Citation: Joly-Mascheroni, R., Abad-Hernando, S., Forster, B. & Calvo-Merino, B. (2021). Embodiment and Multisensory Perception of Synchronicity: Biological Features Modulate Visual and Tactile Multisensory Interaction in Simultaneity Judgements. Multisensory Research, 34(5), pp. 493-510. doi: 10.1163/22134808-bja10020

DOI: 10.1163/22134808-bja10020

Cite as: Joly-Mascheroni R, Abad-Hernando S, Forster B, Calvo-Merino B. Embodiment and Multisensory Perception of Synchronicity: Biological Features Modulate Visual and Tactile Multisensory Interaction in Simultaneity Judgements. Multisens Res. 2021 Feb 1:1-18. doi: 10.1163/22134808-bja10020.

Embodiment and Multisensory Perception of Synchronicity: Biological Features Modulate Visual and Tactile Multisensory Interaction in Simultaneity Judgements

Ramiro Joly-Mascheroni*, Sonia Abad-Hernando, Bettina Forster and Beatriz Calvo-Merino*

City University of London, Northampton Square, EC1V 0HB, London, UK

Received 17 May 2020; accepted 15 December 2020

ORCID:
R Joly-Mascheroni 0000-0002-3539-5810
S Abad-Hernando 0000-0003-2579-8824
B Forster 0000-0001-5126-7854
B Calvo-Merino 0000-0003-4669-4573

* To whom correspondence should be addressed**:** Ramiro Joly-Mascheroni and Beatriz Calvo-Merino, Cognitive Neuroscience Research Unit, Department of Psychology, City, University of London. Northampton Square, EC1V 0HB, London, UK. E-mail: rjoly@city.ac.uk; b.calvo@city.ac.uk

Abstract

The concept of embodiment has been used in multiple scenarios, but in cognitive neuroscience it normally refers to the comprehension of the role of one's own body in the cognition of everyday situations and the processes involved in that perception. Multisensory research is gradually embracing the concept of embodiment, but the focus has mostly been concentrated upon audiovisual integration. In two experiments, we evaluated how the likelihood of a perceived stimulus to be embodied modulates visuotactile interaction in a Simultaneity Judgement task. Experiment 1 compared the perception of two visual stimuli with and without biological attributes (hands and geometrical shapes) moving towards each other, while tactile stimuli were provided on the palm of the participants' hand. Participants judged whether the meeting point of two periodically-moving visual stimuli was synchronous with the tactile stimulation in their own hands. Results showed that in the hand condition, the Point of Subjective Simultaneity (PSS) was significantly more distant to real synchrony (60 ms after the Stimulus Onset Asynchrony, SOA) than in the geometrical shape condition (45 ms after SOA). In experiment 2, we further explored the impact of biological attributes by comparing performance on two visual biological stimuli (hands and ears), that also vary in their motor and visuotactile properties. Results showed that the PSS was equally distant to real synchrony in both the hands and ears conditions. Overall, findings suggest that embodied visual biological stimuli may modulate visual and tactile multisensory interaction in simultaneity judgements.

Keywords: Embodiment, visuotactile interaction, multisensory integration, simultaneity judgement, body, biological motion

1. Introduction

The majority of events we experience every day give rise to sensations in more than one sensory modality. We combine the information provided by our senses, such as the temporal relations between different sensorial events, to construct a single and coherent global percept (De Gelder and Bertelson, 2003; Ernst and Bulthoff, 2004). Depending on the sensory modality with which we perceive these sensorial events, we may experience the occurrences at different speeds, and as a result, the integrated percept can have an impact on whether the sensory events are regarded as having occurred synchronously or asynchronously. This has been generally tested using a Simultaneity Judgement task (see Spence *et al.*, 2003). Previous work, considering judgements of temporal events under multisensory conditions, points to the dominance of auditory signals over visual or tactile modalities (Fendrich and Corballis, 2001, Vroomen *et al*., 2004; Bresciani *et al*., 2006). Multisensory research has also shown that the naturally occurring lags in arrival (physical) and processing (neural) times of different types of information result in differing timings of physical transmission — sound (330 m/s), light (300×10^6 m/s) — as opposed to neural transmission — auditory (10 m/s), visual (50 m/s) and tactile perception (55 m/s). Authors refer to the Horizon of Simultaneity in which audiovisual information arrives synchronously at the primary sensory cortices only if the event occurs at a distance between 10 and 15 metres from the observer. In a shorter distance (less than 10 m), auditory signals would be the first to be processed, over visual and tactile (Keetels & Vroomen, 2012).

Experiments demonstrated that sounds intervening between two lights led to a decline in performance in a temporal order task, and acted as if either the sounds pushed the perception of the lights further apart in time, or the sounds pulled the lights closer together, depending on when the sounds occurred (Morein-Zamir *et al.*, 2003) suggesting a ‘temporal ventriloquism’ phenomenon analogous to spatial ventriloquism.

Furthermore, if participants are requested to judge conflicting multisensory cues, studies showed that a periodic auditory distracter disrupts visual synchronisation, but the converse is not true (Repp and Penel, 2004).

Most of the existing literature on multisensory processing has largely been confined to studies employing rather simplistic visual and auditory events (see Lange *et al.,* 2018). Occasionally, this resulted in the stimulus used being relatively impoverished, often lacking symbolic meaning, and generally possessing minimal ecological value (Smit *et al.*, 2019), such as beeps and flashes, or lights accompanied by an audible click. Studies inspired by the McGurk and Ventriloquist effects (e.g. McGurk and MacDonald, 1976), concentrated on the speech domain (Bertelson *et al.*, 1994; Burnham and Dodd, 2004; Sekiyama *et al.*, 2014). Previous work exploring multisensory integration in general, and simultaneity judgement studies in particular, have implemented a range of stimulus types, such as visual flashes and auditory tones, realistic audiovisual speech (see Keetels and Vroomen, 2005; Arrighi *et al.*, 2006; Enoki *et al.*, 2006; Vatakis and Spence, 2008) and experimental methods such as film soundtracks and their audiovisual integration, or spatial separation between sound and light (see, e.g., Exner, 1875; Dixon and Spitz, 1980; Vroomen, *et al.*, 2004; Vatakis *et al.,* 2008). Some of the most complex stimuli utilised to date have been videos of humans speaking or playing an instrument (see, e.g., Dixon and Spitz, 1980; Hollier and Rimell, 1998; Vatakis and Spence, 2008).

Research has shown that the final percept in a multisensory perception task could be modulated by the individual features of the perceived stimuli. Saygin *et al. (*2008) presented Point Light Displays (PLDs; upright, inverted and scrambled) and asked participants to judge the audio-visual synchrony between a PLD human walking and a beep that either followed the same frequency of the steps of the visual stimuli (match trials) or a different one (mismatch trials). In that study, the authors employed a paradigm that highlights the importance of biological features in multisensory integration. Their bimodal (visual-auditory) novel task showed that participants were more accurate in the upright condition compared with the inverted and scrambled conditions. That advantage disappeared if the upright PLD was not phase-locked with the beep at the first footstep. The authors suggested these results could be attributed to both the human experience in perceiving a simple biological motion, and to the gestalt of the visual stimuli. This study suggests the importance of a motorically encoded perception and biological attributes in multisensory temporal integration. It has been previously shown that biological motion does not necessarily integrate linearly over space and time with constant efficiency, but instead, it adapts to the particular nature of the stimulus in question (Neri *et al.*, 1998).

In Saygin *et al. (*2008), the authors state that the gestalt of the upright walker had a crossmodal consequence for audiovisual temporal judgment, but it is still unclear if similar advantages are found in other sensory modalities. From the existing literature, it is difficult to know whether the physical attributes of the stimuli, coupled with a potential embodiment, could influence temporal judgements between a visual and a tactile event. Here we explored whether biological attributes represented in body-related stimuli would denote a possible relationship between a simultaneity judgement and a potential embodiment. If so, we investigated if this embodiment would be reflected in extra cognitive processes, which could have an effect of delaying the point of subjective simultaneity (PSS). We examined if a plausible motion path, such as hands clapping, would modulate the simultaneity judgement of visual and tactile information, and whether this modulation is present when observing non-biological stimuli. Cognitive approaches to temporal dimensions and human behavioural responses often refer to the ways in which an individual's behaviour emerges from interactions of their own brain, their whole body, and the environment around them. Embodiment is thought of as more than physiological or brain activity (Gibbs, 2005; De Vignemont, 2011) and is constituted by recurring patterns of kinaesthetic, proprioceptive action that provide much of people's felt subjective experience (Medina *et al.*, 2015; Gonzalez-Franco and Berger, 2019). The intuition behind the embodied approach is that cognition is fundamentally integrated with perceptual and motor systems. Such integration does not itself exclude multimodal representations, as long as the function of these representations is to engage appropriate simulations and not to act as independent conceptual representations that help the process in question. In the present study, the task was to perform a simultaneity judgment and, in this context, to assess the possibility of a modulatory effect of an independent conceptual representation, a potential embodiment of body-related visual information in a visuotactile task. The perception of biological and non-biological movement is achieved by a whole sensory experience that encompasses more than just the visual system, but rather in conjunction with several brain areas' activity that is present as a kinaesthetic form and includes all of the aspects of the body in action and the agent observed. "Perception is tightly linked to subjunctive thought processes whereby objects are perceived by imagining how they may be physically manipulated" (Gibbs, 2005).

Over the past two decades, a large portion of research on multisensory perception has typically been focused on vision and audition (e.g., Vroomen and de Gelder, 2000; Shams *et al.*, 2000; Guttman *et al.,* 2005; Zampini *et al.,* 2005b). Audiovisual temporal judgements can be performed between several combinations of features that belong to either the visual domain (e.g., variations in size, luminance, or position) or the auditory domain (e.g., source location, changes in frequency or amplitude; Fujisaki & Nishida, 2007), while work on other sensory modalities has also made important contributions to the field of visual perception in combination with touch and proprioception (Diederich *et al.,* 2003; Azañón & Soto-Faraco, 2008; Smit *et al.*, 2019; Tamé *et al.*, 2019). Findings of multisensory research on auditory perception in combination with touch (Graziano *et al.*, 1999; Kitagawa *et al.*, 2005; Zampini *et al.,* 2007; Simon-Dack & Teder-Sälejärvi, 2008; Fujisaki & Nishida, 2009; Tajadura-Jiménez *et al.*, 2009) have elegantly shown that there are similar principles of multisensory interactions across different sensory modalities, and that particular features of the stimuli may have had an impact on the patterns of those multisensory interactions. Similarly, multisensory interactions have been suggested to have exerted an attenuation of any effect the tactile stimuli may have had when the perception seemed to be accompanied by the embodying of an observed body part (Kilteni & Ehrsson, 2017). Authors such as Alais & Carlile (2005), Spence *et al. (*2003), Körding *et al. (*2007), Guttman *et al. (*2005), and Làdavas & Farnè (2004), showed that visual perception of rhythms, e.g., observing someone tapping, proved to be automatically encoded if that perception belonged to the auditory domain. This finding raised further questions, such as whether a similar specialised mechanism would apply for the processing of a tactile perception that also followed a varying temporal structure.

Surprisingly little empirical work has been conducted to test if the gestalt of a stimulus, that is all of the visual components of the stimuli in question, could have a distinctive impact on the accuracy of judging whether a visual event was simultaneous with a tactile one (Rubichi *et al.,* 2011; Nardini *et al.,* 2013). Here, we explored to what extent the biological attributes of the perceived stimuli could have an impact on the simultaneity judgements we would make concerning those occurrences.

Our working hypothesis was that accuracy on a judgement of synchronicity between two multisensorial events was dependent upon differing levels of embodiment. With this aim, in the first experiment, we manipulated the potential level of embodiment, by employing stimuli conformed by either biological or non-biological attributes (hands and geometrical shapes). The stimuli with biological and non-biological attributes performed a simple movement, similar to that of clapping. We altered the temporal patterns between the visual and tactile events in a bimodal simultaneity judgement task, that has not been explored in this format before. In the second experiment, we explored the specificity of the embodied effect by testing simultaneity judgements between two biological agents in motion. The first condition being again hands, the second condition a body part stimulus that normally does not experience a motion path similar to clapping. To this end, we used images of ears following a similar trajectory path to the apparent motion of the hands clapping. These data will help in explaining a differential effect on visual–tactile interaction during the perception of biological and non-biological stimuli, and the specificity of the embodied response.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Fourteen participants (mean age 21.85, SD 4.14, one male) took part in the experiment. All participants had normal or corrected-to-normal vision, and reported no cognitive, attentional, or neurological abnormalities. Written and informed consent was obtained from all participants and the study was approved by the Psychology Research Ethics Committee at City University London, following guidelines and procedures established in the Declaration of Helsinki.

2.1.2. Stimuli

The images for the visual stimuli were of real hands and geometrical shapes (rectangles), matched in colour, luminosity and size. All sets of images followed a similar horizontal trajectory and were displayed at a 13° visual angle in height when viewed at 60 cm from the computer screen. For the hands condition (H), the visual images were sequential frames of two hands that started on the edges of the screen, met in the middle and then moved back to the starting position, representing a clapping movement as a biological agent in an apparent continuous motion. The hands stimuli were depicted including the forearms, to ensure their biological identity was processed and to enhance any potential embodiment of the hands themselves. Note that the tactile stimulation was provided in the palms of the participants' hands. The size of the geometrical shapes and the ears were determined by, and therefore identical to, the size of the hands excluding the forearms. Therefore, it applied to all stimuli that, when they reached the point of touching each other, the actual area of contact was the same size, regardless of whether they possessed biological features or not. In the geometrical-shapes condition (S), the visual images were frames of two vertical rectangles, that moved through a horizontal apparent motion path, identical to that of the hands, again, both in terms of the apparent movement and the size, starting from the edges of the screen, meeting in the middle and seemingly moving back to the starting position (see Fig. 1A).

The materials and apparatus were identical in Experiment 1 and 2. Participants were requested to seat in a sound-attenuated room, controlled for light and temperature. The tactile stimulators were presented using 12-V solenoids (5 mm in diameter), and were attached with medical tape to the palms of both hands, and placed where the superficial transverse ligament meets the level of the middle finger. The tactile stimuli consisted of a mechanical single brief pulse. The tactile stimulation was a 10 ms tap presented simultaneously to both hands. The vibration intensity was constant, but the tactile tap was presented at various asynchronies. The tactile stimulators were activated when the two 10-V solenoids drove a metal rod with a blunt conical tip in the palmar fascia. In order to mask the sound of the tactile stimulator, participants were requested to use headphones, through which white noise (58 dB SPL) was provided binaurally.

2.1.3. Procedure

We employed a Simultaneity Judgments (SJ) task to assess perceived crossmodal synchronicity between visual and tactile events. The experiment consisted of 336 trials per stimulus and Stimulus Onset Asynchrony (SOA), resulting in a total 8736 trials per participant. On all trials, participants observed a set of images on a computer screen simulating apparent motion. The duration of the visual stimuli was either 50 or 25 ms per image presentation, as stated in the SOA timeline (see Fig. 1A). The images were presented consecutively to give the impression of motion, and shorter SOAs were used close to the synchronicity point to simulate the act of clapping. Tactile stimulation was applied only once per trial during the first 10 ms of the image presentation (see Fig.1E). The tactile Stimulus Onset Asynchronies (SOA) were identical for all conditions and presented randomly at the following time points relative to the middle frame of the two visual images: −275 ms, −225 ms, −175 ms, −125 ms, −75 ms, −25 ms, synchrony (i.e. 0 ms delayed from the visual stimuli), 25 ms, 75 ms, 125 ms, 175 ms, 225 ms, and 275 ms (see Fig. 1A). Stimuli were displayed using E-Prime software (https://www.pstnet.com/eprime.cfm).

Before the experiment began, participants were given a brief verbal description of the study and a set of instructions were displayed on the computer screen. Participants were told that they were going to observe some images, receive a gentle tactile stimulation in the palms of their hands, and that their task was to judge whether or not the meeting point of the visual stimuli in the middle of the screen was simultaneous with the tactile sensation in their hands. They were requested to respond verbally, by saying YES or NO, and the responses were recorded and response times (RTs) were logged using E-Prime. In order for participants to familiarise themselves with the stimuli, and to make sure they understood the task, they began the study with a practice session of two trials.

2.2. Results

Participants' judgements of simultaneity of the visual and tactile stimuli were compared with a normal distribution using the Kolmogorov–Smirnov goodness-of-fit test. Best-fitting Gaussian curves to the data were performed with r^2 values ranging from 0.66 to 0.98 (mean r^2 0.93; SD 0.05). The peaks of these Gaussian curves were taken as an estimate of the Point of Subjective Simultaneity (PSS). Paired *t*-test comparisons were implemented for each experiment from the PSS data per participant and per condition. A paired *t*-test from the standard deviations data was performed to check that significant variance within groups was not due to differences in the variability of responses between stimuli. All *p* values are reported as two-tailed. Results show the mean PSS was higher in accuracy (i.e., closer to synchrony) in the 'rectangles' condition (mean = 45.36 ms, SD = 31.33) compared with the 'hands' condition (mean = 63.57 ms, SD = 33.21). Paired *t*-test comparisons reveal that the difference was significant, t (13) = 3.26, $p = 0.006$ (Fig. 2).

2.3. Conclusion Experiment 1

Overall, there was a consistent and significant difference when comparing simultaneity judgements between tactile stimulation and visual stimuli with or without biological attributes. The PSS for all participants were significantly closer to real synchrony when perceiving geometrical shapes compared to perceiving hands. This suggests that engaging visually with bodily information, coupled with a potential embodiment, modulates visual–tactile simultaneity judgement performance. Possible alternative interpretations of these results should be addressed to explore if the current results are due to the differences in the stimuli attributes, including apparent size of the whole visual percept. To explore the impact of perceiving biological information on this task we developed Experiment 2.

3. Experiment 2

In experiment 2, we compared performance during observation of two visual stimuli with biological attributes (hands and ears) in the same visual–tactile simultaneity judgements task as experiment 1, to assess the impact biological attributes may have on visual–tactile perception. The aim was to explore any difference on the PSS distance to real synchrony when both stimuli are composed of biological attributes such as in the hands and ears conditions.

3.1. Methods

3.1.2. Participants

Fourteen participants (mean age 22.21, SD 4.87, four males) took part in the experiment. All participants had normal or corrected-to-normal vision, and reported no cognitive, attentional, or neurological abnormalities. Written and informed consent was obtained from all participants and the study was approved by the Psychology Research Ethics Committee at City University London, following guidelines and procedures established in the Declaration of Helsinki.

3.1.2. Stimuli

In Experiment 2, the same hands condition (H) as in Experiment 1 was tested against an ears condition (E). The visual display of the ears condition contained images of human ears following the same horizontal motion path and apparent movement as in the hands condition (Fig. 1). The characteristics of the tactile stimulation were the same as in the previous experiment.

3.2. Procedure

We employed the same SJ task as in experiment 1. We assessed perceived crossmodal synchronicity between visual and tactile events; however, this time, the visual images were of stimuli composed with biological attributes (i.e., bodily stimuli; see Fig. 1). On all trials, participants observed the visual images on a computer screen and received a tactile stimulation in the palms of their hands.

3.3. Results

Participants' visual and tactile judgements of simultaneity were compared with a normal distribution using the Kolmogorov–Smirnov goodness-of-fit test. Best-fitting Gaussian curves to the data were performed with r^2 values ranging from 0.86 to 0.98 (mean r^2 = 0.93; SD = 0.03). The peaks of these Gaussian curves were taken as an estimate of the PSS. Paired *t*-test comparisons were implemented for each experiment from the PSS data per participant and per condition. A paired *t*-test from the standard deviations data was performed to check that significant variance within groups was not due to differences in the variability of responses between stimuli. All *p* values are reported as two-tailed. Results show mean PSS was slightly higher in accuracy (i.e. closer to real synchrony) in the ears condition (mean = 23.64 ms, SD = 21.37) compared with the hands condition (mean = 25.86 ms, SD = 22.39). Paired *t*-test comparisons reveal that the difference was not significant, t (13) = 0.36, p = 0.724 (Fig. 2).

3.4. Conclusion Experiment 2

Overall, there was no significant difference when comparing simultaneity judgements of both visual stimuli with biological attributes within participants. The PSS for all participants did not significantly differ in terms of their distance to real synchrony when perceiving ears compared to hands.

4. General Discussion

These studies aimed to investigate how the ability to visually embody a stimulus could modulate multisensory perception. To this end, we compared the effects of processing biological and non-biological visual stimuli on visual–tactile interaction during a simultaneity judgement task.

Our results show that, under certain circumstances of crossmodal perceptual interactions, embodiment-inducing attributes of the stimulus negatively affect timing perception and judgements of synchronicity of occurrence. Data show how visuotactile simultaneity judgments measured by the point of subjective simultaneity were closer to real synchrony in the non-biological condition (shapes) than in the biological or body condition (hands). Interestingly, no differences are found for visuotactile simultaneity judgments when comparing two visual body conditions (hands vs ears). To summarise, when the perception of events demands perception of visual and tactile sensory modalities, and this is coupled with a task to perform a simultaneity judgement, the nature of the visual stimuli (bodily-related or not) has a significant impact on performance.

Our behavioural results are consistent with previous neurophysiological work that showed that an enhancement of ecological validity can have an impact on biological motion perception (Leonardis *et al.*, 2014; Rognini *et al.*, 2019). Traditionally, multisensory integration and the perception of biological motion have been studied as separate psychological domains. Although a variety of stimuli have been employed to test judgements of simultaneity, surprisingly few studies included real-world agents or parts of the body (Virsu *et al.*, 2008) The great majority of studies presented simple light flashes and beeps to test visual and auditory domains (Harrar and Harris, 2005; Chen *et al.*, 2016; 2018).

Previous studies showed that the perception of Point Light Display (PLD) stimuli, as representative of a visual gestalt of a human walker, plays a critical role in their multisensory integration. Therefore, a logical step forward was to test stimuli that enhanced even further the human biological aspect of a PLD, and real images of bodies represented a good candidate. In our investigation, we used real images of human hands, human ears, and geometrical figures such as rectangles, to assess the subjective judgement of simultaneity between the meeting of the images (i.e., touching of the two hands or two rectangles at the centre of the screen) with the tactile stimulation while the tactile perception was either synchronous or asynchronous with the meeting point of the visual stimuli.

Our results confirm the intuition behind the embodied approach. A number of studies using different experimental paradigms and various techniques implicate sensorimotor representations in various cognitive tasks (Toussaint *et al.*, 2010; Ionta *et al.*, 2016). Results here suggest the judgement of simultaneity in a multisensory perception task might invoke a modulation of embodiment induced by the presence of biological features in the stimuli. Extracognitive processes based on object and agents' representations seem to be mutually integrated with perceptual and motor systems and appear to be dependent upon a certain level of embodiment that was not possible to test when the stimuli used were beeps and lights. Findings of both Experiments 1 and 2 suggest participants may engage differently in simultaneity judgements, either when the stimuli possess biological attributes (hand, ears) or not (geometrical shapes). A similar take on this embodiment modulation of visually perceived stimuli has been shown in other domains such as attention (Arslanova *et al.*, 2019), working memory (Galvez-Pol *et al.*, 2018, 2019) or emotion recognition (Pitcher *et al.*, 2008; Sel *et al.*, 2014).

Because of the nature of the apparent movement and the congruency of the visual hands condition and the tactile stimulations on the palm of the hands, there is a possibility that participants resolved to rely on their own internal representation of clapping. From young babies, humans acquire experience of clapping (Spelke, 1976, 1979; Lewkowicz, 1996; Repp, 2004; Kopp and Dietrich, 2013; Kopp, 2014), so it would have felt natural for the participants to rely on their own representation when it came to judging if the events were simultaneous. In this instance, the tactile taps were only simultaneous once every nine times, and therefore the strategy may have been counterproductive. The statistically significant delay in the PSS for the hands, against the rectangles, shows that the discordance between the altered timing of the tactile tap and their visual perception of the hands clapping resulted in too large a discrepancy between their potential internal representation of the clapping action and the most often asynchronous tactile perception. This triggered a delayed sensation of synchronicity well after the actual simultaneous instance at 60 ms after the SOA.

In these studies, we aimed to invoke visual and motor representations of the observed stimuli, restricted by a set of patterns that could be encoded as sequences, and importantly in the case of the hands, which followed a plausible motion pathway. The ears did not comply with this sequence, as they portrayed a simple movement, but following an improbable motion path. Overall, suggesting that the mere presence of bodily stimuli (and not the plausibility of their motion) may be sufficient to engage the embodied process and that may, in turn, modulate the simultaneity judgement. The effect that the potential embodiment of the stimulus may have had in these results, could not have been addressed in previous studies with more simplistic and, importantly, non-biological attributes, such as beeps and lights, which have often been employed in these types of explorations.

There are cognitive models describing neural networks specialised in this type of optic-flow features with increasing complexity along the hierarchy (see Ratcliffe and Newport, 2017 or Giese and Poggio, 2003 for a review), but these models would only apply when the stimuli follow a recognisable motion path, and in the recognition of normal biological movement stimuli. The models predict the existence of neurons in the dorsal pathway that become selectively activated by complex optic-flow patterns that arise for biological movement patterns. In the case of the ears condition, the particular motion path would not be recognised or matched with an internal representation and therefore would have been expected to, even behaviourally, interfere with any perceptual task, such as in this simultaneity or sensorial integration. Furthermore, in the case of the ears condition, it was not possible to rely on a previous experience or a mental representation of two ears meeting and touching each other. The non-reliance on this strategy may serve, at least in part, to explain the reduced effect in Experiment 2.

The results suggest that current models of visuotactile integration need to be extended to account for multisensory integration in dynamic conditions and natural everyday events. Future experiments should explore beyond the constraints of the present study. For example, it would be relevant to know if simultaneity judgements could be additionally modulated or improved when the tactile tap is always congruent with visual perception, e.g., visual perception of ears touching and tactile perception on the ears. On a different direction, it may be necessary to explore if the biological attributes of the perceived stimuli do also modulate the meeting point of those perceived stimuli in a unisensory modality. This would, in consequence, affect the simultaneity judgement point in a multisensory task, but rather in a unisensory modality involving these same stimuli. Finally, of equal interest could be to test how shapes with a different degree of resemblance to a biological stimulus (i.e., the shape formed by the inclusion of more than the forearm) that resembles a human body part, or an avatar (i.e., an artificially designed stimulus) that graphically represents but also differs from a human likeness, would still have a modulatory effect on the perception of simultaneity in a multisensory task.

The present investigation constitutes a small step in a possible new direction, where testable comparisons could be made regarding the positive or negative effects in simultaneity judgements of an enhancement in ecological validity of the stimuli. This study provided a starting point for further examinations of embodiment and its relationship to action perception, multisensory integration and crossmodal asynchronies.

References

Alais, D. and Carlile, S. (2005). Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound, *Proc. Natl Acad. Sci. U. S. A.* **102**, 2244–2247. doi.org/10.1073/pnas.0407034102

Arrighi, R., Alais, D. and Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences, *J. Vision* **6**, 6. doi: 10.1167/6.3.6

Arslanova, I., Galvez-Pol, A., Calvo-Merino, B. and Forster, B. (2019). Searching for bodies: ERP evidence for independent somatosensory processing during visual search for body-related information, *Neuroimage* **195**, 140–149. doi: 10.1016/j.neuroimage.2019.03.037

Azañón, E. and Soto-Faraco, S. (2008). Changing reference frames during the encoding of tactile events, *Curr. Biol.*, **18**, 1044–1049. doi: 10.1016/j.cub.2008.06.045

Barsalou, L. W. (2008) Grounded cognition, *Annu. Rev. Psychol.*, **59**, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Bertelson, P., Vroomen, J., Wiegeraad, G. and de Gelder, B. (1994). Exploring the relation between McGurk interference and ventriloquism, in: *Third International Conference on Spoken Language Processing*, Yokohama, Japan, September 18–22, 1994.

Bresciani, J.-P., Dammeier, F. and Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events, *J. Vision*, **6**, 2. doi.org/10.1167/6.5.2

Burnham, D. and Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect, *Dev. Psychobiol.*, **45**, 204–220. doi.org/10.1002/dev.20032

Burr, D. (1999). Vision: Modular analysis — or not? *Curr. Biol.*, **9**, R90–R92. doi.org/10.1016/S0960-9822(99)80057-8

Burr, D. and Alais, D. (2006). Combining visual and auditory information, *Prog. Brain Res.* **155**, 243–258. doi.org/10.1016/S0079-6123(06)55014-9

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E. and Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation, *Curr. Biol.*, **16**, 1905–1910. doi.org/10.1016/j.cub.2006.07.065

Calvo-Merino, B., Urgesi, C., Orgs, G., Aglioti, S. M. and Haggard, P. (2010). Extrastriate body area underlies aesthetic evaluation of body stimuli. *Exp. Brain Res.* **204**, 447–456. doi.org/10.1007/s00221-010-2283-6

Chen, Y.-C., Shore, D. I., Lewis, T. L. and Maurer, D. (2016). The development of the perception of audiovisual simultaneity, *J. Exp. Child Psychol.*, **146**, 17–33. doi.org/10.1016/j.jecp.2016.01.010

Chen, Y.-C., Lewis, T. L., Shore, D. I., Spence, C. and Maurer, D. (2018). Developmental changes in the perception of visuotactile simultaneity, *J. Exp. Child Psychol* **173**, 304–317. doi.org/10.1016/j.jecp.2018.04.014

De Gelder, B. and Bertelson, P. (2003). Multisensory integration, perception and ecological validity, *Trends Cogn. Sci.* **710**, 460–467.

De Vignemont, F. (2011). Embodiment, ownership and disownership, *Consc. Cogn.* **20**, 82–93. doi.org/10.1016/j.concog.2010.09.004

Diederich, A., Colonius, H., Bockhorst, D., & Tabeling, S. (2003). Visual-tactile spatial interaction in saccade generation. *Experimental Brain Research*, **148,** 328-337. doi. 10.1007/s00221-002-1302-7

Dixon, N. F. and Spitz, L. (1980). The detection of auditory visual desynchrony, *Perception* **9**, 719–727. doi.org/10.1068/p090719

Enoki, K., Washikita, K. and Yamada, M. (2006). Detection threshold of asynchrony between auditory and visual stimuli for various motion patterns of a ball, in: *Proceedings of the 9th Western Pacific Acoustics Conference* (WESPAC IX). Seoul: Acoustical Society of Korea, June 26–28, 2006..

Ernst, M. O. and Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn. Sci*. **8**, 162–169. doi.org/10.1016/j.tics.2004.02.002

Exner, S. (1875). Experimentelle Untersuchung der einfachsten psychischen Processe, *Pflugers Arch. Gesamte Physiol.*, *Menschen Tiere*, **111**, 403–432.

Fendrich, R. and Corballis, P. M. (2001). The temporal cross-capture of audition and vision, *Percept. Psychophys.* **63**, 719–725. doi.org/10.3758/BF03194432

Fujisaki, W. and Nishida, S. (2007). Feature-based processing of audio-visual synchrony perception revealed by random pulse trains, *Vis. Res.* **47**, 1075–1093. doi.org/10.1016/j.visres.2007.01.021

Fujisaki, W. and Nishida, S. Y. (2009). Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception, *Exp. Brain Res.* **198**, 245–259. doi.org/10.1007/s00221-009-1870-x

Galvez-Pol, A., Forster, B. and Calvo-Merino, B. (2018). Modulation of motor cortex activity in a visual working memory task of hand images, *Neuropsychologia* **117**, 75–83. doi.org/10.1016/j.neuropsychologia.2018.05.005

Galvez-Pol, A., Forster, B. and Calvo-Merino, B. (2019). Beyond action observation: neurobehavioral mechanisms of memory for visually perceived bodies and actions, *Neurosci. Biobehav. Rev.* **116**, 508–518.

Gibbs, R. W., Jr (2005). *Embodiment and Cognitive Science*. Cambridge University Press, Cambridge, UK.

Giese, M. A. and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements, *Nat. Rev. Neurosci.* **4**, 179–192. https://doi.org/10.1038/nrn1057

Gonzalez-Franco, M. and Berger, C. C. (2019). Avatar embodiment enhances haptic confidence on the out-of-body touch illusion, *IEEE Trans. Haptics* **12**, 319–326. doi: 10.1109/TOH.2019.2925038

Graziano, M. S. A., Reiss, L. A. J. and Gross, C. G. (1999). A neuronal representation of the location of nearby sounds, *Nature* **397**, 428–430. doi.org/10.1038/17115

Guttman, S. E., Gilroy, L. A. and Blake, R. (2005). Hearing what the eyes see: auditory encoding of visual temporal structure, *Psychol.*, *Sci.* **16**, 228–235. doi.org/10.1111/j.0956-7976.2005.00808.x

Hollier, M. P. and Rimell, A. N. (1998) An Experimental Investigation into Multi-Modal Synchronization Sensitivity for Perceptual Model Development. *Audio Engineering Society Convention 105.* doi:10.1016/j2006.05.078

Harrar, V. and Harris, L. R. (2005). Simultaneity constancy: detecting events with touch and vision, *Exp. Brain Res.* **166**, 465–473. doi.org/10.1007/s00221-005-2386-7

Ionta, S., Villiger, M., Jutzeler, C. R., Freund, P., Curt, A. and Gassert, R. (2016). Spinal cord injury affects the interplay between visual and sensorimotor representations of the body, *Sci. Rep.* **6**, 20144. doi.org/10.1038/srep20144

Keetels, M. and Vroomen, J. (2005). The role of spatial disparity and hemifields in audio-visual temporal order judgments, *Exp. Brain Res.* **167**, 635–640. doi.org/10.1007/s00221-005-0067-1

Keetels, M. and Vroomen, J. (2012). Perception of synchrony between the senses, in: *The Neural Bases of Multisensory Processes*, M. M. Murray and M. T. Wallace (Eds), pp. 147–177. CRC Press/Taylor and Francis.

Kilteni, K. and Ehrsson, H. H. (2017). Body ownership determines the attenuation of self-generated tactile sensations, *Proc. Natl Acad. Sci.. U. S. A.* **114**, 8426–8431. doi.org/10.1073/pnas.1703347114

Kitagawa, N., Zampini, M., & Spence, C. (2005). Audiotactile interactions in near and far space. *Experimental Brain Research*, **166**, 528–537. doi 10.1007/s00221-005-2393-8

Kopp, F. (2014). Audiovisual temporal fusion in 6-month-old infants, *Dev. Cogn. Neurosci.* **9**, 56–67. doi.org/10.1016/j.dcn.2014.01.001

Kopp, F. and Dietrich, C. (2013). Neuronal dynamics of audiovisual synchrony and asynchrony perception in 6-month-old infants, *Front. Psychol.*, **4**, 2. doi.org/10.3389/fpsyg.2013.00002

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B. and Shams, L. (2007). Causal inference in multisensory perception, *PLoS ONE* **29**, e943. https://doi.org/10.1371/journal.pone.0000943

Làdavas, E. and Farnè, A. (2004). Visuo-tactile representation of near-the-body space, *J. Physiol.*, *Paris* **98**, 161–170. doi.org/10.1016/j.jphysparis.2004.03.007

Lange, J., Kapala, K., Krause, H., Baumgarten, T. J. and Schnitzler, A. (2018). Rapid temporal recalibration to visuo–tactile stimuli, *Exp. Brain Res.* **236**, 347–354. doi.org/10.1007/s00221-017-5132-z

Lennie, P. (1998). Single units and visual cortical organization, *Perception* **27**, 889–935. doi.org/10.1068/p270889

Leonardis, D., Frisoli, A., Barsotti, M., Carrozzino, M. and Bergamasco, M. (2014). Multisensory feedback can enhance embodiment within an enriched virtual walking scenario, *Presence (Camb.)* **23**, 253–266. doi.org/10.1162/PRES_a_00190

Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants, *J. Exp. Psychol.*, *Hum. Percept. Perform.* **22**, 1094–1106. doi.org/10.1037/0096-1523.22.5.1094

Livingstone, M. and Hubel, D. (1988). Segregation of form, color, movement and depth: Anatomy, physiology and perception, *Science* **240**, 740–749.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* **264**, 746–748. doi.org/10.1038/264746a0

Medina, J., Khurana, P. and Coslett, H. B. (2015). The influence of embodiment on multisensory integration using the mirror box illusion, *Consc. Cogn.* **37**, 71–82. doi.org/10.1016/j.concog.2015.08.011

Morein-Zamir, S., Soto-Faraco, S. and Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism, *Cogn. Brain Res.* **17**, 154–163. doi.org/10.1016/S0926-6410(03)00089-2

Nardini, M., Begus, K. and Mareschal, D. (2013). Multisensory uncertainty reduction for hand localization in children and adults, *J. Exp. Psychol.*, *Hum. Percept. Perform.* **39**, 773–787. doi.org/10.1037/a0030719.

Neri, P., Morrone, M. C. and Burr, D. C. (1998). Seeing biological motion, *Nature* **395**, 894–896. doi.org/10.1038/27661

Niedenthal, P. M., Mermillod, M., Maringer, M. and Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression, *Behav. Brain Sci.*, **33**, 417–433. doi.org/10.1017/S0140525X10000865

Pitcher, D., Garrido, L., Walsh, V. and Duchaine, B. C. (2008). Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions, *J. Neurosci.* **28**, 8929–8933. doi.org/10.1523/JNEUROSCI.1450-08.2008

Ratcliffe, N. and Newport, R. (2017). The effect of visual, spatial and temporal manipulations on embodiment and action, *Front. Hum. Neurosci.* **11**, 227. doi: 10.3389/fnhum.2017.00227

Repp, B. H. and Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms, *Psychol.*, *Res.* **68**, 252–270. doi.org/10.1007/s00426-003-0143-8

Rognini, G., Petrini, F. M., Raspopovic, S., Valle, G., Granata, G., Strauss, I., Solcà, M., Bello-Ruiz, J., Herbelin, B., Mange, R., D'Anna, E., Di Iorio, R., Di Pino, G., Andreu, D., Guiraud, D., Stieglitz, T., Rossini, P. M., Serino, A., Micera, S. and Blanke, O. (2019). Multisensory bionic limb to achieve prosthesis embodiment and reduce distorted phantom limb perceptions, *J. Neurol.*, *Neurosurg. Psychiatry* **907**, 833–836. doi.org/10.1136/jnnp-2018-318570

Rubichi, S., Riggio, L., Gherri, E. and Nicoletti, R. (2011). Cognitive conflict is an example of action-grounded cognition, *Int. J. Psychol.*, *Stud.* **3**, 28–35. doi: 10.5539/ijps.v3n1p28

Saygin, A. P., Driver, J. and de Sa, V. R. (2008). In the footsteps of biological motion and multisensory perception: judgments of audiovisual temporal relations are enhanced for upright walkers, *Psychol.*, *Sci.* **195**, 469–475. doi.org/10.1111/j.1467-9280.2008.02111

Sekiyama, K., Soshi, T. and Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: a heightened McGurk effect in older adults, *Front. Psychol.*, **5**, 323. doi.org/10.3389/fpsyg.2014.00323

Sel, A., Forster, B. and Calvo-Merino, B. (2014). The emotional homunculus: ERP evidence for independent somatosensory responses during facial emotional processing, *J. Neurosci.* **34**, 3263–3267. doi.org/10.1523/JNEUROSCI.0106-13.2014

Shams, L., Kamitani, Y. and Shimojo, S. (2000). Illusions. What you see is what you hear, *Nature* **408**, 788. doi.org/10.1038/35048669

Simon-Dack, S. L. and Teder-Sälejärvi, W. A. (2008). Proprioceptive cues modulate further processing of spatially congruent auditory information. A high-density EEG study, *Brain Res.* **1220**, 171–178. doi.org/10.1016/j.brainres.2007.09.015

Smit, S., Rich, A. N. and Zopf, R. (2019). Visual body form and orientation cues do not modulate visuo-tactile temporal integration, *PLoS ONE* **14**, e0224174. doi.org/10.1371/journal.pone.0224174

Spelke, E. (1976). Infants' intermodal perception of events, *Cogn. Psychol.*, **8**, 553–560.

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy, *Dev. Psychol.*, **15**, 626–636. doi.org/10.1016/0010-0285(76)90018-9

Spence, C., Baddeley, R., Zampini, M., James, R. and Shore, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one, *Percept. Psychophys.* **65**, 318–328. doi.org/10.3758/BF03194803

Tajadura-Jiménez, A., Kitagawa, N., Väljamäe, A., Zampini, M., Murray, M. M. and Spence, C. (2009). Auditory–somatosensory multisensory interactions are spatially modulated by stimulated body surface and acoustic spectra, *Neuropsychologia* **47**, 195–203. doi.org/10.1016/j.neuropsychologia.2008.07.025

Tamè, L., Azañón, E. and Longo, M. R. (2019). A conceptual model of tactile processing across body features of size, shape, side and spatial location, *Front. Psychol.*, **10**, 291. doi.org/10.3389/fpsyg.2019.00291

Toussaint, L., Robin, N. and Blandin, Y. (2010). On the content of sensorimotor representations after actual and motor imagery practice, *Motor Control* **14**, 159–175. doi.org/10.1123/mcj.14.2.159

Vatakis, A. and Spence, C. (2008). Investigating the effects of inversion on conjugal processing with an audiovisual temporal-order judgment task, *Perception* **371**, 143–160. doi.org/10.1068/p5648

Vatakis, A., Navarra, J., Soto-Faraco, S. and Spence, C. (2008). Audiovisual temporal adaptation of speech: temporal order versus simultaneity judgments, *Exp. Brain Res.* **1853**, 521–529. doi.org/10.1007/s00221-007-1168-9

Virsu, V., Oksanen-Hennah, H., Vedenpää, A., Jaatinen, P. and Lahti-Nuuttila, P. (2008). Simultaneity learning in vision, audition, tactile sense and their cross-modal combinations, *Exp. Brain Res.* **186**, 525–537. doi.org/10.1007/s00221-007-1254-z

Vroomen, J., and Gelder, B. D. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of experimental psychology: Human perception and performance*, **26** (5), 1583. doi.org/10.1037/0096-1523.26.5.1583.

Vroomen, J., Keetels, M., De Gelder, B. and Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony, *Cogn. Brain Res.* **22**, 32–35. doi.org/10.1016/j.cogbrainres.2004.07.003

Zampini, M., Guest, S., Shore, D. I. and Spence, C. (2005b). Audio-visual simultaneity judgments, *Percept. Psychophys*. **67]**, 531–544. doi.org/10.3758/BF03193329

Zampini, M., Torresan, D., Spence, C. and Murray, M. M. (2007). Auditory–somatosensory multisensory interactions in front and rear space, *Neuropsychologia* **45**, 1869–1877. doi.org/10.1016/j.neuropsychologia.2006.12.004

Figure 1. Illustration of the apparent movement of the stimulus and the Stimulus Onset Asynchronies (SOAs). (A) This figure represents the SOA timeline. On all trials, participants observed a set of images on a computer screen simulating apparent motion. The duration of the visual stimuli was 50 ms, except for the images directly before and after synchrony which lasted only 25 ms. The images were presented consecutively in order to give the impression of motion. The tactile stimulation occurs during the first 10 ms of the image presentation, but only once per trial randomly allocated to any of the 13 SOAs. Example of images in the hands (B), shapes (C) and ears (D) condition. Tactile stimulation was applied simultaneously to the palm of the participants' hands (E).

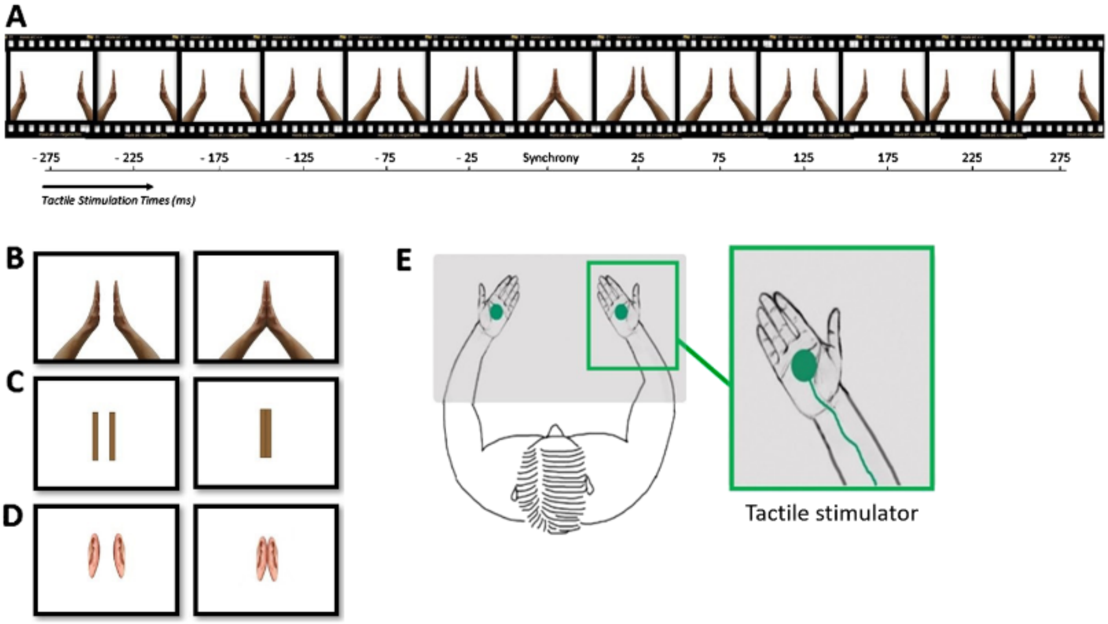

Figure 2. Results from Experiment 1 (Hands and Shapes) and Experiment 2 (Hands and Ears). (A, C) Simultaneity Judgements (SJs) for visuo-tactile stimuli pairs and the best-fitting Gaussian curves were applied to the data (24 trials/data point), showing the proportion of 'simultaneous' responses as a function of delay of the tactile stimulation. The peaks of these curves provide an estimate of the point of subjective simultaneity (PSS). (B, D) Mean values of the PSSs for the visuo-tactile stimulus pairs presented in Experiment 1. PSS is plotted on the *y*-axis for each visual stimulus. Error bars represent SEM.