Brief article

Semantic audio-visual congruence modulates visual sensitivity to biological motion across awareness levels

Stefano Ioannucci [*], Petra Vetter

Visual and Cognitive Neuroscience Lab, Dept. of Psychology, University of Fribourg, Switzerland

A B S T R A C T

Whether cross-modal interaction requires conscious awareness of multisensory information or whether it can occur in the absence of awareness, is still an open question. Here, we investigated if sounds can enhance detection sensitivity of semantically matching visual stimuli at varying levels of visual awareness. We presented biological motion stimuli of human actions (walking, rowing, sawing) during dynamic continuous flash suppression (CFS) to 80 participants and measured the effect of co-occurring, semantically matching or non-matching action sounds on visual sensitivity (d′). By individually thresholding stimulus contrast, we distinguished participants who detected motion either above or at chance level.

Participants who reliably detected visual motion above chance showed higher sensitivity to upright versus inverted biological motion across all experimental conditions. In contrast, participants detecting visual motion at chance level, i.e. during successful suppression, demonstrated this upright advantage exclusively during trials with semantically congruent sounds. Across the whole sample, the impact of sounds on visual sensitivity increased as participants' visual detection performance decreased, revealing a systematic trade-off between auditory and visual processing. Our findings suggest that semantic congruence between auditory and visual information can selectively modulate biological motion perception when visual awareness is minimal or absent, while more robust visual signals enable perception of biological motion independent of auditory input. Thus, semantically congruent sounds may impact visual representations as a function of the level of visual awareness.

1. Introduction

Vision does not work in isolation and is substantially influenced by contextual information from other sensory modalities (see e.g., Calvert, 2004; Gilbert & Li, 2013; Klink et al., 2012; Oliva & Torralba, 2007; Petro et al., 2017; Vetter et al., 2014; Vetter & Newen, 2014). Cross-modal interactions between vision and other sensory modalities usually result in the optimization of visual detection or discrimination, reflected in increased accuracy and decreased reaction times compared to unisensory stimulation (e.g. Bulkin & Groh, 2006; Driver & Noesselt, 2008; Noesselt et al., 2010). Crucially, this occurs not only when sensory cues are clearly detected, but particularly when they are ambiguous, uncertain, or suppressed. Cross-modal interactions influence the perception of ambiguous stimuli and their access to conscious awareness, as demonstrated with binocular rivalry (e.g. Blake et al., 2004; Chen et al., 2011; Chen & Spence, 2010; Conrad et al., 2010, 2013; Guzman-Martinez et al., 2012; Hense et al., 2019; Kang & Blake, 2005; Liaw et al., 2022; Lunghi et al., 2010, 2014, 2015; Pápai & Soto-Faraco,

2017, Parker & Alais, 2006; Plass et al., 2017; Zhou et al., 2010), or Continuous Flash Suppression (CFS; e.g. Aller et al., 2015; Alsius & Munhall, 2013; Hong & Shim, 2016; Lunghi et al., 2017; Lupyan & Ward, 2013; Palmer & Ramsey, 2012; Park et al., 2024; Plass et al., 2014; Salomon et al., 2013; Tan & Yeh, 2015; Yang & Yeh, 2014), although not exclusively with those techniques (e.g. Chen & Spence, 2011; Conrad et al., 2012; Delong & Noppeney, 2021; Faivre et al., 2014; Kvasova et al., 2023; Montoya & Badde, 2023; Ngo & Spence, 2010; Soto-Faraco et al., 2004; Williams et al., 2022). Nevertheless, whether cross-modal interactions require conscious awareness of information from one or all interacting modalities remains debated (Deroy et al., 2016; Mudrik et al., 2014). Many studies demonstrate that conscious information from one sense speeds up the emergence of suppressed information in another sense, usually vision, into awareness. Whether cross-modal interaction enhances actual visual stimulus processing or perceptual sensitivity (d′), has been studied much less. Moreover, cross-modal congruency has often been investigated in terms of spatial and/or temporal co-occurrence of simple stimuli such as

flashes, gratings or beeps. Yet, in everyday life, our brain has adapted to integrate diverse information, including semantic associations with varying degrees of ambiguity across sensory modalities. Such high-level information has been thought to require awareness to be combined and to influence behavior (Mudrik et al., 2014). Higher-order cross-modal associations often involve semantic congruence, reflecting the degree of compatibility between the meaning, or semantics, of multiple stimuli. For example, observing raindrops falling on a windowpane while simultaneously hearing the patter of rain has a high degree of semantic congruence since both the visual and auditory stimuli relate to the same concept. When this happens, congruent multisensory stimuli tend to enhance each other and lead to a more robust and coherent perceptual experience, particularly under noisy conditions (Chen & Spence, 2010; Delong & Noppeney, 2021; Kvasova et al., 2023; Williams et al., 2022; Williams & Störmer, 2024). In contrast, incongruent stimuli might create confusion or conflict, leading to a less clear percept (Noppeney et al., 2010). Moreover, a fundamental principle of audio-visual interaction suggests that the influence of one modality increases as information in another modality becomes less reliable (Alais & Burr, 2004).

In the current study, we made use of the power of semantic congruence and investigated the dynamics of cross-modal interaction under noisy visual circumstances and under varying degrees of visual awareness, for complex stimuli with naturalistic configuration that are semantically congruent rather than just spatio-temporally congruent. We used biological motion stimuli (Troje, 2013) depicting different actions (walking, rowing and sawing) and their corresponding action sounds. The brain is highly susceptible to biological motion (Chang et al., 2018; Grèzes et al., 2001; Grossman et al., 2000; Grossman & Blake, 2002; Saygin, 2007; Vaina et al., 2001) and upright biological motion has preferential access to visual awareness compared to inverted or scrambled biological motion (Sun et al., 2022). We employed continuous flash suppression (CFS) to render visual stimuli noisy and suppress biological motion from visual awareness. With CFS, high contrast and fast flickering masks are displayed to the dominant eye while a low contrast image is presented to the non-dominant eye (Tsuchiya & Koch, 2005). CFS provides a highly sensitive tool for testing the capacity of visual stimuli to emerge into awareness. However, simply measuring breakthrough reaction times has its limitations with respect to assessing the extent of unaware stimulus processing (Alais et al., 2024; Lanfranco et al., 2023; Stein, 2019; Stein et al., 2011). Therefore, in this study, we investigated potentially unaware cross-modal interactions by manipulating visual noise level and suppressing visual stimuli from awareness to differential degrees, focusing specifically on how semantically congruent and incongruent sounds affect visual sensitivity (d'). We hypothesized that 1) upright biological motion would have a detection advantage compared to inverted biological motion under CFS (in line with Sun et al., 2022); 2) sounds would improve visual detection under conditions of visual uncertainty; and 3) semantically congruent action sounds would selectively enhance upright biological motion perception under conditions of limited visual awareness.

2. Methods

2.1. Participants

Eighty healthy adult participants took part in the experiment. As we planned to divide participants into awareness groups ("unaware" and "aware", according to visual stimulus detection performance, see below), sample size was overestimated to ensure adequate power in each group and to detect interaction effects across groups and experimental conditions (requiring at least 27 participants per awareness group for 80 % power, based on effect sizes from Sun et al., 2022; calculated via G*Power, Faul et al., 2007). This overestimation was necessary due to the unpredictable variability in how participants were susceptible to our contrast thresholding procedure (see below), with initial observations

(n = 49) suggesting that approximately 40 % of participants would fall into the "unaware group". Thus, data collection occurred in two phases, with all analyses presented here based on the complete dataset. All participants were naive to the procedure with no history of mental or visual conditions. Participants had normal or corrected-to-normal vision and intact hearing, and signed informed consent to take part in the study, for which they received either monetary compensation (20 CHF) or study credits. All procedures were approved by the Psychology Ethics Committee of the University of Fribourg.

2.2. Apparatus

The experiment was coded in PsychoPy (Peirce et al., 2019) and presented on a LCD monitor (ASUSTeK Computer Inc., Taipei, Taiwan) with a resolution of 1920 by 1080 at 60 Hz refresh rate. Observers viewed a dichoptic display through a mirror stereoscope (ScreenScope, ASC Scientific, Carlsbad, USA), adjusted individually to ensure binocular alignment, and with their heads stabilized on a chin rest at a viewing distance of 50 cm. Sounds were displayed via headphones (Sennheiser electronic GmbH & Co, Wedemark-Wennebostel, Germany).

2.3. Stimuli

The visual stimuli were displayed against a uniform mid-gray background and consisted of three biological motion stimuli (walking, rowing, sawing) of equal duration (2 s) with their corresponding synchronized sounds (e.g., footsteps with walking, the rhythmic sound of oars moving through water with rowing, the periodic sound of a saw cutting wood with sawing), adapted from previous work (Meyer et al., 2011). Each biological motion stimulus comprised 13 point-light markers (dark gray dots, individually adjusted for contrast, see below), positioned on key anatomical landmarks: one on the head and two on the shoulders, elbows, wrists, hips, knees, and ankles. As a control condition, these stimuli were inverted by rotating them by 180 degrees. In congruent trials, each biological motion display was paired with its corresponding action sound. For incongruent trials, these same action sounds were randomly paired with non-matching biological motion displays (e.g., sawing sounds with walking motion). Cross-modal correlations between low-level motion and sound features of the stimuli (individual dot velocities and sound envelope) did not differ significantly between congruent and incongruent stimulus pairs ($T_{(16)} = -1.33$, p = 0.2).

The initial starting position of the biological motion was randomly determined across trials to curtail predictability. To effectively suppress moving stimuli from visual awareness, a colored high-contrast 'Mondrian' mask was generated uniquely for each trial. This mask consisted of 256 stochastically generated square shapes, each varying in size (ranging from 30 to 100 pixels in both x and y dimensions) and color. The squares flashed at a frequency of 12 Hz and, in alternating rows, moved randomly either toward the left or right. This dynamic composition aimed to address the limitations of static CFS masks in suppressing moving stimuli (Moors et al., 2014).

The flickering mask was always presented to the dominant eye, which was determined via a dominant eye test (hole-in-card test; Miles, 1930), and the biological motion stimulus was shown to the non-dominant eye using a stereoscope. To help fuse the two images, black and white bars framed the stimulus field for each eye (Fig. 1). Sound stimuli were paired either congruently or incongruently with the biological motion stimuli, or no sound was presented. Sounds were equalized for overall volume and volume was adjusted to an individual comfort level.

2.4. Experimental design

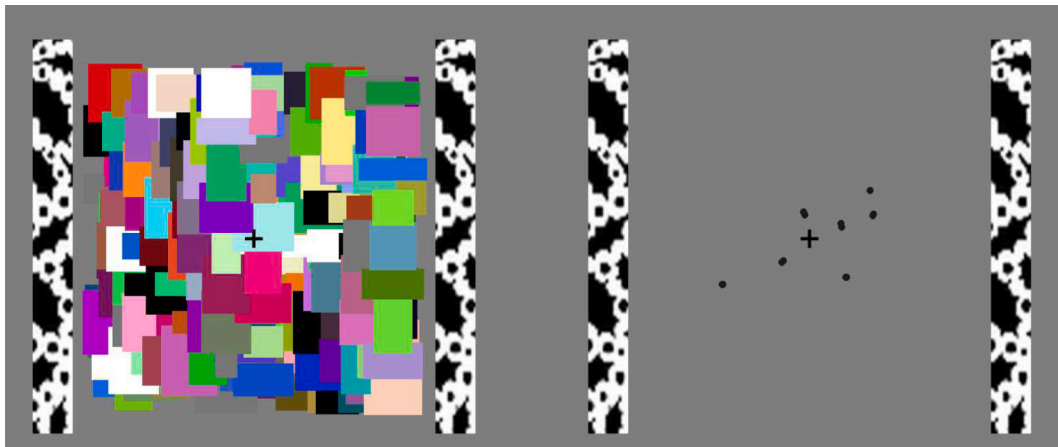A 2 (visual stimulus presence: present or absent) by 2 (biological

Fig. 1. Visual stimulus example. The colorful Mondrian mask is shown to the left (dominant) eye, while an upright biological motion stimulus (rowing; full contrast for illustration) is shown to the right (non-dominant) eye.

motion type: upright or inverted biological motion) by 3 (sound: congruent, incongruent, no sound) design was employed. The experimental run comprised a total of 504 trials in which biological motion stimuli were present in 50 % of trials, of which 50 % had no sound, 25 % had congruent sounds and 25 % incongruent sounds. The trials without visual stimuli were evenly split among those with and without sounds. All sound/stimulus combinations were presented equally often and in pseudo-randomised order.

2.5. Procedure

Before the experimental run, participants underwent a contrast thresholding session where a Bayesian staircase procedure (Ioannucci et al., 2022) was utilized to estimate detection performance at various contrast levels, ranging from 0.02 to 0.17. These contrast values were relative RGB values in PsychoPy's color space, used to create perceptually distinct visibility levels. The thresholding session spanned 180 trials in which biological motion (upright and inverted) was always present together with a white noise (432hz) stimulus. Based on each participant's psychometric curve, we aimed to assign contrast levels that would yield either chance (50 %) or above-chance (80 %) performance, targeting an approximately equal distribution between awareness conditions. However, in the main experimental session, the introduction of meaningful action sounds influenced detection performance differently than the white noise used during thresholding. This led to some participants performing better or worse than predicted by their contrast threshold, resulting in final group sizes of 52 participants performing above chance and 28 at chance level. The intentional variation in contrast thresholds was undertaken to manipulate participant's degree of visual awareness, mitigating problematic post-hoc selection procedures (Shanks, 2017). In the main experimental session, the subject-specific contrast as determined in the thresholding session was maintained.

Each trial started with a central fixation cross for 500 ms which remained on the display during 2000 ms in which the flickering mask and the biological motion stimulus (if present) along with the sound (if present) was presented. During the first 900 ms the contrast of the biological motion stimuli was ramped up to the previously determined threshold.

Participants were instructed to press keys on the keyboard as soon as they perceived any moving dot. If no response was made by the end of a trial, a prompt was presented ('Have you seen any moving dot? Yes/ No'). As a secondary task in each trial, observers were then also required to indicate which movement they might have been presented with (rowing, walking or sawing), regardless of their previous response. No feedback was provided.

Note that this experimental design was not a classic breaking CFS (bCFS) paradigm where the target stimulus is presented until participants report awareness and breakthrough RTs are measured. Instead, this paradigm was optimised for measuring visual sensitivity (d′, see below). Target stimulus and mask presentation ended automatically after 2000 ms, and if participants had not reported awareness of any moving dots before then, they responded to the first task ('Have you seen any moving dot? Yes/No'), before responding to the second task ("Which movement?").

2.6. Data analysis

Trials with outlying reaction times were excluded using the Median Absolute Deviation method (Leys et al., 2013). Our primary dependent measure was visual sensitivity (d′) which we derived from dot detection accuracy using signal detection theory. For each participant and condition combination, hit rates were calculated as the proportion of 'yes' responses on stimulus-present trials, while false alarm rates were calculated as the proportion of 'yes' responses on stimulus-absent trials. To avoid infinite values, a log-linear correction was applied by adding 0.5 to both the number of hits and false alarms and adding 1 to both the number of signal and noise trials. Hit and false alarm rates were then converted to z-scores using the inverse of the cumulative normal distribution. Visual sensitivity was computed as the difference between the z-transformed hit rate and z-transformed false alarm rate. This calculation was performed separately for each combination of biological motion type (upright/inverted) and sound condition (congruent/incongruent/no sound). To assess the impact of the experimental conditions on visual sensitivity, we fitted a linear mixed model with d′ as the dependent variable, including sound condition, biological motion type and awareness group as fixed effects, reaction time as covariate, along with their interactions (d′ ~ 1 + sound condition + motion condition + awareness + rt + sound condition * motion condition + sound condition * awareness + motion condition * awareness + sound condition * motion condition * awareness + (1 + motion condition | participant)). The random effects structure included random intercepts to account for participant-specific baseline differences in sensitivity, as well as random slopes for biological motion condition to allow for individual variations in how participants processed upright versus inverted biological motion. This structure was determined by sequentially adding random slopes for each factor until model convergence failed, with biological motion condition being the fundamental experimental manipulation that could be included while maintaining model convergence. The model was implemented via jamovi (The jamovi project, 2019) with degrees of freedom calculated by the Satterthwaite method (Satterthwaite, 1946).

As a secondary dependent measure, we also derived mean reaction

time (RT) from all trials, merging trials where participants reported the presence of a dot before the trial ended automatically and from trials where participants responded to the first task ('Have you seen any moving dot? Yes/No'). Mean RT was derived for each combination of biological motion type, sound condition and awareness level, as for d' values. We then ran a similar linear mixed model employing reaction time as dependent measure and d' as covariate. Here, in the random part of the model, it was not possible to include biological motion as it prevented model convergence, but it was possible to include sound condition (rt ~ 1 + sound condition + motion condition + awareness + d' + sound condition * motion condition + sound condition * awareness + motion condition * awareness + sound condition * motion condition * awareness + (1 + sound condition | participant)).

3. Results

As we were interested in the extent of cross-modal interaction both in the presence and the absence of awareness, we divided our participant sample into those who detected any visual dot in the first task ("Have you seen any moving dot?") either above chance (aware group) or at chance (unaware group). The accuracy calculation for group classification included both stimulus-present and stimulus-absent trials (each comprising 50 % of total trials). A correct response required either detecting the stimulus when present ('yes' response) or correctly rejecting when absent ('no' response). This balanced design ensured that response bias (e.g., mostly responding 'no' when a stimulus was present) would result in chance-level performance (50 % accuracy). The subdivision of participants into groups was carried out by applying subject-wise one-sample t-tests on mean detection accuracy against chance level (>0.50, one-tailed, p < 0.05), yielding two groups (Fig. 2): at chance performing participants (unaware group, N = 28, 15f, M_{age} = 21.3 ± 1.8) and above-chance performing participants (aware group, N = 52, 32f, M_{age} = 20.9 ± 2.1).

Our linear mixed model examined how visual sensitivity (d') varied across experimental conditions. As expected, the aware group had substantially higher visual sensitivity to the moving dots than the unaware group (β = 2.14, 95 % CI [1.7, 2.5] p < 0.001). This awareness effect interacted with biological motion type (β = 0.13, 95 % CI [0.01, 0.24], p = 0.032), indicating different patterns of biological motion processing (upright or inverted) depending on visual awareness.

Sound presence enhanced overall visual sensitivity (β = −0.08, 95 % CI [−0.13, −0.03], p = 0.002) and interacted with awareness level (β = 0.12, 95 % CI [0.02, 0.21], p = 0.023), showing that auditory information had a differential impact depending on participants' degree of awareness. The key finding emerged in a three-way interaction between

sound congruency, motion type, and awareness level (β = 0.24, 95 % CI [0.008, 0.46], p = 0.043). We examined this interaction through signed-rank tests with Holm correction for multiple comparisons and standardized effect sizes (Cohen's d), comparing upright versus inverted biological motion within each sound condition and awareness group. In the aware group (Fig. 3), upright biological motion showed advantages over inverted motion, whether paired with congruent (Z = 3.16, p_{corr} = 0.034, d = 0.40), incongruent sounds (Z = 3.34, p_{corr} = 0.01, d = 0.46) or in the absence of sounds (Z = 2.19, p_{corr} = 0.044, d = 0.19).

In the unaware group (Fig. 4), visual discrimination of biological motion was enhanced with respect to inverted motion only when paired with congruent sounds (Z = 2.73, p_{corr} = 0.033, d = 0.48), while no significant effects emerged with incongruent sounds (Z = −1.02, p_{corr} = 0.46, d = 0.12) or no sounds (Z = −1.5, p_{corr} = 0.42, d = 0.27).

These results suggest that while aware participants showed detection advantages for upright biological motion with any or no sound, this advantage was present exclusively during congruent sound trials for the participants with minimal or no visual awareness.

Our model also revealed that response times were predictive of visual sensitivity (β = −0.39, 95 % CI [−0.7, −0.01] p = 0.01), with faster responses associated with higher d' values across conditions, suggesting that stronger visual percepts facilitated faster detection responses, regardless of biological motion type, sound and awareness levels. The relationship between response times and visual sensitivity was confirmed by a correlation analysis between these two variables, revealing a small but significant negative correlation ($R_{(79)}$ = −0.13, 95 % CI [−0.22, −0.04], p = 0.004; Fig. S3).

The linear mixed model ran on RTs revealed a significant main effect of sound presence, which generally hastened responses (β = −0.04, 95 % CI [−0.06, −0.02], p < 0.001) and a main effect of biological motion type, with upright biological motion showing generally faster responses (β = −0.02, 95 % CI [−0.03, −0.006], p = 0.004). No other significant effects emerged (β < 0.05, 95 % CI [−0.09, 0.19], p > 0.18). Thus, RTs were overall faster for upright than inverted biological motion, and faster in the presence of any sound, without modulation of sound congruency or awareness level. Plots of mean reaction times as a function of experimental conditions are provided as supplementary figures (Figs. S1, S2).

Lastly, in agreement with the first model, the negative relationship with d' was confirmed as significant (β = −0.03, 95 % CI [−0.06,
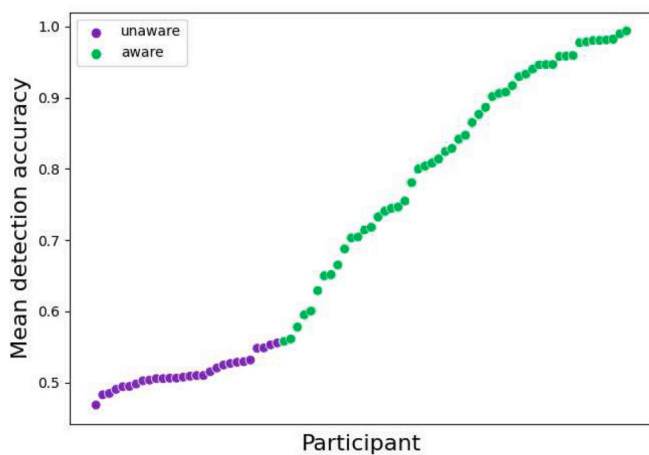

Fig. 2. Sample division into aware and unaware groups based on above or at chance accuracy in detecting any visual stimuli at subject-specific contrast thresholds.
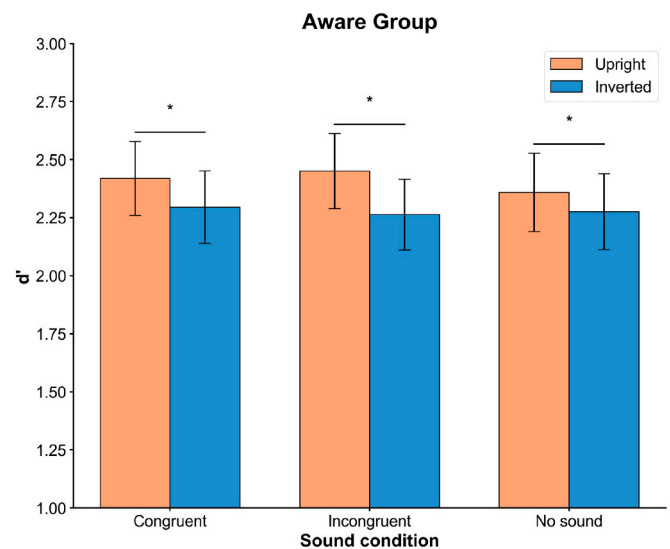

Fig. 3. Barplots with mean d' (y axis) across biological motion type (colors) and sound conditions (x axis) for the aware sample. Error bars represent the within-subjects S.E.M (Cousineau & O'Brien, 2014), * p < .05. For violin plots of the same data, see Supplemental Fig. S4.

Unaware Group

Fig. 4. Barplots with mean d′ values (y axis) across biological motion type (colors) and sound conditions (x axis) for the unaware sample. Error bars represent the within-subjects S.E.M (Cousineau & O'Brien, 2014), * p < .05. For violin plots of the same data, see Supplemental Fig. S5.
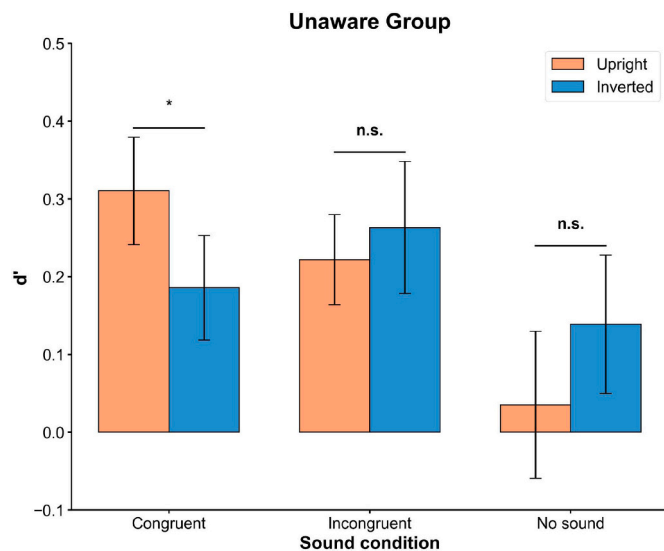
−0.007], p = 0.013), such that lower RTs predicted higher d′ and vice versa.

The differences in visual sensitivity across the aware and unaware group suggest that during awareness, participants relied predominantly on visual information, while during unawareness, participants relied more on auditory information due to the visual stimuli being suppressed. Thus, we examined whether individual differences in visual awareness across the whole sample predicted the extent of auditory influence. We found a moderate significant negative correlation between individual mean detection accuracies and individual differences in d′ between sound presence and absence (Fig. 5; R$_{(78)}$ = −0.3, 95 % CI [−0.49, −0.09], p = 0.007), highlighting that as participants' visual detection performance worsened and reached chance level, the impact of auditory signals tended to increase.

In the second, biological motion categorisation task ("Was the movement rowing, sawing or walking?"), assessed via a one-sample *t*-test against chance level (> 33.3 % accuracy, two-tailed, p < 0.05) participants were not able to report the suppressed movement category above chance in any of the sound conditions, regardless of their group
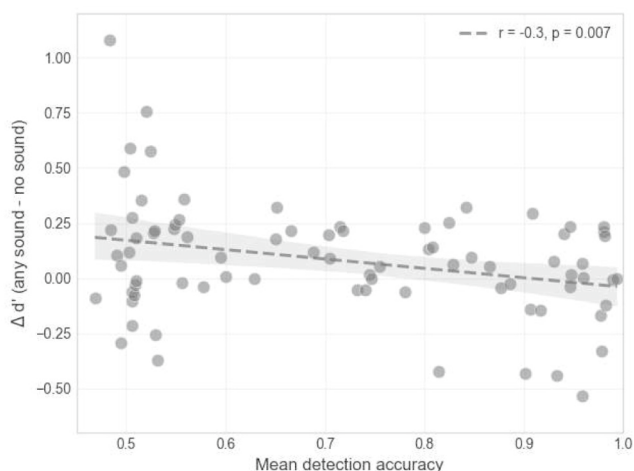

Fig. 5. Scatterplot of the difference in mean d′ values (delta d′) between trials with any sound and no sound (y axis) and mean dot detection accuracy (x axis).

assignment (mean = 34.5 %, std. = 21.5 %, median = 33.3 %). This finding is consistent with our experimental design, since participants were instructed to respond as soon as they detected any moving dot, prioritizing detection over categorisation. Therefore, it is important to distinguish between two types of awareness in our study: awareness of stimulus presence (detection of any moving dots), which we used to classify participants as "aware" or "unaware," and content awareness with recognition of specific biological motion patterns, which was at chance level for both groups. The fact that participants couldn't explicitly identify movement patterns suggests that 1) biological motion categorisation requires more time and visual information than mere dot detection, particularly under noisy conditions of CFS, 2) our paradigm successfully manipulated visual awareness at different levels, and 3) that the observed congruency effects tapped into implicit semantic processing rather than explicit recognition.

4. Discussion

Our findings reveal an interplay between visual awareness and semantic audio-visual integration in biological motion processing. While sounds generally enhanced visual sensitivity - consistent with previous work on cross-modal enhancement (Noesselt et al., 2010) - the nature of this enhancement varied with observers' level of visual awareness.

When any dot of the biological motion stimuli was consciously perceived (i.e. detected above chance level), upright biological motion was detected with higher visual sensitivity than inverted biological motion across all conditions, with concurrent sounds - whether congruent or incongruent - doubling the effect size. The detection advantage of upright biological motion is in line with previous work (Blake & Shiffrar, 2007; Saygin, 2007; Sun et al., 2022; Troje, 2013). Interestingly, our findings of better detection of upright biological motion under no-sound conditions were less prominent than previously reported, with d = 0.19 (our study) as compared to d = 0.56 (Sun et al., 2022). This divergence may be attributed to three methodological factors. First, our larger sample size provides more precise effect size estimates. Second, the introduction of sound conditions in our paradigm may have altered the baseline for visual motion detection, potentially diminishing sensitivity during no-sound trials. Third, we implemented methodological improvements to ensure more robust visual suppression, specifically, a moving dynamic Mondrian mask combined with individual contrast thresholding. These enhancements provided more effective suppression of moving stimuli (Moors et al., 2014) compared to the static CFS mask employed in Sun et al. (2022), allowing us to better isolate the effects of awareness on biological motion processing.

When visual stimuli were suppressed from conscious awareness (i.e. dot detection was at chance), a different result pattern emerged. Although sounds still provided an overall enhancement of visual sensitivity, the detection advantage of upright versus inverted biological motion appeared exclusively during trials with semantically congruent sounds. This finding is in line with previous work demonstrating that semantic congruence can influence visual processing even under conditions of limited awareness (Alsius & Munhall, 2013; Faivre et al., 2014; Lupyan & Ward, 2013; Tan & Yeh, 2015) and highlights the complex interplay between consciousness, semantic processing, and multisensory integration in biological motion perception.

Importantly, while detection of biological motion occurred without explicit categorisation in both groups, only in the unaware group (who was at chance in dot detection) did we observe selective enhancement solely for congruent audiovisual stimuli. This suggests that when basic visual detection is severely compromised, semantic congruence becomes an important factor in cross-modal facilitation. In contrast, when dots were detectable (in the aware group), upright biological motion maintained its detection advantage regardless of sound condition, indicating that once basic visual detection is possible, specialized processing pathways for biological motion operate effectively even without explicit motion categorisation. These findings resonate with discussions on the

extent of high-level processing during CFS (Moors et al., 2019) which emphasize that perceptual processing can occur along a continuum of awareness rather than in binary conscious/unconscious states. Thus, even when explicit biological motion categorisation fails, implicit semantic processing is sufficient to facilitate cross-modal interaction based on semantic congruence in the absence of awareness. Interestingly, our results contrast with some previous work showing limited audio-visual integration during CFS. For instance, Moors et al. (2015) found that suppressed visual looming stimuli were not integrated with auditory looming signals. The difference might lie in the properties of biological motion stimuli, which are processed through specialized neural pathways that may remain partially functional even during interocular suppression. The fact that only upright (but not inverted) biological motion detection was enhanced during congruent sound conditions in the unaware group further supports this interpretation, as upright biological motion maintains its naturalistic configuration while inverted biological motion does not.

The relationship between visual awareness and audiovisual interaction was further supported by two correlational analyses. First, we found that response times predicted visual sensitivity across all conditions, with faster responses associated with higher d′ values. This relationship suggests that stronger visual percepts consistently facilitated quicker detection responses, regardless of experimental conditions. Second, we observed that participants' reliance on auditory information varied systematically with their level of visual awareness. Specifically, as participants' visual detection performance decreased toward chance level, the enhancement of visual sensitivity by sounds (compared to no-sound trials) increased.

The employed experimental approach offers several advantages. Rather than measuring time to breakthrough as in b-CFS (e.g. Alsius & Munhall, 2013; Hong & Shim, 2016; Plass et al., 2014; Yang & Yeh, 2014), or dominance times in binocular rivalry (Conrad et al., 2010; Kang & Blake, 2005; Lunghi et al., 2014; Parker & Alais, 2006), we measured visual sensitivity (d′), providing a direct measure of perceptual processing strength without confounding response bias. This approach also mitigates the potential confounds associated with assessing time to break-through in bCFS paradigms (Alais et al., 2024; Lanfranco et al., 2023; Stein, 2019; Stein et al., 2011). Moreover, our large sample size (N = 80) and unified experimental design allowed us to examine the complex interplay between awareness, biological motion processing, and audiovisual interaction within a single comprehensive model. Therefore, rather than piecing together findings from separate experiments with different paradigms and samples, we provide a more robust and complete understanding of how these factors interact to influence audio-visual perception.

Our results suggest an interesting trade-off between the degree of reliance on visual information and the influence of semantically matching sounds in the discrimination of biological motion. Previous studies showed that a spatially co-occurring tone can thrust a suppressed visual flash into awareness (Aller et al., 2015) and that speech sounds congruent to the movements of a talking face can make that face break through b-CFS faster than incongruent speech (Alsius & Munhall, 2013). Similarly, a congruent tactile stimulus can improve detection sensitivity of a suppressed visual stimulus during binocular rivalry (Lunghi & Alais, 2015). Our observed result patterns across awareness levels reflect fundamental principles of multisensory interaction, particularly the inverse effectiveness principle (Alais & Burr, 2004). For aware participants who could reliably detect some visual information despite CFS, the visual signal was sufficiently robust that they showed an upright biological motion detection advantage regardless of sound condition. In contrast, when visual information was suppressed from awareness, and the visual signal accordingly weak, the auditory signal exerted a stronger influence, but only when semantically congruent with the suppressed visual input. This finding adds nuance to the understanding of stimulus processing during CFS. While CFS has been characterized as causing stimulus fractionation, meaning that visual features are

processed in isolation rather than as coherent wholes during suppression (Moors et al., 2017), our results suggest that certain aspects of complex stimuli, particularly the semantic content and orientation of biological motion, can remain available for cross-modal interaction even during successful suppression from awareness. Our correlation analysis supports this interpretation, showing that the impact of sounds on visual sensitivity systematically increased as participants' visual detection performance decreased.

We extend previous investigations by showing that not only temporo-spatially congruent, but also semantically meaningful sounds selectively allow for a processing advantage of suppressed upright biological motion with respect to inverted biological motion depending on the semantic relatedness to the concurrent sound. That is, not only any sound, but specifically the sound of e.g. rowing selectively enhances the unaware perception of rowing movements in biological motion. This is in agreement with findings that correct verbal labels improve detection of visually suppressed objects, as opposed to incorrect ones (Lupyan & Ward, 2013), or that a congruent soundtrack speeds up the perception of noisy visual scenes during continuous flash suppression (Tan & Yeh, 2015). In both these studies and in ours, combining a clear top-down sound signal with a weak bottom-up visual signal was sufficient to modulate visual sensitivity or emergence into awareness of the semantically congruent visual percept. The directionality of this effect is paramount here, i.e. an unambiguous conscious signal in audition disambiguates, via semantic associations, a noisy or suppressed visual percept, specifically a meaningful biological motion stimulus with respect to a meaningless one. In contrast, other studies have highlighted that a noisy or suppressed signal in one modality cannot necessarily produce cross-modal interactions in another modality and instead requires awareness (Delong & Noppeney, 2021; Montoya & Badde, 2023).

In the present study, the combination of stimulus contrast manipulation with CFS induced weaker, less reliable visual signals (Yuval-Greenberg & Heeger, 2013), which are encoded differently based on their level of awareness (Huang et al., 2023). This likely constrained the propagation of visual information to higher-order association areas of the brain—regions instrumental in conscious perception (Feinstein et al., 2004). Under such high visual uncertainty, we propose that multisensory representations in areas like the parietal cortex (Sereno & Huang, 2014) and perceptual decision-making processes in the frontal cortex (Rahnev et al., 2016) become more strongly influenced by auditory inputs. This ensures that conscious awareness prioritizes the most meaningful, salient, and contextually relevant information. Recently, research has shown that semantically congruent object sounds can improve discrimination of degraded visual objects embedded in noise (Williams et al., 2022; Williams & Störmer, 2024). Our results complement this understanding by showing that in scenarios of visual noise and ambiguity, auditory signals assume greater perceptual weight and play a crucial role in resolving visual ambiguities of stimuli with a naturalistic configuration.

In sum, we demonstrate an interplay between information uncertainty and cross-modal influences, particularly regarding higher-order semantic associations and visual awareness. From an evolutionary perspective, prioritizing meaningful audiovisual cues under visual uncertainty could have been an adaptive survival skill for the detection of threats, like hidden predators, or rewards, such as hidden prey.

CRediT authorship contribution statement

Stefano Ioannucci: Writing – original draft, Visualization, Investigation, Formal analysis, Data curation, Conceptualization. **Petra Vetter:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106181.

Data availability

All stimuli, experimental scripts, analysis and data files are available for download at https://zenodo.org/records/14712975.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology: CB, 14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Alais, D., Coorey, J., Blake, R., & Davidson, M. J. (2024). tCFS: A new 'CFS tracking' paradigm reveals uniform suppression depth regardless of target complexity or salience. *eLife, 12*. https://doi.org/10.7554/eLife.91019.2

Aller, M., Giani, A., Conrad, V., Watanabe, M., & Noppeney, U. (2015). A spatially collocated sound thrusts a flash into awareness. *Frontiers in Integrative Neuroscience, 9*. https://www.frontiersin.org/articles/10.3389/fnint.2015.00016.

Alsius, A., & Munhall, K. G. (2013). Detection of audiovisual speech correspondences without visual awareness. *Psychological Science, 24*(4), 423–431. https://doi.org/10.1177/0956797612457378

Blake, R., & Shiffrar, M. (2007). Perception of Human Motion. *Annual Review of Psychology, 58*(1), 47–73. https://doi.org/10.1146/annurev.psych.57.102904.190152

Blake, R., Sobel, K. V., & James, T. W. (2004). Neural synergy between kinetic vision and touch. *Psychological Science, 15*(6), 397–402. https://doi.org/10.1111/j.0956-7976.2004.00691.x

Bulkin, D. A., & Groh, J. M. (2006). Seeing sounds: Visual and auditory interactions in the brain. *Current Opinion in Neurobiology, 16*(4), 415–419. https://doi.org/10.1016/j.conb.2006.06.008

Calvert, G. (2004). The handbook of multisensory processes. https://mitpress.mit.edu/9780262033213/the-handbook-of-multisensory-processes/.

Chang, D. H. F., Ban, H., Ikegaya, Y., Fujita, I., & Troje, N. F. (2018). Cortical and subcortical responses to biological motion. *NeuroImage, 174*, 87–96. https://doi.org/10.1016/j.neuroimage.2018.03.013

Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition, 114*(3), 389–404. https://doi.org/10.1016/j.cognition.2009.10.012

Chen, Y.-C., & Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1554–1568. https://doi.org/10.1037/a0024329

Chen, Y.-C., Yeh, S.-L., & Spence, C. (2011). Crossmodal constraints on human perceptual awareness: Auditory semantic modulation of binocular rivalry. *Frontiers in Psychology, 2*, 212. https://doi.org/10.3389/fpsyg.2011.00212

Conrad, V., Bartels, A., Kleiner, M., & Noppeney, U. (2010). Audiovisual interactions in binocular rivalry. *Journal of Vision, 10*(10), 27. https://doi.org/10.1167/10.10.27

Conrad, V., Kleiner, M., Bartels, A., O'Brien, J. H., Bülthoff, H. H., & Noppeney, U. (2013). Naturalistic stimulus structure determines the integration of audiovisual looming signals in binocular rivalry. *PLoS ONE, 8*(8), Article e70710. https://doi.org/10.1371/journal.pone.0070710

Conrad, V., Vitello, M. P., & Noppeney, U. (2012). Interactions between apparent motion rivalry in vision and touch. *Psychological Science, 23*(8), 940–948. https://doi.org/10.1177/0956797612438735

Cousineau, D., & O'Brien, F. (2014). Error bars in within-subject designs: A comment on Baguley (2012). *Behavior Research Methods, 46*(4), 1149–1151. https://doi.org/10.3758/s13428-013-0441-z

Delong, P., & Noppeney, U. (2021). Semantic and spatial congruency mould audiovisual integration depending on perceptual awareness. *Scientific Reports, 11*(1), Article 1. https://doi.org/10.1038/s41598-021-90183-w

Deroy, O., Faivre, N., Lunghi, C., Spence, C., Aller, M., & Noppeney, U. (2016). The complex interplay between multisensory integration and perceptual awareness. *Multisensory Research, 29*(6–7), 585–606. https://doi.org/10.1163/22134808-00002529

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron, 57*(1), 11–23. https://doi.org/10.1016/j.neuron.2007.12.013

Faivre, N., Mudrik, L., Schwartz, N., & Koch, C. (2014). Multisensory integration in complete unawareness: Evidence from audiovisual congruency priming. *Psychological Science, 25*(11), 2006–2016. https://doi.org/10.1177/0956797614547916

Faul, F., Erdfelder, E., Lang, A. G., et al. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146

Feinstein, J. S., Stein, M. B., Castillo, G. N., & Paulus, M. P. (2004). From sensory processes to conscious perception. *Consciousness and Cognition, 13*(2), 323–335. https://doi.org/10.1016/j.concog.2003.10.004

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews. Neuroscience, 14*(5), 350–363. https://doi.org/10.1038/nrn3476

Grèzes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C., & Decety, J. (2001). Does perception of biological motion rely on specific brain regions? *NeuroImage, 13*(5), 775–785. https://doi.org/10.1006/nimg.2000.0740

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience, 12*(5), 711–720. https://doi.org/10.1162/089892900562417

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35*(6), 1167–1175. https://doi.org/10.1016/s0896-6273(02)00897-8

Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J., & Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitude-modulation rate. *Current Biology, 22*(5), 383–388. https://doi.org/10.1016/j.cub.2012.01.004

Hense, M., Badde, S., & Röder, B. (2019). Tactile motion biases visual motion perception in binocular rivalry. *Attention, Perception & Psychophysics, 81*(5), 1715–1724. https://doi.org/10.3758/s13414-019-01692-w

Hong, S. W., & Shim, W. M. (2016). When audiovisual correspondence disturbs visual processing. *Experimental Brain Research, 234*(5), 1325–1332. https://doi.org/10.1007/s00221-016-4591-y

Huang, Z., Urale, P. W. B., Morgan, C. A., Rees, G., & Schwarzkopf, D. S. (2023). The role of awareness in shaping responses in human visual cortex. *Royal Society Open Science, 10*(8), Article 230380. https://doi.org/10.1098/rsos.230380

Ioannucci, S., Borragán, G., & Zénon, A. (2022). Passive visual stimulation induces fatigue under conditions of high arousal elicited by auditory tasks. *Journal of Experimental Psychology. General, 151*(12), 3097–3113. https://doi.org/10.1037/xge0001224

Kang, M.-S., & Blake, R. (2005). Perceptual synergy between seeing and hearing revealed during binocular rivalry. *Psichologija, 32*, 7–15.

Klink, P. C., van Wezel, R. J. A., & van Ee, R. (2012). United we sense, divided we fail: Context-driven perception of ambiguous visual stimuli. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 367*(1591), 932–941. https://doi.org/10.1098/rstb.2011.0358

Kvasova, D., Spence, C., & Soto-Faraco, S. (2023). *Not so fast: Orienting to crossmodal semantic congruence depends on task relevance and perceptual load.* https://doi.org/10.20350/digitalCSIC/15662

Lanfranco, R. C., Rabagliati, H., & Carmel, D. (2023). Assessing the influence of emotional expressions on perceptual sensitivity to faces overcoming interocular suppression. *Emotion, 23*(7), 2059–2079. https://doi.org/10.1037/emo0001215

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Liaw, G. J., Kim, S., & Alais, D. (2022). Direction-selective modulation of visual motion rivalry by collocated tactile motion. *Attention, Perception, & Psychophysics, 84*(3), 899–914. https://doi.org/10.3758/s13414-022-02453-y

Lunghi, C., & Alais, D. (2015). Congruent tactile stimulation reduces the strength of visual suppression during binocular rivalry. *Scientific Reports, 5*, 9413. https://doi.org/10.1038/srep09413

Lunghi, C., Binda, P., & Morrone, M. C. (2010). Touch disambiguates rivalrous perception at early stages of visual analysis. *Current Biology, 20*(4), R143–R144. https://doi.org/10.1016/j.cub.2009.12.015

Lunghi, C., Lo Verde, L., & Alais, D. (2017). Touch accelerates visual awareness. *I-Perception, 8*(1), Article 2041669516686986. https://doi.org/10.1177/2041669516686986

Lunghi, C., Morrone, M. C., & Alais, D. (2014). Auditory and tactile signals combine to influence vision during binocular rivalry. *The Journal of Neuroscience, 34*(3), 784–792. https://doi.org/10.1523/JNEUROSCI.2732-13.2014

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences, 110*(35), 14196–14201. https://doi.org/10.1073/pnas.1303312110

Meyer, G. F., Greenlee, M., & Wuerger, S. (2011). Interactions between auditory and visual semantic stimulus classes: Evidence for common processing networks for speech and body actions. *Journal of Cognitive Neuroscience, 23*(9), 2291–2308. https://doi.org/10.1162/jocn.2010.21593

Miles, W. R. (1930). Ocular dominance in human adults. *The Journal of General Psychology, 3*(3), 412–430. https://doi.org/10.1080/00221309.1930.9918218

Montoya, S., & Badde, S. (2023). Only visible flicker helps flutter: Tactile-visual integration breaks in the absence of visual awareness. *Cognition, 238*, Article 105528. https://doi.org/10.1016/j.cognition.2023.105528

Moors, P., Gayet, S., Hedger, N., Stein, T., Sterzer, P., van Ee, R., Wagemans, J., & Hesselmann, G. (2019). Three criteria for evaluating high-level processing in CFS. *Trends in Cognitive Sciences, 23*(4), 267–269. https://doi.org/10.1016/j.tics.2019.01.008

Moors, P., Hesselmann, G., Wagemans, J., & van Ee, R. (2017). Continuous flash suppression: Stimulus fractionation rather than integration. *Trends in Cognitive Sciences, 21*(10), 719–721. https://doi.org/10.1016/j.tics.2017.06.005

Moors, P., Huygelier, H., Wagemans, J., & de-Wit, L., & van Ee, R.. (2015). Suppressed visual looming stimuli are not integrated with auditory looming signals: Evidence

from continuous flash suppression. *i-Perception, 6*(1), 48–62. https://doi.org/10.1068/i0678

Moors, P., Wagemans, J., & de-Wit, L.. (2014). Moving stimuli are less effectively masked using traditional continuous flash suppression (CFS) compared to a Moving Mondrian Mask (MMM): A test case for feature-selective suppression and retinotopic adaptation. *PLoS ONE, 9*(5), Article e98298. https://doi.org/10.1371/journal.pone.0098298

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences, 18*, 488–496. https://doi.org/10.1016/j.tics.2014.04.009

Ngo, M. K., & Spence, C. (2010). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics, 72*(6), 1654–1665. https://doi.org/10.3758/APP.72.6.1654

Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H.-J., & Driver, J. (2010). Sound-induced enhancement of low-intensity vision: Multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *The Journal of Neuroscience, 30*(41), 13609–13623. https://doi.org/10.1523/JNEUROSCI.4524-09.2010

Noppeney, U., Ostwald, D., & Werner, S. (2010). Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *Journal of Neuroscience, 30*(21), 7434–7446. https://doi.org/10.1523/JNEUROSCI.0455-10.2010

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11*(12), 520–527. https://doi.org/10.1016/j.tics.2007.09.009

Palmer, T. D., & Ramsey, A. K. (2012). The function of consciousness in multisensory integration. *Cognition, 125*(3), 353–364. https://doi.org/10.1016/j.cognition.2012.08.003

Pápai, M. S., & Soto-Faraco, S. (2017). Sounds can boost the awareness of visual eventsthrough attention without cross-modal integration. *Scientific Reports, 7*(1), 41684. https://doi.org/10.1038/srep41684

Park, M., Blake, R., & Kim, C.-Y. (2024). Audiovisual interactions outside of visual awareness during motion adaptation. *Neuroscience of Consciousness, 2024*(1), niad027. https://doi.org/10.1093/nc/niad027

Parker, A. L., & Alais, D. M. (2006). Auditory modulation of binocular rivalry. *Journal of Vision, 6*(6), 855. https://doi.org/10.1167/6.6.855

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Petro, L. S., Paton, A. T., & Muckli, L. (2017). Contextual modulation of primary visual cortex by auditory signals. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 372*(1714). https://doi.org/10.1098/rstb.2016.0104

Plass, J., Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Suzuki, S. (2014). Lip reading without awareness. *Psychological Science, 25*(9), 1835–1837. https://doi.org/10.1177/0956797614542132

Plass, J., Guzman-Martinez, E., Ortega, L., Suzuki, S., & Grabowecky, M. (2017). Automatic auditory disambiguation of visual awareness. *Attention, Perception, & Psychophysics, 79*(7), 2055–2063. https://doi.org/10.3758/s13414-017-1355-0

Rahnev, D., Nee, D. E., Riddle, J., Larson, A. S., & D'Esposito, M. (2016). Causal evidence for frontal cortex organization for perceptual decision making. *Proceedings of the National Academy of Sciences, 113*(21), 6059–6064. https://doi.org/10.1073/pnas.1522551113

Salomon, R., Lim, M., Herbelin, B., Hesselmann, G., & Blanke, O. (2013). Posing for awareness: Proprioception modulates access to visual consciousness in a continuous flash suppression task. *Journal of Vision, 13*(7), 2. https://doi.org/10.1167/13.7.2

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*(6), 110–114. https://doi.org/10.2307/3002019

Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain, 130*(9), 2452–2461. https://doi.org/10.1093/brain/awm162

Sereno, M. I., & Huang, R.-S. (2014). Multisensory maps in parietal cortex. *Current Opinion in Neurobiology, 24*, 39–46. https://doi.org/10.1016/j.conb.2013.08.014

Shanks, D. R. (2017). Regressive research: The pitfalls of post hoc data selection in the study of unconscious mental processes. *Psychonomic Bulletin & Review, 24*(3), 752–775. https://doi.org/10.3758/s13423-016-1170-y

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Cross-modal dynamic capture: Congruency effects in the perception of motion across sensory modalities. *Journal of Experimental Psychology. Human Perception and Performance, 30*(2), 330–345. https://doi.org/10.1037/0096-1523.30.2.330

Stein, T. (2019). The breaking continuous flash suppression paradigm: Review, evaluation, and outlook. In G. Hesselmann (Ed.), *Transitions between Consciousness and Unconsciousness.* New York: Routledge.

Stein, T., Hebart, M., & Sterzer, P. (2011). Breaking continuous flash suppression: A new measure of unconscious processing during interocular suppression? *Frontiers in Human Neuroscience, 5.* https://www.frontiersin.org/articles/10.3389/fnhum.2011.00167.

Sun, Y., Wang, X., Huang, Y., Ji, H., & Ding, X. (2022). Biological motion gains preferential access to awareness during continuous flash suppression: Local biological motion matters. *Journal of Experimental Psychology: General, 151*, 309–320. https://doi.org/10.1037/xge0001078

Tan, J.-S., & Yeh, S.-L. (2015). Audiovisual integration facilitates unconscious visual scene processing. *Journal of Experimental Psychology. Human Perception and Performance, 41*(5), 1325–1335. https://doi.org/10.1037/xhp0000074

The jamovi project. (2019). *Jamovi* (1.1) [computer software]. https://www.jamovi.org.

Troje, N. F. (2013). What is biological motion? Definition, stimuli, and paradigms. In *Social perception: Detection and interpretation of animacy, agency, and intention* (pp. 13–36). Boston Review. https://doi.org/10.7551/mitpress/9780262019279.003.0002.

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience, 8*, 1096–1101. https://doi.org/10.1038/nn1500

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences of the United States of America, 98*(20), 11656–11661. https://doi.org/10.1073/pnas.191374198

Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition, 27*, 62–75. https://doi.org/10.1016/j.concog.2014.04.007

Vetter, P., Smith, F. W., & Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. *Current Biology: CB, 24*(11), 1256–1262. https://doi.org/10.1016/j.cub.2014.04.020

Williams, J. R., Markov, Y. A., Tiurina, N. A., & Störmer, V. S. (2022). What you see is what you hear: Sounds alter the contents of visual perception. *Psychological Science, 33*(12), 2109–2122. https://doi.org/10.1177/09567976221121348

Williams, J. R., & Störmer, V. S. (2024). Cutting through the noise: Auditory scenes and their effects on visual object processing. *Psychological Science, 35*(7), 814–824. https://doi.org/10.1177/09567976241237737

Yang, Y.-H., & Yeh, S.-L. (2014). Unmasking the dichoptic mask by sound: Spatial congruency matters. *Experimental Brain Research, 232*(4), 1109–1116. https://doi.org/10.1007/s00221-014-3820-5

Yuval-Greenberg, S., & Heeger, D. J. (2013). Continuous flash suppression modulates cortical activity in early visual cortex. *Journal of Neuroscience, 33*(23), 9635–9643. https://doi.org/10.1523/JNEUROSCI.4612-12.2013

Zhou, W., Jiang, Y., He, S., & Chen, D. (2010). Olfaction modulates visual perception in binocular rivalry. *Current Biology: CB, 20*(15), 1356–1358. https://doi.org/10.1016/j.cub.2010.05.059