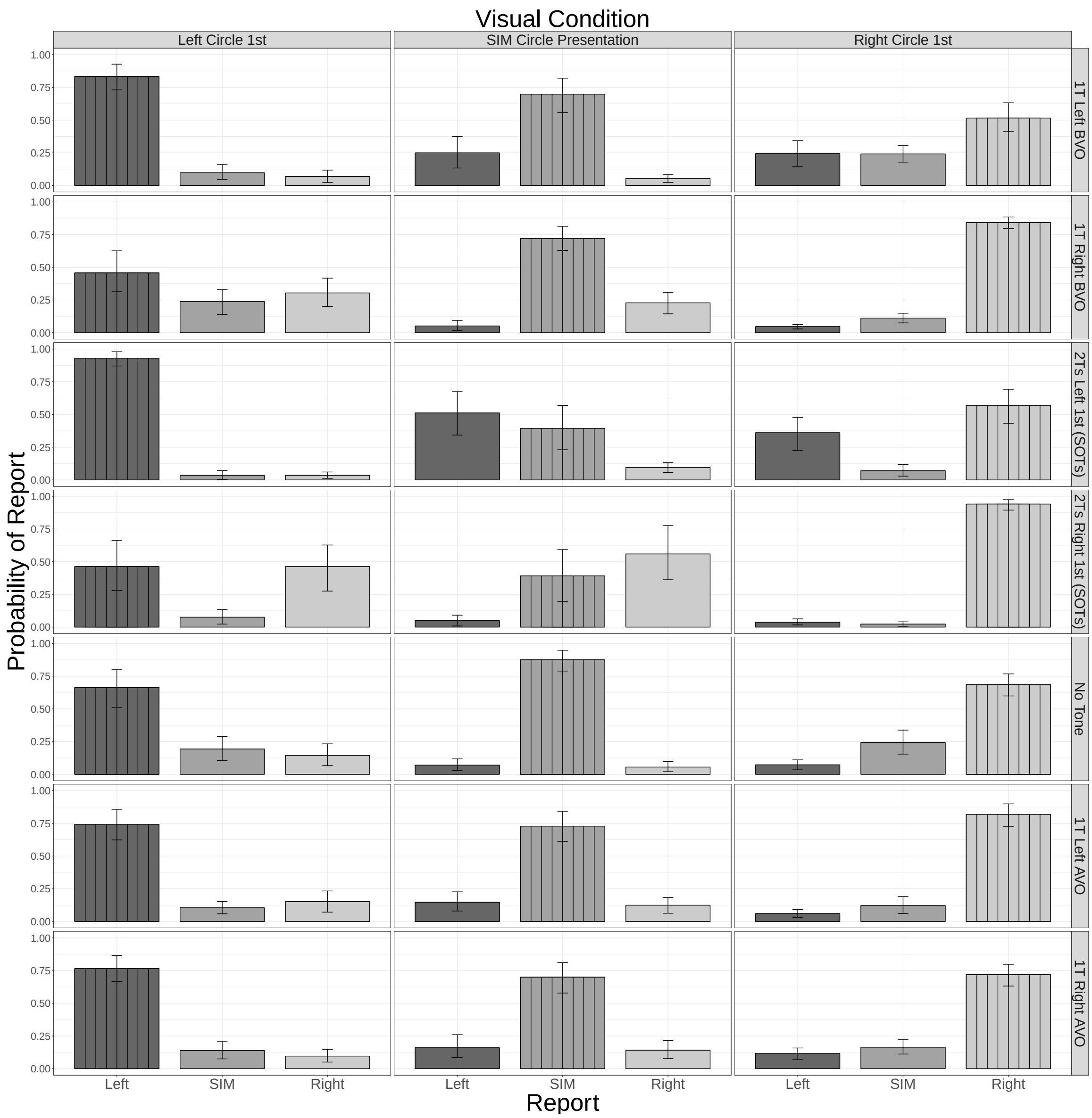

Figure 12. Experiment 4(i) report probability: The 3 visual conditions are labelled at the top of the grid horizontally. The leftmost column denotes sequential presentation of circles, where the first circle was presented to the left of fixation. The rightmost column denotes sequential presentation of circles, where the first circle was presented to the right of fixation. The central column denotes simultaneous presentation of circles, where a circle was presented to both the left and the right of fixation simultaneously. The 7 auditory conditions are labelled vertically on the rightmost edge of the grid, denoting (from top-to-bottom) the presentation of: 1 tone (1T) presented to left ear before visual onset (BVO); 1T presented to right ear BVO; a tone presented to the left ear followed by a tone presented to the right ear; a tone presented to the right ear followed by a tone presented to the left ear; no tones; 1T presented to the left ear after visual onset (AVO); 1T presented to the right ear AVO respectively. Error bars are bootstrapped within-subject 95% confidence intervals. Reports are labelled on the x-axis with reports corresponding to the actual presentation order of visual stimuli highlighted with vertical hatching.

Appendix D

Below is sample data taken from a the same pilot as reported in Appendix C.

There were various featural and spatial manipulations of auditory stimuli. The methods and basic design were the same as in Experiment 1. Ten participants, 7 male, and 3 female (mean age = 23.8, SD = 4.44), 9 of whom were naïve to the purpose of the experiment and 1 of whom was the experimenter, participated in the experiment. Using the same exclusion criteria as all previous experiments resulted in 7 participants being included in the analysis below.

There was no statistical difference in report bias corresponding to the actual presentation of visual stimuli (reporting simultaneity) in the simultaneous visual condition when 2 different tones (one a sine-wave, the other a white noise burst) were presented to analogous central space compared to baseline (no tone). $t(6) = 2.15$, $p = .08$, $d = 0.81$, $SE = 0.05$. The $BF = 1.48$, which provides anecdotal evidence indicating a slight increase in report bias corresponding to the actual presentation of visual stimuli when 2 featurally different tones were presented to analogous central space. This trend in report bias is more consistent with a single tone than with 2 tones that are identical, thus supporting the notion that featurally distinct auditory stimuli presented to analogous central space are deemed to be from different sources and subsequently only one, or neither, is bound with the visual stimuli in temporal perception. In turn this abolishes the temporal fission illusion via 2 centrally presented tones.

Table 7 shows that report biases consistent with temporal ventriloquism were no longer statistically different from control conditions. Figure 13 shows a trend towards increased simultaneity report bias in both of the sequential visual conditions when 2 centrally presented auditory stimuli were featurally distinct, compared to controls (no tones). This is consistent with the abolished

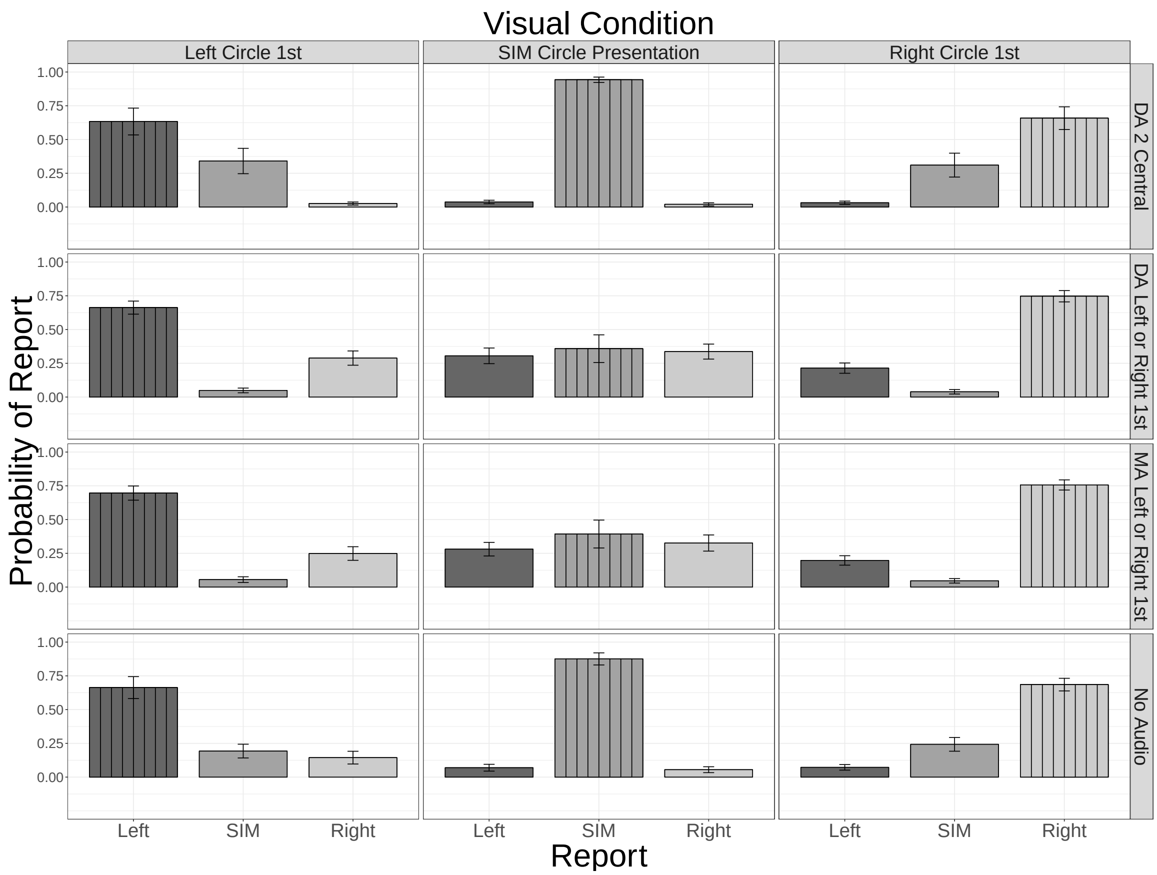

Figure 13. Experiment 4(ii) report probability: The 3 visual conditions are labelled at the top of the grid horizontally. The leftmost column denotes sequential presentation of circles, where the first circle was presented to the left of fixation. The rightmost column denotes sequential presentation of circles, where the first circle was presented to the right of fixation. The central column denotes simultaneous presentation of circles, where a circle was presented to both the left and the right of fixation simultaneously. The 4 auditory conditions are labelled vertically on the rightmost edge of the grid, denoting (from top-to-bottom) the presentation: 2 different auditory stimuli (DA) presented to analogous central space; 2 different auditory stimuli (DA), one presented to the left ear and the other presented to the right ear, or vice versa; matching auditory stimuli (MA) one presented to the left ear and the other presented to the right ear, or vice versa; and no auditory stimuli presented. Error bars are bootstrapped within-subject 95% confidence intervals. Reports are labelled on the x-axis with report corresponding to the actual presentation order of visual stimuli highlighted with vertical hatching.

weaker form of temporal fission where only 1 tone is likely to be integrated with the visual stimuli. While the increase in simultaneity report bias is not supported statistically it is worth noting that the recommend sample size to detect the temporal fusion illusion is 22 for a power estimate of 95.12%.

Table 7

FDR corrected t-tests for reports associated with temporal fission via centrally presented tones, and reports associated with temporal ventriloquism-like effects in Exp 4(ii).

Condition 1	Condition 2						
Visual : Auditory : Report	Visual : Auditory : Report	t	$p - value$	df	d	SE	BF
Left : No Beep : Left	Left : 2 C-DA : Left	0.17	p = .870	6	0.06	.09	3.575557e-01
Right : No Beep : Right	Right : 2 C-DA : Right	0.13	p = .899	6	0.05	.09	3.557655e-01
Left : No Beep : SIM	Left : 2 C-DA : SIM	1.37	p = .220	6	0.52	.11	7.015374e-01
Right : No Beep : SIM	Right : 2 C-DA : SIM	0.52	p = .622	6	0.20	.11	3.949842e-01

Therefore it would be expected, given a sufficiently large sample, that temporal fusion would be present in this condition as opposed to report biases consistent with temporal ventriloquism.