
Article:
Geangu, Elena orcid.org/0000-0002-0398-8398, Roberti, Elisa and Turati, Chiara (2021) Do infants represent human actions cross-modally? An ERP visual-auditory priming study. Biological psychology. 108047. ISSN 0301-0511

https://doi.org/10.1016/j.biopsycho.2021.108047

Biological Psychology

Do infants represent human actions cross-modally? An ERP visual-auditory priming study.

--Manuscript Draft--

Manuscript Number:	BIOPSY-D-20-00322R3
Article Type:	Research Paper
Section/Category:	Event-Related Potential
Keywords:	infancy, multisensory, human action, auditory, visual, ERPs
Corresponding Author:	Elena Geangu, PhD University of York York, UNITED KINGDOM
First Author:	Elena Geangu, PhD
Order of Authors:	Elena Geangu, PhD
	Elisa Roberti
	Chiara Turati
Abstract:	Recent findings indicate that 7-months-old infants perceive and represent the sounds inherent to moving human bodies. However, it is not known whether infants integrate auditory and visual information in representations of specific human actions. To address this issue, we used ERPs to investigate infants' neural sensitivity to the correspondence between sounds and images of human actions. In a cross-modal priming paradigm, 7-months-olds were presented with the sounds generated by two types of human body movement, walking and handclapping, after watching the kinematics of those actions in either a congruent or incongruent manner. ERPs recorded from frontal, central and parietal electrodes in response to action sounds indicate that 7-months-old infants perceptually link the visual and auditory cues of human actions. However, at this age, these percepts do not seem to be integrated in cognitive multimodal representations of human actions.
Suggested Reviewers:	Moritz Daum, PhD Professor, Universitat Zurich daum@psychologie.uzh.ch expert in developmental psychology
	Monica Gori, PhD Italian Institute of Technology monica.gori@iit.it expert in multisensory integration development
	Stefanie Peykarjou, PhD Heidelberg University stefanie.peykarjou@psychologie.uni-heidelberg.de expert in infant social development and ERP methodology
	Laurie Bayet, PhD American University bayet@american.edu expert in infant socio-emotional development and ERP methodology
Response to Reviewers:	

Dear Prof. Dien,

Thank you and the reviewer for dedicating time and effort to review our manuscript entitled “Do infants represent human actions cross-modally? An ERP visual-auditory priming study.”, which we submitted for consideration to Biological Psychology journal. Your constructive and supportive comments are very much appreciated.

We further revised the manuscript in light of your comments. Please find below detailed information about how we addressed each of the points that were raised.

We hope our responses are satisfactory, and the revised manuscript is according to the journal’s standards.

Best wishes

Elena Geangu, Elisa Roberti and Chiara Turati

Response to Reviewers

Reviewer #1

The authors did a good job addressing many of the comments and suggestions from me and from the second reviewer.

Response: Thank you for the appreciative comments.

However, their revisions mainly involve changes to the introduction and discussion; they chose not to follow many of the requests or suggestions raised regarding data analysis and ERP methodology.

Response: Revisions were made to the methods and results sections as well. For example, we clarified our analyses and terminology, we analysed the optic flow in the PLDs stimuli, we included examples of the stimuli, we provided new information about the variance in the number of trials, and we revised the figures as requested. For a minority of issues, we clarified or justified why we did not implement the suggested changes (e.g., Comment 13 and 14). Specifically, for the three issues Reviewer 1 highlighted here (i.e. inter-editor reliability, variance in rejected channels, and correction for multiple comparisons), we had originally provided justifications for why we did not implement the suggestions in the revision.

For instance, in my comment #9, the authors assume that ERPs are "robust to interrater reliability" and declined to calculate this. There is no evidence that interrater reliability is not an issue in infant ERP data processing; the cited study by Dominguez-Martinez actually shows the opposite. Calculating and reporting this can only strengthen confidence in the findings reported here, so I see no reason why the authors would refuse other than to save time. I still think this would be a good idea given concerns about the robustness/replicability of the results.

Response: We analysed the reliability with which we rejected/included trials in the analysis within a subsample of our participants. Five participants with a total of 529 trials were inspected again by a second coder, whose decisions for which trials to be included/rejected were made following the same criteria reported in the manuscript. Cohen Kappa was used to calculate reliability based on cases of agreement and disagreement between the two data editors. At a Kappa = .9, the two coders are in near perfect agreement. This is unsurprising given that according to our data editing protocols,

although independent editing takes place, this follows agreed and strict criteria. This information is now included in the revised manuscript on p. 16

Similarly, in response to my #10 comment, the authors refuse to report information about how they handled noisy/bad channels and whether these differed between conditions. I am puzzled by the justification that this is not a requirement in the field, as I frequently read ERP papers that report variance in rejected channels in designs with multiple conditions. To me, this raises a red flag.

Response: We searched papers published recently, and as far as we can tell, what is systematically reported for how noisy/bad channels are handled is the maximum number of bad channels that is the threshold for keeping/rejecting a trial and further subjected to interpolation (please see example references below). We reported this information in our original manuscript. To our knowledge, there is no clear guidance for the extent of the difference in the number of bad channels subjected to interpolation that would influence the analysis of the ERPs, and whether this may vary as a function of sample size and/or the total number of channels used in a study. Therefore we had not reported variance in our previous revision. Nonetheless, to our best understanding of the reviewer's request, we revised the manuscript to include more information about the bad channels that were further subjected to interpolation as follows:

(p. 16) "On average across conditions, for the trials included in the final analysis, 8.5 (SD = 3.0) bad channels were interpolated. For each condition, the number of interpolated bad channels varied as follows: $M_{Clap\text{-}Clap}$ = 9.3 (SD = 2.6; Min = 1.5; Max = 17. 5); $M_{Clap\text{-}Walk}$ = 8.5 (SD = 3.4; Min = 1.6; Max = 16.2); $M_{Walk\text{-}Clap}$ = 8.6 (SD = 2.7; Min = 2.0; Max = 16.5); $M_{Walk\text{-}Walk}$ = 7.7 (SD = 3.3; Min = 1.1; Max = 15.4). There was a small significant difference (M_{diff} = 1.67, p = .004) between Clap-Clap and Walk-Walk trials. No significant differences emerged between these two conditions in the ERP analysis as described in the results section. The channels used for the detection of eye movements (32, 25, 21, 17, 14, 8 and 1) were marked as bad for the entire data set and rejected from any further analysis (not subjected to interpolation)."

[1] Arslan, M., Warreyn, P., Dewaele, N., Wiersema, J. R., Demurie, E., & Roeyers, H. (2020). Development of neural responses to hearing their own name in infants at low and high risk for autism spectrum disorder. *Developmental Cognitive Neuroscience*, *41*, 100739.

[2] Conte, S., Richards, J. E., Guy, M. W., Xie, W., & Roberts, J. E. (2020). Face-sensitive brain responses in the first year of life. *NeuroImage*, *211*, 116602.

[3] Ventura-Bort, C., Wirkner, J., Dolcos, F., Wendt, J., Hamm, A. O., & Weymar, M. (2019). Enhanced spontaneous retrieval of cues from emotional events: An ERP study. *Biological Psychology*, *148*, 107742.
[4] Smith, E. S., Crawford, T. J., Thomas, M., & Reid, V. M. (2020). The Influence of Maternal Schizotypy on the perception of Facial Emotional Expressions during Infancy: an Event-Related Potential Study. *Infant Behavior and Development*, *58*, 101390.
[5] Sirri, L., Linnert, S., Reid, V., & Parise, E. (2020). Speech intonation induces enhanced face perception in infants. *Scientific reports*, *10*(1), 1-9.

Finally, in my comment #13 and comment #4 from Reviewer 2, we both raised the issue of multiple comparisons in the authors' ANOVA approach and they also declined to make any changes regarding this issue. I do not find their response convincing—that they do not need to correct for inflated p-values because they only performed planned comparisons.

Response: In our response we stated:

"In order to balance between possible Type 1 and Type 2 errors, we performed only planned comparisons as indicated by our hypotheses, and the p-values are reported uncorrected".

There is sufficient literature for us to perform planned comparisons that are relevant for testing the study hypotheses, rather than all possible comparisons leading to Type 2 errors. We reported the p values in the most transparent way of reporting the data, in the uncorrected version. This allows the readers to apply any corrections they want and interpret the results in light of that. We revised the manuscript to include a correction for false discovery rate (Benjamini-Hochberg), as per Reviewer 2's suggestions, in addition to the uncorrected p-values (p. 21-24). We also included a statement about this approach:

(p. 18) "The p-values are reported both uncorrected and with the Benjamini-Hochberg correction for false discovery rate (Benjamini-Hochberg, 1995) in parentheses."

Highlights

- We investigated 7-months-old infants' multimodal representations of human actions using ERPs
- Infants process human handclapping and walking action sounds as distinct
- Infants perceptually link videos and sounds of human walking
- Infants do not appear to have cognitive auditory-visual representations of human actions

Do infants represent human actions cross-modally? An ERP visual-auditory priming study.

Geangu, Elena[a*#]; Roberti, Elisa[b,c#]; and Turati, Chiara[b,c#]

[a]University of York, Heslington, York, YO10 5DD, United Kingdom,
[b]Università degli Study di Milano - Bicocca, Piazza dell'Ateneo Nuovo, 1, 20126, Milan, Italy
[c]NeuroMi, Milan Center for Neuroscience, Milano, Italy
[#]Authors are presented in the order of their contribution

[*]Corresponding author: Geangu Elena
University of York
Heslington, York, YO10 5DD, UK
Email: elena.geangu@york.ac.uk

Roberti Elisa
[b]Università degli Study di Milano – Bicocca
[c]NeuroMi, Milan Center for Neuroscience, Milano, Italy
Piazza dell'Ateneo Nuovo, 1, 20126
Milan, Italy
e.roberti@campus.unimib.it

Turati Chiara
[b]Università degli Study di Milano – Bicocca
[c]NeuroMi, Milan Center for Neuroscience, Milano, Italy
Piazza dell'Ateneo Nuovo, 1, 20126
Milan, Italy
chiara.turati@unimib.it

Abstract

Recent findings indicate that 7-months-old infants perceive and represent the sounds inherent to moving human bodies. However, it is not known whether infants integrate auditory and visual information in representations of specific human actions. To address this issue, we used ERPs to investigate infants' neural sensitivity to the correspondence between sounds and images of human actions. In a cross-modal priming paradigm, 7-months-olds were presented with the sounds generated by two types of human body movement, walking and handclapping, after watching the kinematics of those actions in either a congruent or incongruent manner. ERPs recorded from frontal, central and parietal electrodes in response to action sounds indicate that 7-months-old infants perceptually link the visual and auditory cues of human actions. However, at this age these percepts do not seem to be integrated in cognitive multimodal representations of human actions.

Introduction

Listening to the sounds people make while they move and act upon the surrounding environment provides us with rich social information. Very often, before seeing a person entering the room, human adults can detect based on the sound of the footsteps who that person is, whether that person is female or male (Bartsch, van der Zwan, Cottrell, & Brooks, 2007; Li, Logan, & Pastore, 1991), and even how that person may be feeling (Sievers et al., 2013). Furthermore, the addition of this auditory information to visual cues about human actions gives adults an advantage for navigating their social environment. Auditory-visual information, rather than visual or auditory information alone, about people's actions best supports how adults coordinate their actions (Bischoff et al., 2014) and how well they can learn from each other (Haslinger et al., 2005; Hauf, Elsner & Aschersleben, 2004; Murgia et al., 2016; O, Law, & Rymal, 2015).

A substantial body of literature indicates that the presence of multi-sensory information is also important for how infants perceive different aspects of their environment (Robinson & Sloutsky, 2010). A basic multisensory integration is already present at birth (Lewkowicz, Leo & Simion, 2010), and develops steadily in the first year of life. At first, different modalities are associated through low-level cues such as temporal synchrony (Bahrick & Lickliter, 2000, 2002; Bremner, 2017; Bremner, Lewkowicz, & Spence, 2012; Geangu, 2009; Lewkowicz, 2000; Patterson & Werker, 2003). As infants grow older, their expanding motor, cognitive and social abilities allow them an increasingly active exploration of the multimodal environment that surrounds them (Liszkowski & Tomasello, 2011; Nishiyori, Bisconti, Meehan & Ulrich, 2016; van Elk, van Schie, Hunnius, Vesper & Bekkering, 2008). As a consequence of this acquired experience, infants no longer solely rely on low-level attributes, but become capable to detect higher level multimodal cues (Walker-Andrews, 1983; Lewkowicz & Ghazanfar, 2009), such as affect (Kahana-Kalman & Walker-Andrews, 2001) and gender (Patterson & Werker, 2002). The existent evidence indicates that perceptual discrimination is enhanced by multimodal stimuli also in younger infants. For instance, 4-5-months olds are better able to discriminate their mother's

face from that of a stranger when they can hear the sound of her voice (Burnham, 1993), and they are better able to detect changes in affective prosody when both face and voice are present compared to voice alone (Flom & Bahrick, 2007).

Although human action sounds are present in infants' acoustic environment from birth, we have limited evidence for how the infant brain develops to process and integrate this information into their multimodal representations of people (Geangu, Quadrelli, Lewis, Macchi Cassia, & Turati, 2015; Quadrelli & Turati, 2016; Quadrelli, Geangu, & Turati, 2019). The aim of the present study is to investigate the neurocognitive processes involved in relating auditory and visual cues inherent to the human body movement, in order to better understand how infants develop multisensorial representations of human actions.

Already at birth, infants' auditory system is sufficiently developed to support the segregation of concurrent streams of sounds, and hence prepared for perceiving and representing distinct social sounds from their surrounding environment (Hepper & Shahidullah, 1994; Draganova et al., 2018; Graven & Browne, 2008; Winkler et al., 2003). They also have well developed abilities to process acoustic properties such as intensity and frequency, temporal relations, and melody (Baruch et al., 2004; Berg & Boswell, 1998; Nazzi et al., 1998; Plantiga & Trainor, 2005; Trainor & Trehub, 1992), which contribute to the extraction of the complex acoustic features and their integration into coherent percepts (Geangu et al., 2015; Gervain, Werker, Black & Geffen, 2015; Gervain, Werker, & Geffen, 2014; Gervain & Geffen, 2019). Importantly, already at birth, the infant brain appears to process those acoustic properties that are relevant for the efficient discrimination and perceptual categorization of natural sounds, such as the similarity in the acoustic patterns at different levels of observation, or scale-invariance (Gervain, Werker, Black & Geffen, 2015; Gervain, Werker, & Geffen, 2014; Gervain & Geffen, 2019). For example, Gervain and colleagues (2016) showed that the newborns' hemodynamic responses recorded at scalp locations corresponding to brain areas involved in the auditory processing of temporally modulated events (i.e., left temporal cortex) and the extraction of

perceptual categories (i.e., the left frontal cortex, with Broca's area) differentiate between scale-invariant and variable-scale sounds. At the age of 5-months, infants show behavioural responses which suggest that they are able to form perceptual categories of scale-invariant but not variable-scale sounds (Gervain et al., 2014). A pattern of neural responses consistent with the perception and representation of the sounds inherent to moving human bodies as belonging to a distinct super-ordinate category of 'living' entities has also been described in infants (Geangu et al., 2015a; Quadrelli et al., 2019). ERP studies suggest that by the age of 7-months, infants appear to process human action sounds, such as footsteps and handclapping, as a distinct grouping of human sounds, alongside human vocalizations. For instance, infants' ERPs recorded at left fronto-temporal scalp regions, which were proposed to reflect enhanced sensory processing (de Haan & Nelson, 1997, Grossmann et al., 2006), differentiate between human action sounds and other types of natural and mechanical sounds (Geangu et al., 2015a). Further differentiations between human action sounds and other types of sounds were also found for the ERP components known to be associated with global-level category formation (i.e., the frontal late positive component – LPC, Quinn, Westerlund & Nelson., 2006), perhaps reflecting the integration of the many stimulus features (e.g., animacy and action features) in distinctive representations, and the ERP components shown to discriminate between visual depictions of biological and non-biological motion (i.e., the parietal negative slow wave - NSW, Hirai et al., 2003, Hirai and Hiraki, 2005, Marshall and Shipley, 2009). By the age of 14-months, the sounds of specific human actions, such as walking and hand-clapping, appear to elicit distinct patterns of sensorimotor activation as indicated by the μ rhythm suppression at central scalp regions (Quadrelli et al., 2019). Taken together, the existent evidence indicates that from an early age, the infant brain processes the acoustic properties characteristic to natural sounds, and from at least the age of 7-months, infants group human action sounds into categories that are distinct from other natural and mechanical sounds (Geangu et al., 2015a; Quadrelli et al., 2019).

Infants are sensitive to how the images and sounds of human body movements synchronise in time, as revealed by studies using point light displays (PLD). PLDs are obtained by representing human movement through points of light placed on the main joints (Johansson 1973), in the absence of any other information about body shape. As such, PLDs specifically convey information about the biological motion characteristic to performing those actions. These stimuli are readily perceived as depicting biological motion even by newborns (Simion, Regolin & Bulf, 2008; Bardi, Regolin & Simion, 2011; Bidet-Ildei, Kitromilides, Orliaguet, Pavlova & Gentaz, 2014), while older infants appear to extract richer information about the type of movement performed by human bodies (Marshall & Shipley, 2009; Missana, Rajhans, Atkinson & Grossmann, 2014; Bhatt, Hock, White, Jubran & Galati, 2016), such as the emotional expression (Missana et al., 2014). Falck-Ytter and colleagues (2011) presented 5-months-old infants with video recordings of point light displays (PLDs) depicting the kinematics of a human adult clapping hands together with the corresponding audio recordings of that action. The synchronicity with which the auditory and visual information occurred was manipulated, as well as the defining dynamic features of biological motion. Infants showed a visual preference for the PLDs where the kinematics specific for hand clapping were intact and which also occurred in temporal synchrony with the corresponding sound. Furthermore, the ERPs in response to hand clapping sounds presented in temporal synchrony and asynchrony with the video recording of the action indicate that the cross-modal temporal information is extracted at pre-attentive sensorial level and integrated into a coherent percept relatively fast (Kopp, 2014; Kopp & Dietrich, 2013). In particular, the auditory ERP components associated with these functions and occurring within 300 ms from stimulus onset (N100 and P200) tended to peak faster and have higher amplitude when infants perceive hand clapping sounds in temporal synchrony with the corresponding visual stream compared to recordings where this temporal relation was perturbed (Kopp, 2014; Kopp & Dietrich, 2013). Therefore, it seems that infants extract the amodal temporal properties from both visual and auditory depictions of human actions, and are sensitive

to crossmodal violations in this information. When the audiovisual synchrony is present, infants can better discriminate between similar human actions (Bahrick, Walker, & Neisser, 1981).

However, while these studies provide important information about infants' sensitivity to the multimodal nature of human actions, they have primarily investigated infants' responses to one single type of action, and this was mainly to hand clapping.

As such, it is not known whether infants can integrate auditory and visual information into multi-modal representations of specific human actions. For example, it remains unknown whether infants process the visual information of a walking person as belonging to the sound of footsteps as opposed to the sound coming from clapping hands, or whether the video activates a cognitive multi-modal representation of walking actions. In this regard, cross-modal priming paradigms are particularly useful for studying the effects of multimodal representation on stimulus processing.

Cross-modal priming paradigms have been highly informative for determining if visual and auditory information can activate amodal social representations (e.g., Bristow et al., 2008; Grossmann, Striano, & Friederici, 2006; Friedrich & Friederici, 2004). In these paradigms, prime and target events are presented in different sensorial modalities, but refer to the same specific social dimension of interest. For example, a prime stimulus may convey visual information about the expression of a discrete emotion (e.g., facial expression of emotion), while the following target stimulus conveys emotional information in the auditory domain (e.g., vocal prosody). ERP studies with both adults and infants which manipulated the congruency between the information provided by the two stimulus events, have shown that the processing of the target can be influenced by the information conveyed by the prime both in adults (e.g., Carroll & Young, 2005; Paulmann & Pell, 2010; Steinbeis & Koelsch, 2011) and infants (e.g., Grossmann et al., 2006; Bristow et al., 2009; Friedrich & Friederici, 2004). These effects demonstrate that individuals may categorize their surrounding environment relying on representations that encompass both

the visual and the auditory information reflected by the prime and target. According to a spreading activation account (Bower, 1991; Fazio, Sanbonmatsu, Powell, & Kardes, 1986), the prime activates the mental representation reflecting links previously established between the prime and the target. As a consequence, the information corresponding to the matching target is more accessible than for the unrelated targets (Kiefer, 2002). Several ERP components have been documented to reflect how the visual context created by the prime influences the processing of the auditory target. Similar to the effects reported by Kopp and colleagues (2013; 2014), some studies with infants and adults have reported significant priming effects already at the sensorial and perceptual stages of processing the auditory targets, as reflected in the frontocentral N100 and P200 (e.g., Friedrich & Friederici, 2011; Friedrich & Friederici, 2015; Hyde et al., 2011; Garrido-Vasquez et al., 2018; Paulmann & Pell, 2010; Yeh et al., 2016). For example, in 6-months-old infants, spoken words primed congruently by images of objects typically reference to by that word lead to increased amplitude in the time window corresponding to the N100-P200 compared to words preceded by the image of non-corresponding objects (Friedrich & Friederici, 2011). These effects have been interpreted to reflect the learned perceptual association between the visual and auditory information, and the fact that the presentation of the prime builds the expectation for a specific acoustic information to occur (Friedrich & Friederici, 2004; Paulmann & Pell, 2010; Garrido-Vasquez et al., 2018).

There are also late ERP components that have been shown to index amodal stages reflecting auditory-visual integration. Similar to the parietal N400 component found in adults to reflect semantic relations between the prime and the target (Kiefer, 2002; Kutas & Federmeier, 2011), an ERP component with negative polarity was also described in infants and young children, however at later time windows (400 - 900 ms), particularly when the auditory targets are longer (Friedrich & Friederici, 2011). Although it is more difficult to establish the extent to which this infant late parietal negative component (labelled here as LNC) indexes the same semantic processing as the N400 described in adults (Juottonen, Revonsuo & Lang, 1996; Federmeier,

Van Petten, Schwartz & Kutas, 2003), systematic investigations in infant word acquisition have shown that the infant parietal LNC is more likely to reflect cognitive representations that integrate the visual and auditory information, rather than simple perceptual associations (Friedrich & Friederici, 2011; 2017). The parietal LNC has been reported in several infant studies to be sensitive to the congruency between social sounds and images (Friedrich & Friederici, 2004, 2005; Grossmann, Striano & Friederici, 2006; Kushnerenko et al., 2008). ERP components with late latency recorded at frontal and central locations that were previously associated with categorical representations in infants, including auditory category distinctions (e.g., LPC; Geangu et al., 2015a; Quinn, Westerlund & Nelson., 2006), were also reported to be sensitive to priming. For example, both human emotional vocalizations (Grossman et al., 2006) and vowel utterances (Bristow et al., 2009) tend to elicit in infants increased positive amplitude of the LPC when they are primed by illustrations of the facial movements typically associated with their production. Spoken words primed by images of the objects they typically refer to also elicit in toddlers increased LPC amplitude compared to incongruently primed words (Friedrich & Friederici, 2004).

To our knowledge, no study has yet investigated whether infants group auditory and visual information into basic-level representations of specific human actions. Considering the relevance of multisensorial representations of social agents for social development, the present study investigated the contextual effects that the visual depictions of human action kinematics have on 7-months-old infants' ERP responses to human action sounds. In a cross modal priming paradigm, infants listened to human hand clapping and walking (i.e., footsteps) sounds which were preceded congruently or incongruently by PLDs depicting those actions. We focused on these two types of human actions for several reasons. First, there is a high chance that most infants of this age have experience with perceiving audio-visually these actions (Geangu et al., 2015a). Within each modality separately, by the age of 7-months, infants also appear to distinguish either between different human body movements and postures (visual, e.g., Geangu

& Vuong, 2020; Geangu, Senna, Croci, Turati, 2015b; Ichikawa, Kanazawa, & Yamaguchi, 2011; Marshall and Shipley, 2009; Missana et al., 2014) and/or between human actions and other categories of living entities (auditory, e.g., Geangu et al., 2015a). Furthermore, infants seem to be sensitive to certain relations between the information about human actions conveyed through these two modalities, such as temporal synchrony (Bahrick et al., 1981; Falck-Ytter et al., 2011; Kopp, 2014; Kopp & Dietrich, 2013). In order to minimize possible effects of temporal synchrony, we chose for the present study to not overlap the presentation of the visual prime and auditory target stimuli (Schirmer, Kotz & Friederici, 2002; Friedrich & Friederici, 2004; 2005; von Koss Torkildsen, Syversen, Simonsen, Moen & Lindgren, 2007), and to include multiple exemplars of PLDs and action sounds that were only related in terms of the type of action they represented. In line with a spreading activation account (Bower, 1991; Fazio et al., 1986), if infants represent human actions by integrating both visual and auditory information, we predict that the neural responses to the target human action sounds will be influenced by the prior presentation of the point light displays (PLDs) depicting the kinematics of those actions. Given the previous priming effects reported in infants and adults (Bristow et al., 2009; Kopp, 2014; Kopp & Dietrich, 2013; Friedrich & Friederici, 2004; 2005, 2017; Pizzamiglio et al., 2005), we expect increased amplitude of the fronto-central N100, P200, and LPC components in response to hand clapping and footstep sounds primed congruently by the corresponding PLDs compared to when they are primed by the PLDs of non-corresponding actions. Furthermore, we hypothesized an increased negativity of the parietal LNC in response to human actions sounds primed incongruently by the PLDs, relative to the congruently primed target sounds (Bristow et al., 2009; Friedrich & Friederici, 2004; 2005, 2017; Koelsch, Kasper, Sammler, Schulze, Gunter, & Friederici, 2004; Kutas, & Federmeier, 2011). Several previous studies on cross-modal processing in infancy report that the effect of the match between the prime and the target on different ERP components varies across the left and right hemispheres, and the midline scalp locations (e.g., Bristow et al., 2009; Friedrich & Friederici, 2004; 2005; 2006; Vogel, Monesson,

& Scott, 2012). For example, the early sensorial and perceptual ERP responses to both visual and auditory targets differentiate between congruous and incongruous cross-modal priming at the electrodes located over the left hemisphere and midline (Vogel et al., 2012; Friedrich & Friederici; 2004; 2006). Furthermore, auditory ERP components have also been show to appear first during neural maturation and to be more prominent at midline locations (e.g., Bishop, 2007; Marshall et al., 2009; Novak, Kurtzberg, Kreuzer, & Vaughan, 1989; deRegnier, Nelson, Thomas, Wewerka, & Georgieff, 2000). In light of these previous findings, we also anticipated that the cross-modal priming effects in the present study may vary across the hemispheres and the midline scalp locations.

Method

Participants

The final sample consisted of twenty 7-month-old infants (M_{age} = 212 days, SD= 10 days, Min_{age} = 197 days, Max_{age} = 229 days; 7 males). Twenty-nine additional infants were tested, but excluded from the final sample because of loss of attention (n=4), fussiness and crying (n=10) or artifacts resulting in insufficient analysable trials (n=15, please see the *Data Analysis* for further information). This rejection rate is analogous to other studies which used paradigms that are highly demanding for infants' attention, due to the length of the trials and the requirements for artifact free data (e.g., Righi, Westerlund, Congdon, Troller-Renfree & Nelson, 2014). The sample size is in line with previous infant research in which priming paradigms were employed with similar age groups (e.g., Peykarjou, Wissner & Pauen, 2017; Peykarjou, Wissner & Pauen, 2020). Participants were recruited from a small urban area in the North of England (UK). They did not have any history of neurological or significant medical condition, were born full term (between 37 and 42 gestational weeks), with a normal birth weight (>2500 g) and had intact vision and hearing abilities. Prior to the testing sessions, all parents were provided with

information about the study and gave their written informed consent, according to the ethical standards of the Declaration of Helsinki (BMJ 1991; 302:1194). The study was approved by the University ethics committee. Parents received £10 as a reimbursement for their travel expenses, while infants received a diploma as a reminder of their participation.

Task

A cross-modal priming paradigm was implemented. Each trial began with a visual prime PLD depicting the kinematics specific to one of two types of human actions, walking or hand clapping, which lasted 1500ms. After the prime offset, a static white bull's eye was presented in the centre of the screen for 1900 ms in order to maintain infants' engagement with the task and avoid eye-movements. The auditory target stimulus (i.e., footstep or hand clapping sound) was presented 500 ms after the onset of the bull's eye, and lasted 1400 ms. A similar inter-stimulus interval of 500 ms has already been employed in priming studies with infants (Bristow et al., 2008; Kopp & Dietrich, 2013), and is reasonably short to allow the priming effect on the target (Spruyt, Hermans, De Houwer, Vandromme & Eelen, 2007). The intertrial stimulus was represented by a fixation cross located in the center of the screen. The duration of the intertrial varied randomly between 1000-1200 ms (see Figure 1A for an illustration of the trial structure). The procedure continued with the next trial if the infant showed visual engagement with the stimuli. If needed, a non-social audio-video clip was played between trials in order to re-gain infants' attention before continuing the stimulus presentation.

The auditory targets were primed either congruently or incongruently by the visual stimulus in terms of action category, resulting in four conditions: clap congruent ($\mathbf{C}_{lap}\mathbf{C}_{lap}$, hand clapping visual prime - hand clapping auditory target); clap incongruent ($\mathbf{W}_{alk}\mathbf{C}_{lap}$**,** walking visual prime – hand clapping auditory target); walk congruent ($\mathbf{W}_{alk}\mathbf{W}_{alk}$, walking visual prime - walking

auditory target; walk incongruent (**C**$_{lap}$**W**$_{alk}$, hand clapping visual prime - walking auditory target) (Figure 1B).

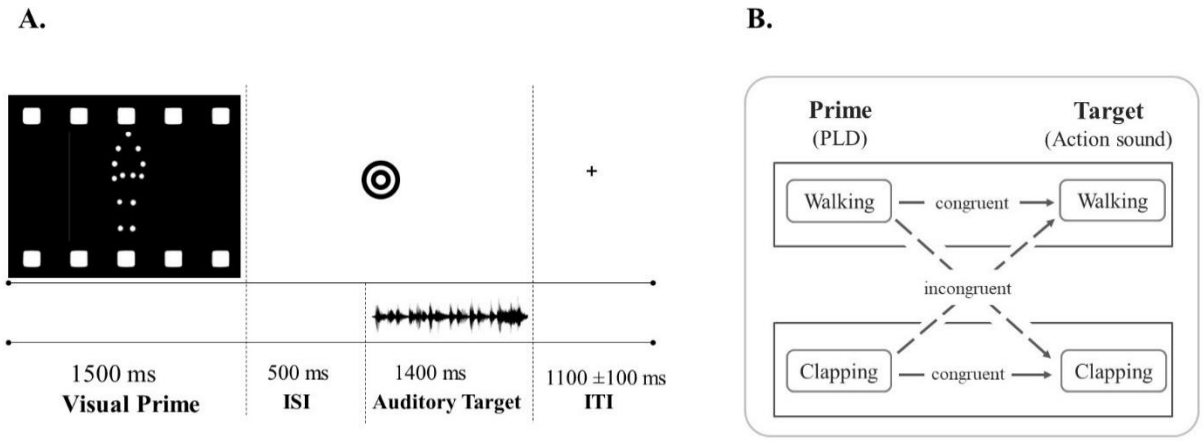

Figure 1. Example of a trial structure (A) and the schematic illustration of the prime-target stimuli combinations (B).

Stimuli

Four different exemplars of *auditory target stimuli* for each action category (i.e., hand clapping and footsteps) were included in the study. Each stimulus represented the sound of an action performed by one individual. The sounds were extracted from an existing database (Geangu et al., 2015a; Engel, Frum, Puce, Walker & Lewis, 2009), were edited to 1400 ms duration and balanced for perceived overall loudness using Audacity 2.1 (all sounds were 44100 Hz). The first 100 ms were ramped and the whole duration was normalized in MATLAB R2012b (MathWorks, Natick, MA) adding a bandpass filter from 1 to 10,000 Hz with a digital finite impulse response filter in order to remove any potential background noise (for further details on the spectral properties of the sounds, see Geangu et al., 2015a). The sound stimuli were rated by 10 adults, and only those identified correctly more than 80% were included in the study. The target stimuli were presented from two speakers placed equidistantly at the left and right side of the monitor.

The *priming visual stimuli* consisted of point-light displays (PLDs) depicting human adults either walking or clapping hands. Four exemplars were used for each action. Each of the PLDs

contained 13 points of light placed at the major joints within the body (head, shoulders, elbows, wrists, hips, knees, and ankles). The walking PLD stimuli were extracted from an existing database (Vanrie & Verfaillie, 2004), while the clapping PLD stimuli were recorded in our lab using the Xsens body motion tracking system (Xsens Inc., Netherlads). The 3D coordinates extracted from the motion sensors placed at each of the 13 points were used to create the PLDs animations with the Biomotion Toolbox (van Boxtel & Lu, 2013) implemented in MATLAB (MathWorks, Natick, MA). The PLD animations as produced by the BiomotionToolbox have similar height (M_{clap} = 463.5 pixels; M_{walk} = 463.5 pixels) and width (M_{clap} = 646.5 pixels; M_{walk} = 646.5 pixels). The size and color of the points of light, and the color of the background were maintained constant across the stimuli. We further tested whether low-level image motion features differed between exemplars (on average) from each action category by computing the optic flow for each PLD. The optic flow measures the displacement of each point on consecutive frames of the video. The magnitude of the displacements are then averaged across points and frame pairs. More specifically, we computed the optic flow for each PLD based on the Lucas-Kanade method (Lucas & Kanade, 1981), and then compared the magnitude between walking and clapping actions. While there is a small numeric difference in the mean magnitude of optic flow between the clap (M = .56 pixels) and the walk (M = .78 pixels) PLDs, this difference is not significant (p = .12). Each PLD video had a duration of 1500 ms.

Examples of the visual prime and auditory target stimuli can be found here .

Procedure

The visual primes were presented on a 17-inch CRT computer monitor, with a dimension of 24x16cm, forming a horizontal visual angle of 13.2° and a vertical visual angle of 19.3°.

The stimuli were presented in a randomized order using MATLAB R2012b (MathWorks, Natick, MA), so that no more than three prime-target pairs of the same type were presented

consecutively, and no more than three of the same PLD or sound exemplars could be repeated consecutively. Participants sat on their parents' laps, at a viewing distance from the monitor of approximately 70 cm. Testing was performed in a dimly lit and sound attenuated room.

The average time of stimuli presentation was approximately 10 minutes, including a break when needed, and the procedure was stopped if the infants showed signs of distress. The average number of trials presented was 101.15 (*SD*=17.7): $\mathbf{C}_{lap}\mathbf{C}_{lap}$ – M = 25.9, SD = 4.5; $\mathbf{W}_{alk}\mathbf{C}_{lap}$ - M = 25.2, SD = 4.6; $\mathbf{W}_{alk}\mathbf{W}_{alk}$ - M = 25.2, SD = 4.5; $\mathbf{C}_{lap}\mathbf{W}_{alk}$ – M = 24.8, SD = 4.7. In order to minimize body movement and exclude the influence of other vocal stimulations, parents were instructed to avoid talking, pointing to the screen or otherwise socially interact with their infants. Trials where these instructions were not respected were excluded from analysis.

Electroencephalogram recording and ERP analysis

EEG was recorded continuously using a 128-electrode HydroCel Geodesic Sensor Net (Electrical Geodesic Inc., Eugene, OR), amplified through an EGI NetAmps 300 amplifier and recorded with EGI (Electrical Geodesics, Inc.) software. Impedances of the electrodes were checked before the beginning of the recording and were considered acceptable if lower than 50 kΩ. The signal was referenced online to the vertex electrode (Cz) and data was sampled at 500 Hz. Further offline processing was performed using NetStation v4.5.4 (Eugene, OR). The signal was bandpass filtered (0.3-30 Hz). Data were segmented into trials with 200 ms baseline before and 1400 ms following the target stimulus onset (i.e. the action sounds), and then the signal was corrected with respect to the average voltage of the baseline. All the channels where the signal exceeded ±200 μV at any electrode were marked as bad, and each trial where more than 18 channels were bad was automatically rejected. Artifacts, such as blinks, eye movements or other movements which cannot be automatically individuated, were manually checked and, if necessary, rejected. Trials in which more than 18 bad channels were identified, were excluded

from further analysis. On the remaining trials, individual bad channels were replaced using spherical spline interpolation, and then the signal of each channel was re-referenced to the average of all channels. On average across conditions, for the trials included in the final analysis, 8.5 (SD = 3.0) bad channels were interpolated. For each condition, the number of interpolated bad channels varied as follows: $M_{Clap\text{-}Clap}$ = 9.3 (SD = 2.6; Min = 1.5; Max = 17. 5); $M_{Clap\text{-}Walk}$ = 8.5 (SD = 3.4; Min = 1.6; Max = 16.2); $M_{Walk\text{-}Clap}$ = 8.6 (SD = 2.7; Min = 2.0; Max = 16.5); $M_{Walk\text{-}Walk}$ = 7.7 (SD = 3.3; Min = 1.1; Max = 15.4). There was a small significant difference (M_{diff} = 1.67, p = .004) between Clap-Clap and Walk-Walk trials. No significant differences emerged between these two conditions in the ERP analysis as described in the results section. The channels used for the detection of eye movements (32, 25, 21, 17, 14, 8 and 1) were marked as bad for the entire data set and rejected from any further analysis (not subjected to interpolation). The rejected trials had, in average, 35.16 bad channels (SD = 11.25): $M_{Clap\text{-}Clap}$ = 35.16 (SD = 11.25); $M_{Clap\text{-}Walk}$ = 39.23 (SD = 13.7); $M_{Walk\text{-}Clap}$ = 36.44 (SD = 9.83); $M_{Walk\text{-}Wwalk}$ = 33.72 (SD = 11.26). Only the trials in which the prime was watched for more than 50% were included in the final analysis. This criterion ensured that the infants had the chance to extract from the PLD the relevant information for identifying the type of action (Hoehl & Wahl, 2012). From all trials excluded from the final analysis, an average of 29.4% trials per condition (SD = 18.03) were in part rejected because the participants did not watch the prime for a minimum of 50% of its duration ($\mathbf{C}_{lap}\mathbf{C}_{lap}$ – M = 31.4, SD = 21.2; $\mathbf{W}_{alk}\mathbf{C}_{lap}$ – M = 30.8, SD = 20.8; $\mathbf{W}_{alk}\mathbf{W}_{alk}$ – M = 29.2, SD = 14.18; $\mathbf{C}_{lap}\mathbf{W}_{alk}$ – M = 26.4, SD = 15.9). Across participants, an average of 10 trials per condition (SD = 1.7) contributed to the average ERPs ($\mathbf{C}_{lap}\mathbf{C}_{lap}$ – M = 10.7, SD = 1.7, Min = 8, Max =14; $\mathbf{W}_{alk}\mathbf{C}_{lap}$ – M = 10.1, SD = 1.7, Min = 7, Max =15; $\mathbf{W}_{alk}\mathbf{W}_{alk}$ – M = 10.6, SD = 1.4, Min = 8, Max =13; $\mathbf{C}_{lap}\mathbf{W}_{alk}$ – M = 10, SD = 1.4, Min = 7, Max =12). There were no significant differences between conditions in terms of the number of trials contributing to the final analysis (p >.09). The number of rejected trials is mainly due to the length of the trials (approximately 5 seconds). Especially in the second half of the procedure when infants' attention naturally diminishes, eye

and body movement happened more frequently during the target stimulus presentation. The trials (N=529) of five participants were manually inspected for artifacts by a second coder. Cohen Kappa was used to calculate reliability between the original and second editor based on cases of agreement and disagreement for the accepted/rejected trials. At a Kappa = .9, the two coders are in near perfect agreement. The number of trials that contribute to the final analyses is in line with previous research using similar paradigms (Bristow et al., 2009; Crespo-Llado, Vanderwert & Geangu, 2018; Hendrickson, Love, Walenski & Friend, 2019). Individual participant averages where computed separately for each channel across all trials within each condition.

The visual inspection of the waveforms indicated the presence of ERP morphologies similar to those typically reported in cross-modal priming or auditory studies in infancy and adulthood (e.g., Bristow et al., 2009; Crespo-Llado et al., 2018; Friedrich & Friederici, 2004; 2005; Grossmann et al., 2006; Hendrickson et al., 2019; Ho et al., 2015; Kokinous et al., 2015; Otte, Donkers, Braeken & Van den Bergh, 2015; Pizzamiglio et al., 2005; Pell et al., 2015; Yeh, Geangu, & Reid, 2016), with the N100-P200 complex and the LPC at fronto-central regions of interest (ROIs), and the LNC at parietal ROIs. At frontal ROIs, the ERP morphology also indicated the presence of a positive deflection around 300ms, similar to a P300. Although the presence of a frontal P300 component is not systematically reported in previous infant crossmodal priming studies, some studies with adults have indicated that the P300 may reflect processing of human action sounds linked to the STS and the premotor cortex (Pizzamiglio et al., 2005). In light of these data and the ERP morphology in the present study, we also explored possible priming effects on the P300. The choice of time- windows for the analysis was informed by converging evidence from previous studies and visual inspection of the waveforms (Kappenman & Luck, 2012): 50-110ms (N100 – e.g., Crespo-Llado et al., 2018; Ho et al., 2015; Kokinous et al., 2015; Yeh, Geangu, & Reid, 2016); 120-220 ms (P200 – e.g., Kopp & Dietrich, 2013; Crespo-Llado et al., 2018; Paulmann, Bleichner, & Kotz, 2013; Yeh et al., 2016); 250-450ms (P300 – e.g., Pizzamiglio et al., 2005); 450-800 ms (LPC - Friedrich & Friederici, 2004;

2005; Grossmann et al., 2006; Grossman, 2013); and 400-850 ms (LNC - Bristow et al., 2009; Friedrich & Friederici, 1998; 2004; 2005). In order to test our predictions regarding the variation of priming effects as a function of location, we included in the analysis clusters of electrodes corresponding to the left and right hemisphere, and the midline. The clusters of electrodes corresponding to each ROI were as follows: frontal (left: electrode 19, 20, 23, 24, 27, 28; right: electrode 3, 4, 117, 118, 123, 124; midline: electrode 16, 11); central (left: electrode 29, 30, 35, 36, 37, 41, 42; right: electrode 87, 93, 103, 104, 105, 110, 111; midline: electrode REF, 55); and parietal (left: electrode 53, 54, 60, 61, 67; right: electrode 77, 78, 79, 85, 86; midline: electrode 62, 72). We restricted the midline electrode groups to only those electrodes that are located on the midline according to the 10-10 system (Acharya et al., 2016; Luu & Ferree, 2005) in order to minimize the risk that possible effects specific to this location may be influenced by those that characterise the right and/or the left hemisphere. Unequal numbers in the electrode groups are not uncommon (e.g., Kopp & Dietrich, 2013; Dien, 2017; Vogel et al., 2012). Besides testing our predictions, the inclusion in the analysis of the midline clusters can also inform the integration of the present findings with those from other previous auditory studies which indicate that the midline electrodes show first during maturation and more prominently different auditory ERP components of interest (e.g., Bishop, 2007; Marshall et al., 2009; Novak, Kurtzberg, Kreuzer, & Vaughan, 1989; deRegnier, Nelson, Thomas, Wewerka, & Georgieff, 2000) and are also more likely to be sensitive to age related changes and various risk factors (van den Heuvel, 2015; deRegnier et al., 2002; Paul et al., 2013; Marshall et al., 2009; Nelson et al., 2003; Siddappa et al., 2004), as well as with studies that include midline electrodes only (e.g., Kokinous et al., 2015; Yeh et al., 2016; also for recommendations on this topic from Picton et al., 2000).

The ERP parameter included in the analysis was the mean amplitude for each time window of interest, due to its robustnes against noise (Clayson, Baldwin, & Larson, 2013; Kappenman & Luck, 2012).

Results

In order to analyze the presence of a cross-modal priming effect, for each ROI, we conducted a 2 (Target: clap, walk) x 2 (Prime: clap, walk) x 3 (Location: left, midline, right) repeated-measures ANOVAs on the mean amplitude of each component of interest. This analytic approach allows to test both the effects of prime-target congruency (suggested by a significant interaction between the prime and target factors), and the possible independent effects of the prime and of the target. This design allows direct comparisons between trials where the auditory target stimuli are preceded by different primes. In order to balance between Type 1 and Type 2 errors, planned comparisons based on the study hypotheses were performed for all significant main effects and interactions (Saville, 1990). The p-values are reported both uncorrected and with the Benjamini-Hochberg correction for false discovery rate (Benjamini-Hochberg, 1995) in parentheses. All statistical tests were interpreted at .05 level of significance (two-tailed). Greenhouse-Geisser correction was applied whenever the assumption of Sphericity was violated (indicated by ε).

Figure 2 and 3 show the ERPs for all ROIs included in the analysis, with the N100, P200, P300 and LPC components visible in the frontal and central ROIs, and the LNC visible in the parietal clusters. Next, we will present the analyses for each of these components.

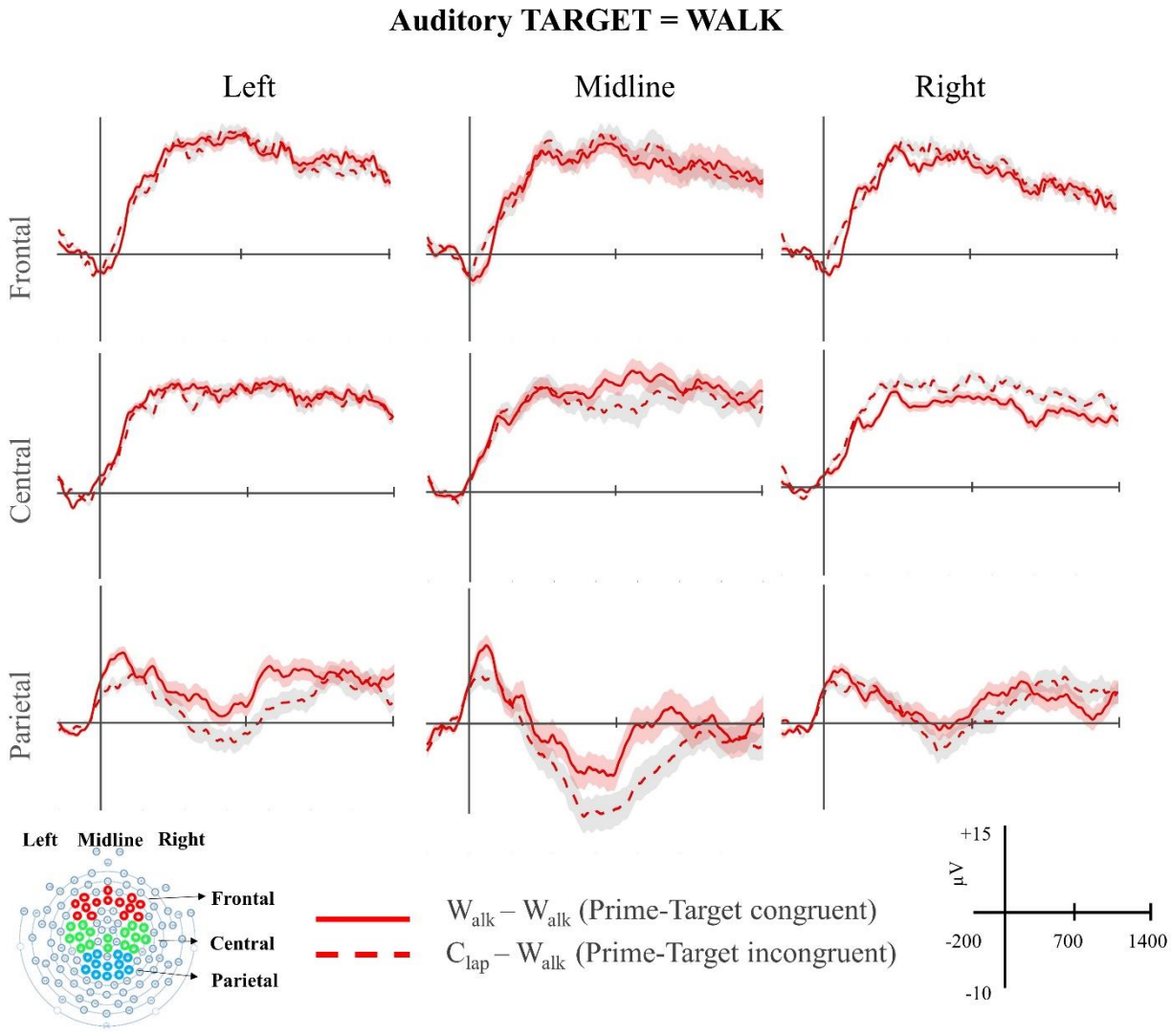

Figure 2. The ERPs for auditory targets representing walking actions at frontal, central, and parietal ROIs (left, midline, and right sites). The clusters of electrodes corresponding to each ROI were: frontal (left: electrode 19, 20, 23, 24, 27, 28; right: electrode 3, 4, 117, 118, 123, 124; midline: electrode 16, 11); central (left: electrode 29, 30, 35, 36, 37, 41, 42; right: electrode 87, 93, 103, 104, 105, 110, 111; midline: electrode REF, 55); and parietal (left: electrode 53, 54, 60, 61, 67; right: electrode 77, 78, 79, 85, 86; midline: electrode 62, 72). The congruency of the trials is established relative to the target stimuli: solid lines represent trials where auditory target is preceded by a visual prime depicting the same human action (W_{alk}-W_{alk}), dashed lines represent trials where auditory target is incongruent with visual prime (C_{lap}-W_{alk}). The prime-target congruency effect was statistically tested through the Prime x Target interaction. The *SEM* for each data point in the ERP time series is represented as the shaded area.

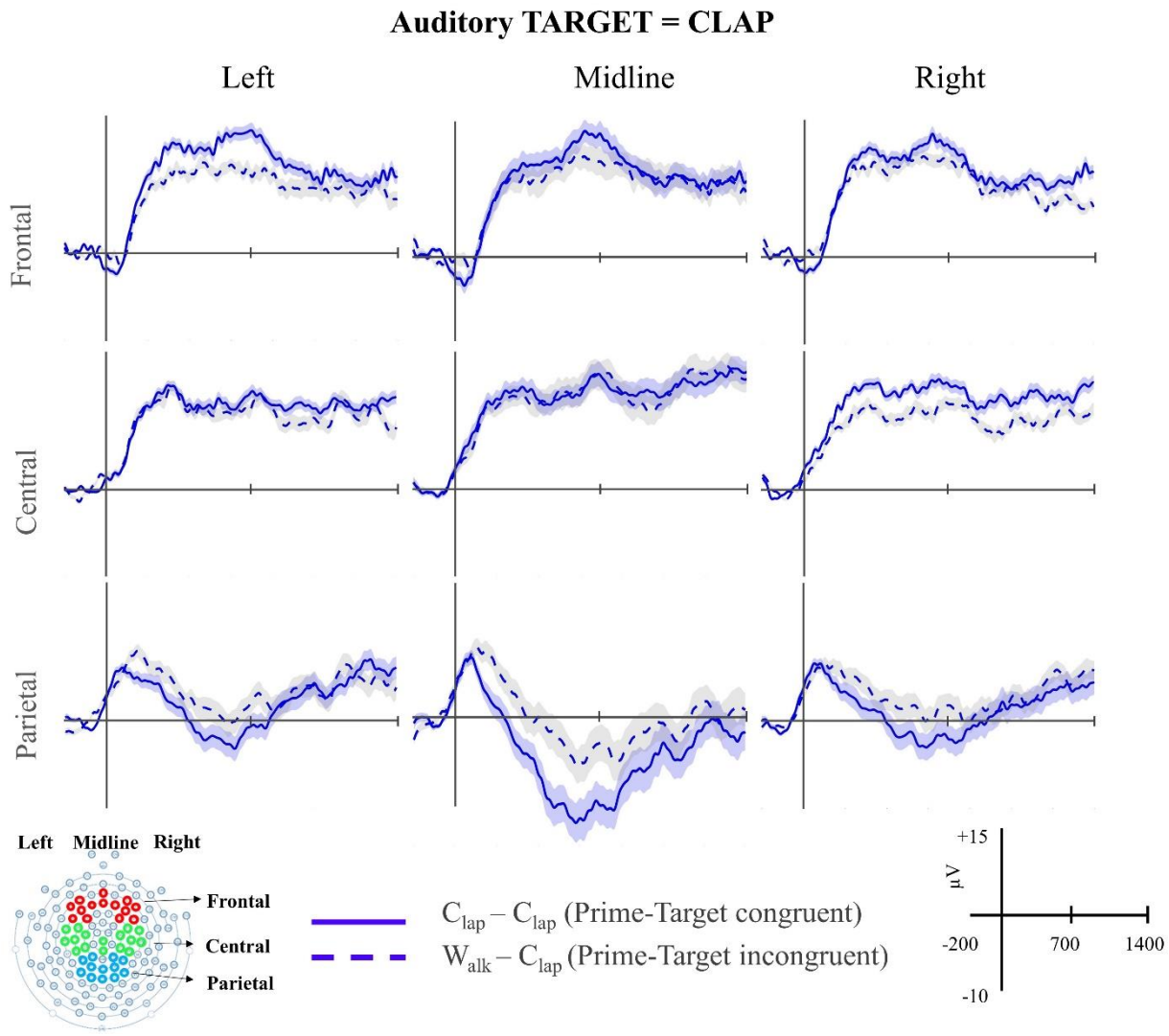

Figure 3. The ERPs for auditory targets representing clapping actions at frontal, central, and parietal ROIs (left, midline, and right sites). The clusters of electrodes corresponding to each ROI were: frontal (left: electrode 19, 20, 23, 24, 27, 28; right: electrode 3, 4, 117, 118, 123, 124; midline: electrode 16, 11); central (left: electrode 29, 30, 35, 36, 37, 41, 42; right: electrode 87, 93, 103, 104, 105, 110, 111; midline: electrode REF, 55); and parietal (left: electrode 53, 54, 60, 61, 67; right: electrode 77, 78, 79, 85, 86; midline: electrode 62, 72). The congruency of the trials is established relative to the target stimuli: solid lines represent trials where auditory target is preceded by a visual prime depicting the same human action (C_{lap} - C_{lap}), dashed lines represent trials where auditory target is incongruent with visual prime (W_{alk} - C_{lap}). The prime-target congruency effect was statistically tested through the Prime x Target interaction. The *SEM* for each data point in the ERP time series is represented as the shaded area.

N100 (50 - 110 ms)

At frontal locations, a main effect of location was observed, $F_{(1,19)} = 3.585$; $p = .037$; $\eta_p^2 = .159$, with the electrodes located over the midline recording reduced negative amplitude ($M = .846$ μV; $SD = .741$ μV) compared to the electrodes located over the right hemisphere ($M = -.3866$ μV; $SD = .829$ μV; $p = .01$) (Figure 3 - A). A significant interaction between target and prime also emerged, $F_{(1,19)} = 6.366$; $p = .021$; $\eta_p^2 = .251$. Planned comparisons showed that at all locations, walking sounds elicited increased negative voltage when they were congruently primed by walking PLDs ($M = -.478$ μV; $SD = .923$ μV) compared to when they were incongruently primed by clapping PLDs ($M = 1.904$ μV; $SD = .760$ μV, $t(19) = -2.744$, $p = .011$ (.026), $d = -0.61$) (Figure 3 – B, C). Walking sounds primed by clapping PLDs also elicited reduced negative amplitude ($M = 1.904$ μV; $SD = .760$ μV) compared to the clapping sounds primed by clapping PLDs ($M = -.768$ μV; $SD = 1.018$ μV, $t(19) = 2.651$, $p = .008$ (.016), $d = 0.59$) (Figure 3 - C). No main effect of the prime or of the target was found (all $p > .15$).

At the central ROI, a main effect of location emerged, $F_{(2,38)} = 5.87$; $p = .006$; $\eta_p^2 = .24$, which was further qualified by a significant interaction with the type of auditory target, $F_{(2,38)} = 4.88$; $p = .013$; $\eta_p^2 = .20$. Planned comparisons revealed that for walking sounds, reduced amplitude was recorded from the electrodes located in the right hemisphere ($M = 3.18$ μV, $SD = 2.84$ μV) compared to those located at midline ($M = 4.83$μV, $SD = 3.78$ μV; $t(19) = -2.359$, $p = .012$ (.024), $d = -0.53$). For the clapping sounds, both the midline ($M = 5.16$ μV, $SD = 4.19$ μV; $t(19) = -2.359$, $p < .001$ (.003), $d = 0.9$) and the right ($M = 4.27$ μV, $SD = 2.81$ μV; $t(19) = -4.10$, $p = .016$ (.032), $d = 0.6$) hemisphere electrodes recorded increased amplitude compared to the electrodes located over the left hemisphere ($M = 2.70$ μV, $SD = 2.44$ μV) (Figure 3). No other significant main effects or interactions were recorded for N100 at either frontal or central ROI ($p > 0.38$).

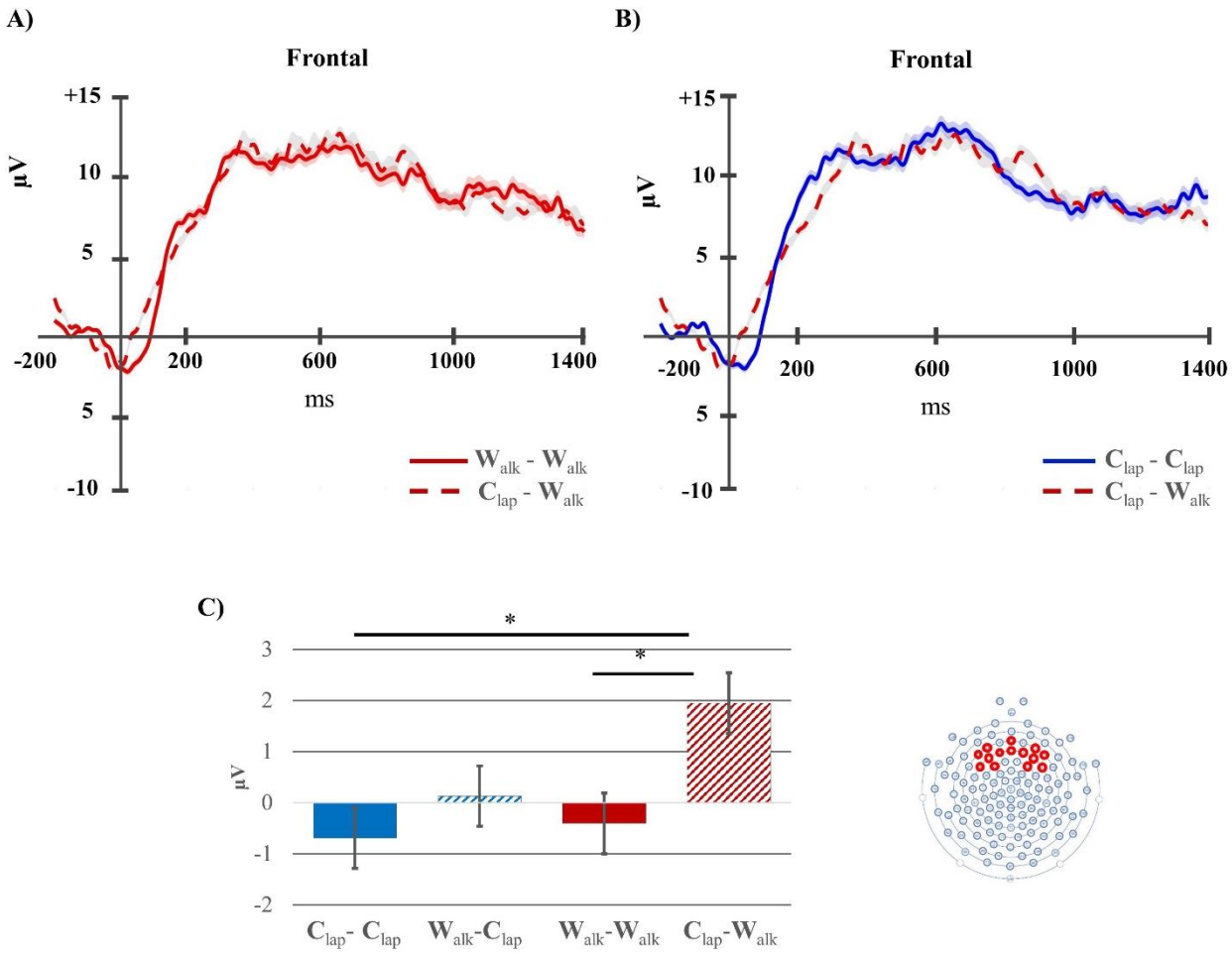

Figure 4. Average N100 (50 - 110 ms) amplitudes at frontal locations. **(A)** at all frontal locations, walking sounds elicited increase negative voltage when primed by walking PLDs (*M* = -.478 μV; *SD* = .923 μV) compared to when primed by clapping PLDs (*M* = 1.904 μV; *SD* = .760 μV); **(B)** walking sounds primed by clapping PLDs also elicited reduced negative amplitude (*M* = 1.904 μV; *SD* =.760 μV) compared to the clapping sounds primed by clapping PLDs (*M* = -.768 μV; *SD* = 1.018 μV, *t*(19) = 2.651, *p* = .008, *d* = 0.59). **(C)** The bar chart summarizes A and B. The *SEM* for each data point in the ERP time series is represented as the shaded area.

P200 (120-220 ms)

For this component, no significant main effect or interactions were observed for either the frontal (all ps > 0.12) or the central ROI (all p > 0.14).

P300 (250 - 450 ms)

For the P300, no significant main effect or interactions were observed for either the frontal (all ps > 0.2) or the central ROI (all ps > 0.19).

LPC (450-800 ms)

At frontal ROI, no significant main effect or interactions were observed for the mean amplitude of the LPC (all ps > 0.06). However, at the central ROI, two significant interactions were observed. On one hand side, target interacted significantly with location, $F(2,38) = 3.29$, $p = 0.048$, $\eta_p^2 = 0.15$. Irrespective of the type of prime, walking sounds elicited increased positive amplitude (M = 10.95 µV, SD = 4.64 µV) compared to the clapping sounds (M = 39.14 µV, SD = 4.74 µV; $t(19) = 2.1$, $p = .049$ (.15), $d = 0.47$), and this difference was specific to the cluster of electrodes over the left hemisphere (Figure 4). The omnibus ANOVA also indicated a significant interaction between the prime and location, $F(2, 38) = 4.15$, $p = 0.023$, $\eta_p^2 = 0.18$. However, none of the planned comparisons indicated significant differences between conditions (all ps > .1).

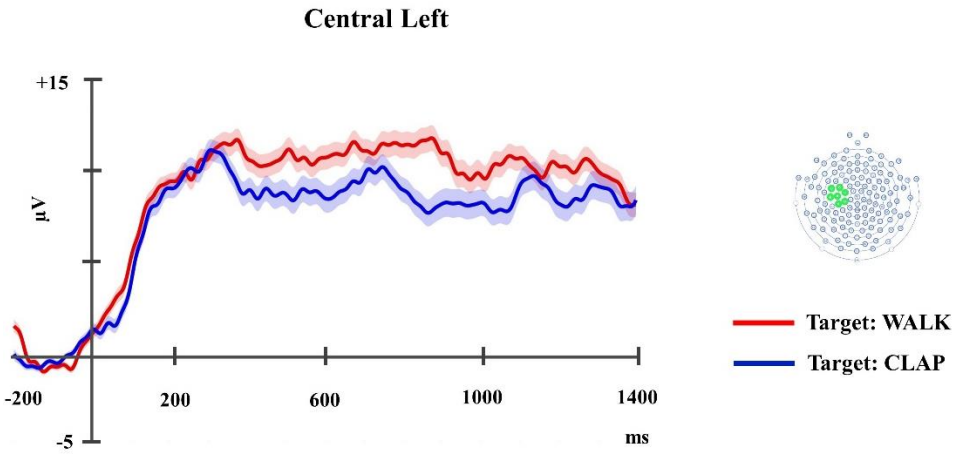

Figure 5. Average LPC (450-800 ms) amplitudes at the central locations. For the left hemisphere, irrespective of the prime, walking sounds elicited increased LPC amplitude (M = 10.95 µV, SD

= 4.64 μV) compared to the clapping sounds (M = 39.14 μV, SD = 4.74 μV). The *SEM* for each data point in the ERP time series is represented as the shaded area. Please note that the negative is plotted down.

LNC (400-850 ms)

At parietal ROIs, the 2 (Target: clap, walk) x 2 (Prime: clap, walk) x 3 (Location: left, right, midline) ANOVA showed a main effect of location, F (2, 38) = 23.75; p < 0.001; η_p^2 = 0 .56, with increased negative amplitude recorded at midline (M = -6.21 μV; SD = 10.4 μV) compared to both left (M = 0.58 μV; SD = 7.64 μV; t(18) = -5.69, p < .001 (<.001), d = -1.3) and right (M = -0.07 μV; SD = 8.62 μV; t(18) = -5.28, p < .001 (<.001), d = -1.21) locations. A significant main effect of prime was also observed, F (1, 19) = 12.93; p = 0.002; η_p^2 = 0.41. These main effects were qualified by their interaction, F (2, 29.07) = 5.01, p = 0.012, η_p^2 = 0.21, ε = 0.77. Planned comparisons showed that for left and midline locations, auditory targets primed by clapping PLDs elicited significantly increased negative voltage (M_{left} = -.92 μV, SD = 6.87 μV; $M_{midline}$ = -8.98 μV, SD = 9.29 μV) compared to targets primed by walking PLDs (M_{left} = 2.07 μV, SD = 6.91 μV; $M_{midline}$ = -3.449 μV, SD = 9.757 μV), t(33.8) = -2.539, p = 0.016 (.032), d = 0.58 and t(33.8) = -4.687, p <0.001 (=.001), d = 0.57, respectively (Figure 5). Furthermore, in the case of both types of primes, the auditory targets elicited significantly more negative voltage at the midline electrode cluster ($M_{PrimeClap}$ = -8.98 μV, SD = 9.29 μV; $M_{PrimeWalk}$ = -3.45 μV, SD = 9.76 μV) compared to the electrode clusters over the left ($M_{PrimeClap}$ = -.92 μV, SD = 6.87 μV, t(19) = 7.25, p < 0.001, d = -1.61; $M_{PrimeWalk}$ = 2.073 μV, SD = 6.91 μV, t(19) = 4.13, p < 0.001, d = 0.87) and right ($M_{PrimeClap}$ = -1.22 μV, SD = 7.34 μV, t(19) = 5.94, p = 0.001, d = -1.28; $M_{PrimeWalk}$ = 1.08 μV, SD = 7.99 μV, t(19) = 3.61, p = 0.002 (.004), d = -0.77) hemispheres. No other significant main effects or interactions were observed for the parietal LNC (all ps > 0.361).

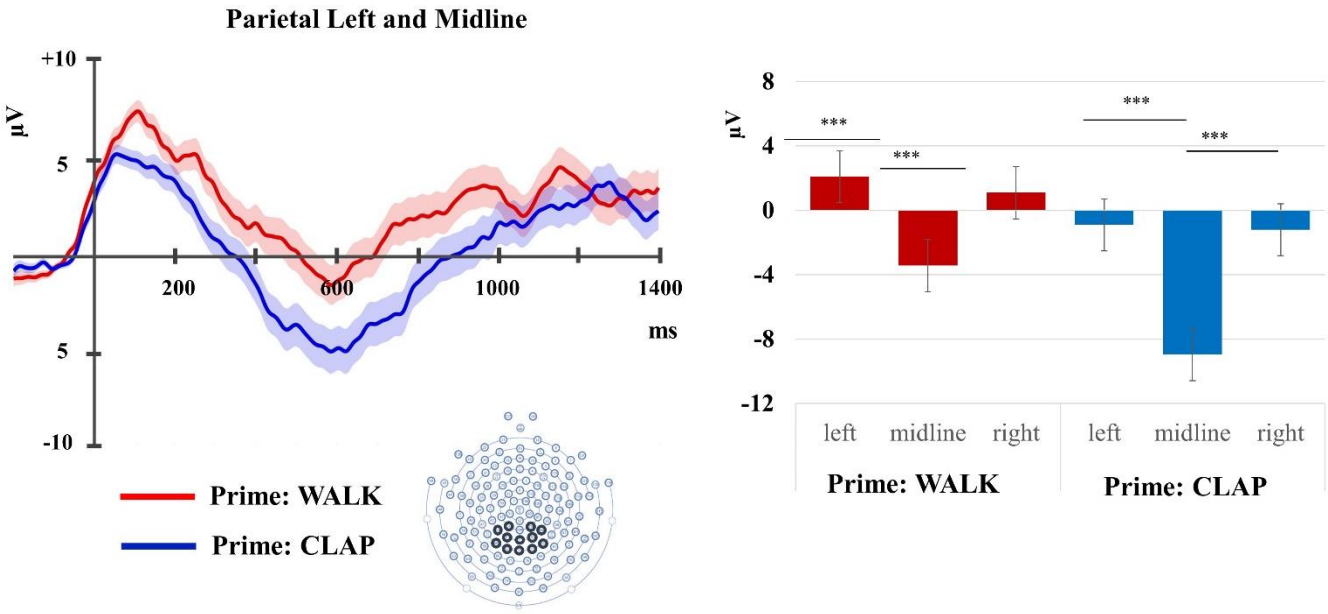

Figure 6. Average LNC (400-850 ms) amplitudes for the parietal electrodes. Auditory targets primed by clapping PLDs elicited significantly increased negative voltage at left (M_{left} = -.92 μV, *SD* = 6.87 μV) and midline electrodes ($M_{midline}$ = -8.98 μV, *SD* = 9.29 μV) compared to targets primed by walking PLDs (M_{left} = 2.07 μV, *SD* = 6.91 μV; $M_{midline}$ = -3.449 μV, *SD* = 9.757 μV). For both types of primes, the auditory targets elicited more negative voltage at the midline electrode cluster ($M_{PrimeClap}$ = -8.98 μV, *SD* = 9.29 μV; $M_{PrimeWalk}$ = -3.45 μV, *SD* = 9.76 μV) compared to the electrode clusters over the left ($M_{PrimeClap}$ = -.92 μV, *SD* = 6.87 μV; $M_{PrimeWalk}$ = 2.073 μV, *SD* = 6.91 μV) and right ($M_{PrimeClap}$ = -1.22 μV, *SD* = 7.34 μV; $M_{PrimeWalk}$ = 1.08 μV, *SD* = 7.99 μV) hemispheres. The *SEM* for each data point in the ERP time series is represented as the shaded area. The negative is plotted down.

Discussion

The sounds associated with human actions provide rich social information (Sievers et al., 2013; Lewis et al., 2011; 2006; Pizzamiglio et al., 2005). Previous research has shown that this auditory information is integral to how human adults represent people, together with information from other modalities (Robinson & Sloutsky, 2010). By the age of 7-months infants differentiate between sounds produced by 'living' vs. 'non-living' entities, and process action sounds as a distinct grouping of human produced sounds alongside vocalizations (Geangu et al., 2015a). Furthermore, infants appear to be sensitive to the temporal synchrony between the visual and

auditory cues of simple human actions, such as hand-clapping (Kopp & Dietrich, 2013; Kopp, 2014). The present study aimed to extend this knowledge by finding out whether infants integrate auditory and visual information in representations of specific human actions. In a cross-modal priming paradigm, 7-months-old infants were presented with the sounds generated by two specific types of human body movement, walking and handclapping, after watching the kinematics of those actions in an either congruent or incongruent manner. The ERP responses to the auditory stimuli indicate emerging audio-visual representations of human walking and hand-clapping actions.

According to a spreading activation account (Bower, 1991; Fazio et al., 1986), we predicted that, if infants represent human actions by integrating both visual and auditory information, the ERP responses to the the target human action sounds will be influenced by the prior presentation of the point light displays (PLDs) depicting the kinematics of those actions. In line with our predictions, the amplitude of one of the early auditory ERP components, the frontal N100, from the N100-P200 complex, varied as a function of the congruency between the visual prime and the auditory target. This effect, however, was only specific to the walking footstep sounds. When preceded by congruent walking PLDs, the footstep sounds triggered increased negative amplitude in the 50-150 ms time window at frontal locations compared to when the sounds were preceded by non-matching PLDs. Typically, the fronto-central N100 is considered to be an exogenous component which reflects pre-attentive neural sensitivity to various acoustic properties of the stimuli, such as frequency (Schröger, Näätänen, & Paavilainen, 1992; Woods & Clayworth, 1985). The N100 differentiates between human produced sounds that are acoustically different, such as different emotional vocalizations (e.g., Pell et al., 2015; Crespo-Llado et al., 2018). However, several studies with infants and adults have also shown prime-target congruency effects on the N100 for audio-visually presented congruent and incongruent speech (Friedrich & Friederici, 2004; 2005), speech prosody (Zinchenko et al., 2015, 2017), and emotional vocalizations (Garrido-Vàsquez et al., 2018; Jessen & Kotz, 2011; Kokinous et al.,

2015). These findings point towards the existence of perceptual representations which include both visual and auditory information. The visual prime activates this perceptual representation, with further top-down effects on the early pre-attentive sensorial processing of the target. Some previous studies show a reduction in the N100 amplitude in response to sounds preceded by related images (e.g., Garrido-Vasquez et al., 2018; Pourtois, de Gelder, Vroomen, Rossion & Crommelick, 2000), regarded to indicate a facilitatory priming effect through a cross-modal prediction mechanism. According to this interpretation, the target is easier to process as a result of the already activated perceptual representation (van Wassenhove et al., 2005; Jessen & Kotz, 2013; Ho et al., 2015). However, similar to the present findings, several of the previous studies report an increase in the N100 amplitude to sounds preceded by related visual targets (Friedrich & Friederici, 2004; 2005; Ho et al., 2015; Pourtois et al., 2000). It is thus possible that the congruent visual priming may enhance the sensorial processing of footstep sounds in 7-months-old infants. In the current study we presented infants with 4 different exemplars of walking sounds and 4 different exemplars for each type of PLDs, and the auditory and visual stimuli were extracted from different recordings of these actions. Given that the primes and the targets were randomly paired for each trial, one possible interpretation of the N100 effects could be that they reflect the influence of perceptual representations which include auditory and visual properties common across many exemplars of human walking. It is also important to note, that although some studies report cross-modal priming effects for N100 at frontal locations (e.g., Friedrich & Friederici, 2004; 2005; Kokinous et al., 2014; Kopp, 2014), like in the present study, others report these effects mostly at centro-parietal locations (e.g., Jessen & Kotz, 2011; Yeh et al., 2016). An analyses of these differences seems to suggest that the N100 cross-modal priming effects are more likely to be observed for anterior/frontal scalp regions in infants (although see Kokinous et al., 2014 for similar findings in adults), while in adults these tend to be more characteristic for centro-parietal regions (Friedrich & Friederici, 2004; 2005; Ho et al., 2015; Pourtois et al., 2000; Vroomen & Stekelenburg, 2007). The infant frontal N100 has also been shown to differentiate

between other types of social sounds (i.e., emotional non-verbal vocalizations – Crespo-Llado et al., 2018). These topographical differences may be related to developmental changes in neurocognitive functions that occur throughout childhood (Friedrich & Friederici, 2004; 2005). Furthermore, there are also some variations in the time windows for which these effects are shown (e.g., Friedrich & Friederici, 2004; 2005; Kopp and Dietrich, 2013). In part these may be explained by age differences (Friedrich & Friederici, 2004; 2005). However, it is also very much likely that the type of stimuli (e.g., emotional vs non-emotional; verbal vs non-verbal) and the experimental paradigm (e.g., the overlap between the prime and the trial; the interstimulus onset asynchrony) also contribute to these variations. For example, Kopp and Dietrich (2013) observed in 6-months-old infants approximately 60 ms difference in the latency of N100 between the study employing an experimental paradigm with synchronous onset of the visual and auditory stimuli, and the one where the two onsets were asynchronous. It would be important in future studies to address more systematically the sources of these topographical and latency variations.

Although we did not anticipate differences in the priming effects between the two types of action sounds, the N100 responses to the hand-clapping sounds were not modulated by the preceding PLDs. This suggests that 7-months-old infants may not necessarily represent the perceptual relation between the hand-clapping sounds and the kinematics corresponding to this type of action. Prior research has shown that infants are sensitive to the statistical regularities with which the stimuli occur in their environment, and that learning and using these environmental regularities are important for the development of several cognitive abilities, including object recognition (Gopnik et al., 2001; Jusczyk & Aslin, 1995; Kirkham et al., 2002; Smith & Yu, 2008). One speculative explanation for the lack of sensitivity to the match between the PLDs and the clapping sounds could be that infants did not have sufficient opportunities to learn the audio-visual correspondences specific to this type of action. Head-mounted camera recordings show that throughout the first year of life, dramatic changes occur in the frequency with which different types of objects appear in infants' view (Fausey et al., 2016; Smith et al.,

2018). For example, while from birth infants have people constantly in their view, specific body parts such as peoples' hands become more frequent only as infants get closer to the second half of the first year of life (Fausey et al., 2016; Smith et al., 2018). Furthermore, hand-clapping actions represent only a small subset of all hand actions infants may have observed, which may not provide sufficient opportunities to encode the relation between the visual and auditory cues (Fitzpatrick, Schmidt, & Lockman, 1996; Harwood, Schoelmerich, Schulze, and Gonzales, 1999; Jones; 2017; Jones & Yoshida, 2011). In contrast, walking people are a more frequent occurrence in infants' environment, and it could be that this also translates into infants having sufficient opportunities for learning the association between the visual and auditory cues of this type of body movement, which supports automatic detection of audio-visual mismatches. Infants also tend to be frequently carried by walking adults, which could provide opportunities for relating the auditory cues of footsteps with the proprioceptive and vestibular information that results from experiencing the rhythm of a walking body. Previous research has shown that 7-months-old infants chose to listen for longer to auditory rhythms that have the same tempo/beat as the one to which they previously experienced through body bouncing, compared to a rhythm they did not experience (Phillips-Silver & Trainor, 2005). Learning the auditory-proprioceptive regularities specific to walking could potentially facilitate establishing new associations with cues from other modalities, such as vision. Further research is important to test the role of the statistical regularities of the visual and auditory information about human actions infants are exposed to in their everyday life for the development of multisensory representations.

Some previous studies showed priming effects for the fronto-central P200, out of the N100-P200 complex (Kopp, 2014; Friedrich & Friederici, 2004), while others reported cross-modal priming effects selectively for the N100 (e.g., Garrido-Vasquez et al., 2018), like in the present study. One characteristic of these latter studies is that the relation between the prime and the target is unambiguous and easy to establish within the experimental paradigm. When the prime is more ambiguous (Yeh et al., 2016; Zinchenko et al., 2017) and/or establishing the relation

between the prime and the target is more challenging (e.g., the difference between the primes is ambiguous), the congruency effects are shifted towards P200 (Garrido-Vasquez et al., 2018; Zinchenko et al., 2017). The fact that in our study the congruency between the PLDs and the walking sounds was evident for the N100, suggests that 7-months-old infants reliably extract from action kinematics the information that allows them to discriminate between hand-clapping and walking actions. Furthermore, the perceptual link between the auditory and visual information specific to walking is well established, and it is rapidly processed. This may have been further facilitated by the temporal separation between the visual primes and the auditory targets, and by the dynamic information provided by both stimulus events (Addabbo, Longhi, Marchis, Tagliabue, & Turati, 2018; Jeschonek, Pauen, & Bobocsai, 2013; Kaiser, Crespo-Llado, Turati, & Geangu, 2017). Our findings add to those reported by Kopp and colleagues (2013, 2014), to suggest that the infant auditory N100 is not only sensitive to the temporal predictability of the human action sounds by the corresponding images of the moving body, but may also be sensitive to whether the visual information predicts what type of sound will occur. That being said, these findings should be interpreted with caution until future research replicates and extends these effects to other types of human actions.

In contrast to auditory-visual integration at perceptual stages, our results suggest that integration at higher amodal stages does not occur yet at the age of 7-months. We hypothesized that if infants represent human actions by integrating both visual and auditory information, then the LPC in response to the target human action sounds will be influenced by the prior presentation of the point light displays (PLDs) depicting the kinematics specific to those actions. As indicated by some of the previous research that used a priming paradigm for investigating language and face processing in infants (e.g., Friedrich & Friederici, 2004; Grossmann et al., 2006), we anticipated that hand-clapping and walking sounds preceded by matching PLDs will elicit a more positive LPC compared to trials of incongruent priming. In contrast to our hypotheses, the anterior-central LPC did not differentiate between congruently and incongruently primed human

actions sounds. Instead, in line with previous studies that link the LPC with global-level category formation in infants (Quinn, Westerlund & Nelson., 2006; Nelson, Thomas, de Haan, & Wewerka, 1998) and auditory category distinctions (e.g., living vs. non-living sounds - Geangu et al., 2015a; positive vs. negative emotional vocalizations – Crespo-Llado et al., 2018; Pell et al., 2015; Schirmer & Kotz, 2006), 7-months-old infants in the present study tended to record an increased LPC voltage to walking relative to hand-clapping sounds, irrespective of the preceding visual information. Thus, while these findings do not provide support for the modulation of the categorical auditory processing by the corresponding visual information, they suggest that in addition to representing human actions sounds as belonging to a broader category of living entities (Geangu et al., 2015a), by the age of 7-months infants tend to represent distinctly the auditory information characteristic to specific types of human actions.

As described by several infant visual-auditory priming studies (Friedrich & Friederici, 2004; 2011; 2017), a negative deflection in the waveform was evident for the electrodes located parietally, with the peak amplitude visible around 600ms after the onset of the human action sounds. This parietal late negative component (LNC) was proposed to reflect the processing of the relations between the prime and the target that go beyond simple perceptual associations, possibly as a precursor of the semantic N400 described in older children and adults (Friedrich & Friederici, 2004; 2011). We hypothesized that if infants integrate the visual information related to action kinematics and the sound produced by those actions, then an increased negativity of the parietal LNC in response to human actions sounds primed incongruently by the PLDs, relative to the congruently primed target sounds (Bristow et al., 2009; Friedrich & Friederici, 2004; 2005, 2017; Koelsch et al., 2004; Kutas, & Federmeier, 2011). Instead, our results showed that all action sounds primed by the clapping PLDs elicited a more negative LNC compared to the actions sounds preceded by the walking PLDs. This indicates that the processing of the auditory information as reflected by the parietal LNC is largely dominated by the visual cues of the kinematics specific to hand-clapping and walking actions, therefore audio-visual integration does

not occur beyond the early sensorial stages of processing. Our findings are consistent with the evidence showing a protracted development of multisensory integration, and that different neurocognitive mechanisms for multisensory processing may be present at different points during development (Burr & Gori, 2012; Bremner, Lewkowicz, & Spence, 2012; Gori et al., 2008; Nardini, Jones, Bedford, & Braddick, 2008). The dominance of one sense over another, particularly the dominance of vision, has been shown for several tasks in infancy, including orientation in the peripersonal space (Bremner et al. 2008a,b) and matching emotion expression for face and voice (Otte et al., 2015). While basic forms of integration may support a simpler pre-attentive mechanism that guides orientation to audiovisual stimuli at a very early age (Neil et al. 2006), more complex integration processes required for non-reflexive tasks may still be absent (Barutchu et al. 2009, 2010).

One possible explanation for the predominant effect of the visual information extracted from the PLDs on the processing of human action sounds as reflected by the parietal LNC could be that, in comparison, the auditory information is less precise (Burr & Gori, 2012). However, the central LPC suggests that infants process hand-clapping sounds as distinct from walking, and we know that the development of infants' auditory acuity tends to precede vision (Slater, 1998). Furthermore, studies have shown that even in the cases where the information from one modality is less precise for the task at hand, relative to information from other senses, older children will still rely predominantly on it, if it is more direct and robust (Gori et al., 2008). Thus, it could be that the predominant influence of visual information on the LNC might not be necessarily due to its higher precision. Another possible explanation could be related to infants' emerging abilities to deal with conflicting information (Gerardi-Caulton, 2000; Holmboe, Fearon, Csibra, Tucker, & Johnson, 2008), which may be overtaxed by the processing demands characteristic to the priming task. Given that the visual prime has temporal precedence relative to the auditory target, the corresponding representations may have an advantage in terms of activation strength. With immature abilities for processing the conflicting information between

the prime and the target, 7-months-old infants may end up relying on the information that is more strongly activated. Differences in how different actions are visually represented, possibly as a result of differences in the opportunities infants have had to observe these actions in their environment as discussed above, may further explain the direction of the difference in the amplitude of the parietal LNC. If one action (based on our speculations, clapping) is more poorly visually represented compared to the other one, then it is plausible to expect that this poorer representation will be also reflected in its influence on the target processing at different levels including those reflected by the parietal LNC. The increased LNC amplitude in response to actions sounds primed by clapping PLDs compared to actions sounds primed by walking PLDs may reflect more effortful processing (Bower, 1991; Fazio et al., 1986). The fact that the N100 differentiated between the prime-target match/mismatch for the walking sounds with no evidence to indicate a similar effect for clapping sounds, could be regarded as further indication in this respect. Further research is needed to test these possible explanations. For example, one could test infants' neural responses to auditorily primed human action PLDs, in order to establish the role of the temporal precedence of the information from one sense relative to another, which is typical to the situations that infants encounter in their everyday life. Even when the onset of stimulation via different modalities is synchronous, there are differences in how fast the information reaches the receptors and also in how fast the neural processing takes place, which can lead to one type of sensory information to precede the other (e.g., Vroomen & Keetels, 2010). Thus, studying the impact of stimulation asynchrony on multisensory processing is relevant for understanding how this ability develops.

Based on the present findings alone, it is not possible to establish whether the parietal LNC indexes similar semantic processing as the N400 (Juottonen, Revonsuo & Lang, 1996; Federmeier, Van Petten, Schwartz & Kutas, 2003), and whether the pattern of priming results characterised by the dominance of one sense represents an immature form of multisensory processing that precedes the adult-like multisensory integration (Burr & Gori, 2012; Gori et al.,

2008). The difference in the priming effects compared to the frontal N100, suggests that these ERP components may reflect different computations relevant for multisensory processing. Studies investigating the semantic processing of spoken words have shown that in younger infants (12-months-old) the early anterior negativity shows cross-modal priming in the absence of a similar effect on the parietal LNC, for which the prime-target matching effects tend to be only observed in older 19-months-old infants (e.g., Friedrich & Friederici, 2004, 2005). Future longitudinal studies are needed in order to establish whether there is a continuity between the infant LNC in response to cross-modal priming and the semantic N400 observed later in childhood and adulthood. Longitudinal studies would also provide opportunities to replicate and elucidate why in younger infants (7-months-old in the present study, 12-months-old in Friedrich & Friederici (2005)) the effects of the match between the visual prime and the auditory target appears to be more restricted to the early ERP components such as the N100.

In summary, our study shows that the integration of visual and auditory cues into multimodal representations of human actions begins to emerge in infancy. At the age of 7-months, the multimodal processing appears to be mostly based on perceptual associations between image and sound, and may only characterize actions that are more frequently present in infants' environment. At this age, however, infants do not appear to integrate the sound and image of body kinematics into specific cognitive representations of human actions. When human action sounds appear in the context of seeing the body movement, the way they are represented is predominantly driven by what infants see. Our findings also expand the existent evidence about infants' neural processing of human action sounds. By the age of 7-months, infants not only process human action sounds as belonging to a broader category of 'living' sounds, but also tend to group them according to specific actions. The present study suggests that infants are sensitive to the change in the sound people make while in action and can establish some basic relations between the visual and auditory streams of information. It would be interesting to find out in

future research whether human action sounds also contribute to how infants infer people's intentions from their actions and how infants engage in successful social interactions.

Acknowledgements

We would like to express our gratitude to all families who dedicated their time to participate in this research project with their children. Without their continuous interest in our work and desire to help, these findings would have not been possible. The authors would also like to thank Barrie Usherwood for his technical assistance with the study implementation, and to Dr. Quoc Vuong for insightful discussions and support. The authors wish to declare no conflict of interests. The work for this study was partially supported by an Wellcome Trust Centre for Future Health Grant and an EPSRC IAA 2017 Grant awarded to Elena Geangu at University of York (UK).

CRediT roles

Elena Geangu: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing - original draft; Writing - review & editing.

Elisa Roberti: Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Validation; Visualization; Roles/Writing - original draft; Writing - review & editing.

Chiara Turati: Conceptualization; Formal analysis; Writing - review & editing

Bibliography

Acharya, J. N., Hani, A. J., Cheek, J., Thirumala, P., & Tsuchida, T. N. (2016). American Clinical Neurophysiology Society guideline 2: guidelines for standard electrode position nomenclature. *The Neurodiagnostic Journal*, *56*(4), 245-252.

Addabbo, M., Longhi, E., Marchis, I. C., Tagliabue, P., & Turati, C. (2018). Dynamic facial expressions of emotions are discriminated at birth. *PloS one*, *13*(3), e0193868.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental psychology*, *36*(2), 190.

Bahrick, L. E., & Lickliter, R. (2002). Intersensory redundancy guides early perceptual and cognitive development. *Advances in child development and behavior*, *30*, 153-189.

Bahrick, L. E., Walker, A. S., & Neisser, U. (1981). Selective looking by infants. *Cognitive Psychology*, *13*(3), 377-390.

Bardi, L., Regolin, L., & Simion, F. (2011). Biological motion preference in humans at birth: Role of dynamic and configural properties. *Developmental science*, *14*(2), 353-359.

Bartsch, L., Van der Zwan, R., Cottrell, D., & Brooks, A. (2007, January). Auditory biological motion processing: The eyes alone don't have it!. In *Psychology Making an Impact: Proceedings of the 42nd APS Annual Conference, Brisbane* (pp. 7-11).

Baruch, C., Panissal-Vieu, N., & Drake, C. (2004). Preferred perceptual tempo for sound sequences: comparison of adults, children, and infants. *Perceptual and motor skills*, *98*(1), 325-339.

Barutchu A, Crewther D.P, & Crewther S.G. (2009). The race that precedes coactivation: development of multisensory facilitation in children. *Developmental Science*, 12:464–73.

Barutchu A, Danaher J, Crewther S.G, Innes-Brown H, Shivdasani M.N, & Paolini A.G. (2010). Audiovisual integration in noise by children and adults. *Journal of Experimental Child Psychology,* 105:38–50.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: series B (Methodological)*, *57*(1), 289-300.

Berg, K. M., & Boswell, A. E. (1998). Infants' detection of increments in low-and high-frequency noise. *Perception & psychophysics*, *60*(6), 1044-1051.

Bhatt, R. S., Hock, A., White, H., Jubran, R., & Galati, A. (2016). The development of body structure knowledge in infancy. *Child Development Perspectives*, *10*(1), 45-52.

Bidet-Ildei, C., Kitromilides, E., Orliaguet, J. P., Pavlova, M., & Gentaz, E. (2014). Preference for point-light human biological motion in newborns: contribution of translational displacement. *Developmental psychology*, *50*(1), 113.

Bischoff, M., Zentgraf, K., Pilgramm, S., Stark, R., Krüger, B., & Munzert, J. (2014). Anticipating action effects recruits audiovisual movement representations in the ventral premotor cortex. *Brain and cognition*, *92*, 39-47.

Bishop, D. V. M. (2007). Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: where are we, and where should we be going?. *Psychological bulletin*, *133*(4), 651.

Bower, G. H. (1991). Mood congruity of social judgments. In J. P. Forgas (Ed.), *Emotion and social judgments* (International Series in Experimental Social Psychology) (pp. 31-53). Oxford: Pergamon.

Bremner, A. J. (2017). Multisensory development. In Hopkins, B., Geangu, E., & Linkenauger, S. (Eds.). *The Cambridge encyclopedia of child development*. Cambridge University Press.

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (Eds.). (2012). *Multisensory development*. Oxford University Press.

Bremner A.J, Holmes N.P, Spence C. (2008a). Infants lost in (peripersonal) space? *Trends in Cognitive Sciences*. 12:298–305.

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (Eds.). (2012). *Multisensory development*. Oxford University Press.

Bremner A.J, Mareschal D, Lloyd-Fox S, Spence C. (2008b). Spatial localization of touch in the first year of life: Early influence of a visual spatial code and the development of remapping across changes in limb position. *Journal of Experimental Psychology. General.* 137:149–62.

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J. F. (2009). Hearing faces: how the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, *21*(5), 905-921.Burnham, D. (1993). Visual recognition of mother by young infants: facilitation by speech. *Perception*, *22*(10), 1133-1153.

Burnham, D. (1993). Visual recognition of mother by young infants: facilitation by speech. *Perception*, *22*(10), 1133-1153.

Burr, D. & Gori, M. (2012). Multisensory integration develops late in humans. In Murray MM, Wallace MT (Eds.). *The Neural Bases of Multisensory Processes.* Boca Raton (FL): CRC Press/Taylor & Francis. Chapter 18.

Carroll, N. C., & Young, A. W. (2005). Priming of emotion recognition. *The Quarterly Journal of Experimental Psychology Section A*, *58*(7), 1173-1197.

Clayson, P. E., Baldwin, S. A., & Larson, M. J. (2013). How does noise affect amplitude and latency measurement of event- related potentials (ERPs)? A methodological critique and simulation study. *Psychophysiology*, *50*(2), 174-186.

Crespo-Llado, M. M., Vanderwert, R. E., & Geangu, E. (2018). Individual differences in infants' neural responses to their peers' cry and laughter. *Biological psychology*, *135*, 117-127.

Dehaene-Lambertz, G., & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of cognitive neuroscience*, *16*(8), 1375-1387.

de Haan, M., & Nelson, C. A. (1997). Recognition of the mother's face by six- month- old infants: A neurobehavioral study. *Child development*, *68*(2), 187-210.

deRegnier, R. A., Nelson, C. A., Thomas, K. M., Wewerka, S., & Georgieff, M. K. (2000). Neurophysiologic evaluation of auditory recognition memory in healthy newborn infants and infants of diabetic mothers. *The Journal of pediatrics*, *137*(6), 777-784.

deRegnier, R. A., Wewerka, S., Georgieff, M. K., Mattia, F., & Nelson, C. A. (2002). Influences of postconceptional age and postnatal experience on the development of auditory recognition memory in the newborn infant. *Developmental psychobiology*, *41*(3), 216-225.

Dien, J. (2017). Best practices for repeated measures ANOVAs of ERP data: Reference, regional channels, and robust ANOVAs. *International Journal of Psychophysiology*, *111*, 42-56.

Draganova, R., Schollbach, A., Schleger, F., Braendle, J., Brucker, S., Abele, H., ... & Preissl, H. (2018). Fetal auditory evoked responses to onset of amplitude modulated sounds. A fetal magnetoencephalography (fMEG) study. *Hearing research*, *363*, 70-77.

Engel, L. R., Frum, C., Puce, A., Walker, N. A., & Lewis, J. W. (2009). Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage*, *47*(4), 1778-1791.

Falck-Ytter, T., Bakker, M., & von Hofsten, C. (2011). Human infants orient to biological motion rather than audiovisual synchrony. *Neuropsychologia*, *49*(7), 2131-2135.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of personality and social psychology*, *50*(2), 229.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101-107.

Federmeier, K. D., Van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, words, sentences: age-related changes across levels of language processing. *Psychology and aging*, *18*(4), 858.

Fitzpatrick, P., Schmidt, R. C., & Lockman, J. J. (1996). Dynamical patterns in the development of clapping. *Child development*, *67*(6), 2691-2708.

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental psychology*, *43*(1), 238.

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of cognitive neuroscience*, *16*(8), 1465-1477.

Friedrich, M., & Friederici, A. D. (2005). Phonotactic knowledge and lexical-semantic processing in one-year-olds: Brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience*, *17*(11), 1785-1802.

Friedrich, M., & Friederici, A. D. (2011). Word learning in 6-month-olds: fast encoding–weak retention. *Journal of Cognitive Neuroscience*, *23*(11), 3228-3240.

Friedrich, M., & Friederici, A. D. (2017). The origins of word learning: Brain responses of 3-month- olds indicate their rapid association of objects and words. *Developmental Science*, *20*(2), e12357.

Gaetano, J. (2013). Holm-Bonferroni sequential correction: an EXCEL calculator-ver. 1.2.

Galati, G., Committeri, G., Spitoni, G., Aprile, T., Di Russo, F., Pitzalis, S., & Pizzamiglio, L. (2008). A selective representation of the meaning of actions in the auditory mirror system. *Neuroimage*, *40*(3), 1274-1286.

Garrido-Vásquez, P., Pell, M. D., Paulmann, S., & Kotz, S. A. (2018). Dynamic facial expressions prime the processing of emotional prosody. *Frontiers in human neuroscience*, *12*, 244.

Geangu, E. (2008). Notes on self awareness development in early infancy. *Cognition, Brain, Behavior*, *12*(1), 103.

Geangu, E., Quadrelli, E., Lewis, J. W., Cassia, V. M., & Turati, C. (2015a). By the sound of it. An ERP investigation of human action sound processing in 7-month-old infants. *Developmental cognitive neuroscience*, *12*, 134-144.

Geangu, E., Senna, I., Croci, E., & Turati, C. (2015). The effect of biomechanical properties of motion on infants' perception of goal-directed grasping actions. *Journal of experimental child psychology*, *129*, 55-67.

Geangu, E., & Vuong, Q. C. (2020). Look up to the body: An eye-tracking investigation of 7-months-old infants' visual exploration of emotional body expressions. *Infant Behavior and Development*, *60*, 101473.

Gerardi-Caulton, G. (2000). Sensitivity to spatial conflict and the development of self-regulation in children 24–36 months of age. *Developmental Science, 3* (4), 397-404. 10.1111/1467-7687.00134

Gervain, J., & Geffen, M. N. (2019). Efficient neural coding in auditory and speech perception. *Trends in neurosciences*, *42*(1), 56-65.

Gervain, J., Werker, J. F., Black, A., & Geffen, M. N. (2016). The neural correlates of processing scale-invariant environmental sounds at birth. *Neuroimage*, *133*, 144-150.

Gervain, J., Werker, J. F., & Geffen, M. N. (2014). Category-specific processing of scale-invariant sounds in infancy. *PLoS One*, *9*(5), e96278.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental psychology*, *37*(5), 620.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, *18*(9), 694-698.

Graven, S. N., & Browne, J. V. (2008). Auditory development in the fetus and infant. *Newborn and infant nursing reviews*, *8*(4), 187-193.

Grossman, T. (2013). The early development of processing emotions in face and voice. In *Integrating face and voice in person perception* (pp. 95-116). Springer, New York, NY

Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision research*, *41*(10-11), 1475-1482.

Grossmann, T., & Johnson, M. H. (2007). The development of the social brain in human infancy. *European Journal of Neuroscience*, *25*(4), 909-919.

Grossmann, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science*, *9*(3), 309-315.

Guillem, F., Bicu, M., & Debruille, J. B. (2001). Dissociating memory processes involved in direct and indirect tests with ERPs to unfamiliar faces. *Cognitive Brain Research*, *11*(1), 113-125.

Harwood, R. L., Schoelmerich, A., Schulze, P. A., & Gonzalez, Z. (1999). Cultural differences in maternal beliefs and behaviors: A study of middle- class Anglo and Puerto Rican mother- infant pairs in four everyday situations. *Child development*, *70*(4), 1005-1016.

Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., & Ceballos-Baumann, A. O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *Journal of cognitive neuroscience*, *17*(2), 282-293.

Hauf, P., Elsner, B., & Aschersleben, G. (2004). The role of action effects in infants' action control. *Psychological Research*, *68*(2-3), 115-125.

Hendrickson, K., Love, T., Walenski, M., & Friend, M. (2019). The organization of words and environmental sounds in the second year: Behavioral and electrophysiological evidence. *Developmental science*, *22*(1), e12746.

Hepper, P. G., & Shahidullah, B. S. (1994). The development of fetal hearing. *Fetal and Maternal Medicine Review*, *6*(3), 167-179.

Hirai, M., Fukushima, H., & Hiraki, K. (2003). An event-related potentials study of biological motion perception in humans. *Neuroscience letters*, *344*(1), 41-44.

Hirai, M., & Hiraki, K. (2005). An event-related potentials study of biological motion perception in human infants. *Cognitive Brain Research*, *22*(2), 301-304.

Ho, H. T., Schröger, E., & Kotz, S. A. (2015). Selective attention modulates early human evoked potentials during emotional face–voice processing. *Journal of Cognitive Neuroscience*, *27*(4), 798-818.

Hoehl, S., & Wahl, S. (2012). Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, *37*(3), 187-209.

Holmboe, K., Fearon, R.P., Csibra, G., Tucker, L.A., Johnson, M.H. (2008). Freeze -Frame: A new infant inhibition task and its relation to frontal cortex tasks during infancy and early childhood. *Journal of Experimental Child Psychology*, 100 (2), 89-114.

Hyde, D. C., Jones, B. L., Flom, R., & Porter, C. L. (2011). Neural signatures of face–voice synchrony in 5- month- old human infants. *Developmental Psychobiology*, *53*(4), 359-370.

Hyde, D. C., Porter, C. L., Flom, R., & Stone, S. A. (2013). Relational congruence facilitates neural mapping of spatial and temporal magnitudes in preverbal infants. *Developmental cognitive neuroscience*, *6*, 102-112.

Ichikawa, H., Kanazawa, S., & Yamaguchi, M.K., (2011). The movement of internal facial features elicits 7-to 8-month-old infants' preference for face patterns. *Infant and child development, 20,* 464-674.

Jeschonek, S., Pauen, S., & Babocsai, L. (2013). Cross-modal mapping of visual and acoustic displays in infants: The effect of dynamic and static components. *European Journal of Developmental Psychology*, *10*(3), 337-358.

Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *Neuroimage*, *58*(2), 665-674.

Jessen, S., & Kotz, S. A. (2013). On the role of crossmodal prediction in audiovisual emotion perception. *Frontiers in Human Neuroscience*, *7*, 369.

Johansson, G. (1973). Visual perception or biological motion and a model for its analysis. *Perception and Psychophysics, 14*, 201–211.

Jones, S. (2017). Can newborn infants imitate?. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1-2), e1410.

Jones, S., & Yoshida, H. (2011). Imitation in infancy and the acquisition of body knowledge. *Early development of body representations*, *13*, 207.

Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research*, *4*(2), 99-107.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants′ detection of the sound patterns of words in fluent speech. *Cognitive psychology*, *29*(1), 1-23.

Kappenman, E.S., & Luck., S.J. (2012). *The Oxford handbook of event-related potential components*. Oxford University Press, Oxford, UK.

Kahana- Kalman, R., & Walker- Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child development*, *72*(2), 352-369.

Kaiser, J., Crespo-Llado, M. M., Turati, C., & Geangu, E. (2017). The development of spontaneous facial responses to others' emotions in infancy: An EMG study. *Scientific reports*, *7*(1), 1-10.

Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: Further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive Brain Research*, *13*(1), 27-39.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature neuroscience*, *7*(3), 302-307.

Kokinous, J., Kotz, S. A., Tavano, A., & Schröger, E. (2015). The role of emotion in dynamic audiovisual integration of faces and voices. *Social Cognitive and Affective Neuroscience*, *10*(5), 713-720.

Kopp, F. (2014). Audiovisual temporal fusion in 6-month-old infants. *Developmental cognitive neuroscience*, *9*, 56-67.

Kopp, F., & Dietrich, C. (2013). Neural dynamics of audiovisual synchrony and asynchrony perception in 6-month-old infants. *Frontiers in psychology*, *4*, 2.

Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, *5*(3), 108-125.

Kusak, G., Grune, K., Hagendorf, H., & Metz, A. M. (2000). Updating of working memory in a running memory task: an event-related potential study. *International Journal of Psychophysiology*, *39*(1), 51-65.

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences*, *105*(32), 11442-11445.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647.

Lewis, J. W., Phinney, R. E., Brefczynski-Lewis, J. A., & DeYoe, E. A. (2006). Lefties get it "right" when hearing tool sounds. *Journal of cognitive neuroscience*, *18*(8), 1314-1330.

Lewis, J. W., Talkington, W. J., Puce, A., Engel, L. R., & Frum, C. (2011). Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. *Journal of cognitive neuroscience*, *23*(8), 2079-2101.

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychological bulletin*, *126*(2), 281.

Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in cognitive sciences*, *13*(11), 470-478.

Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: newborns match nonhuman primate faces and voices. *Infancy*, *15*(1), 46-60.

Li, X., Logan, R. J., & Pastore, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. *The Journal of the Acoustical Society of America*, *90*(6), 3036-3049.

Liszkowski, U., & Tomasello, M. (2011). Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cognitive Development*, *26*(1), 16-29.

Lucas, B., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679.

Luu, P., & Ferree, T. (2005). Determination of the HydroCel Geodesic Sensor Nets' average electrode positions and their 10 – 10 international equivalents. *Technical note from Electrical Geodesics, Inc.*

Marshall, P. J., Reeb, B. C., & Fox, N. A. (2009). Electrophysiological responses to auditory novelty in temperamentally different 9- month- old infants. *Developmental Science*, *12*(4), 568-582.

Marshall, P. J., & Shipley, T. F. (2009). Event-related potentials to point-light displays of human actions in 5-month-old infants. *Developmental Neuropsychology*, *34*(3), 368-377.

Missana, M., Altvater-Mackensen, N., & Grossmann, T. (2017). Neural correlates of infants' sensitivity to vocal expressions of peers. *Developmental cognitive neuroscience*, *26*, 39-44.

Missana, M., Rajhans, P., Atkinson, A. P., & Grossmann, T. (2014). Discrimination of fearful and happy body postures in 8-month-old infants: an event-related potential study. *Frontiers in Human Neuroscience*, *8*, 531.

Murgia, M., Santoro, I., Tamburini, G., Prpic, V., Sors, F., Galmonte, A., & Agostini, T. (2016). Ecological sounds affect breath duration more than artificial sounds. *Psychological research*, *80*(1), 76-81.

Nardini M, Jones P, Bedford R, Braddick O. (2008). Development of cue integration in human navigation. *Current* Biology, *18,* 689–93.

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, *21*(4), 779-784.

Neil, P. A., Chee- Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental science*, *9*(5), 454-464.

Nelson, C. A., & Collins, P. F. (1991). Event-related potential and looking-time analysis of infants' responses to familiar and novel events: Implications for visual recognition memory. *Developmental Psychology*, *27*(1), 50.

Nelson, C. A., Thomas, K. M., De Haan, M., & Wewerka, S. S. (1998). Delayed recognition memory in infants and adults as revealed by event-related potentials. *International Journal of Psychophysiology*, *29*(2), 145-165.

Nelson, C. A., Wewerka, S. S., Borscheid, A. J., deRegnier, R. A., & Georgieff, M. K. (2003). Electrophysiologic evidence of impaired cross-modal recognition memory in 8-month-old infants of diabetic mothers. *The Journal of pediatrics*, *142*(5), 575-582.

Nishiyori, Ryota, Silvia Bisconti, Sean K. Meehan, and Beverly D. Ulrich. "Developmental changes in motor cortex activity as infants develop functional motor skills." *Developmental Psychobiology* 58, no. 6 (2016): 773-783.

Novak, G. P., Kurtzberg, D., Kreuzer, J. A., & Vaughan Jr, H. G. (1989). Cortical responses to speech sounds and their formants in normal infants: maturational sequence and spatiotemporal analysis. *Electroencephalography and clinical neurophysiology*, *73*(4), 295-305.

Noy, D., Mouta, S., Lamas, J., Basso, D., Silva, C., & Santos, J.A. (2017). Audiovisual integration increases the intentional step synchronization of side-by-side walkers. Human Movement Science, *56*(B), 71-87. https://doi.org/10.1016/j.humov.2017.10.007.

O, J., Law, B., and Rymal, A. (2015). Now hear this: auditory sense may be an undervalued component of effective modeling and imagery interventions in sport. *Open Psychol. J.* 8, 203–211. doi: 10.2174/1874350101508010203

Otte, R. A., Donkers, F. C. L., Braeken, M. A. K. A., & Van den Bergh, B. R. H. (2015). Multimodal processing of emotional information in 9-month-old infants II: Prenatal exposure to maternal anxiety. *Brain and cognition*, *95*, 107-117.

Patterson, M. L., & Werker, J. F. (2003). Two- month- old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191-196.

Paul, J. A., Logan, B. A., Krishnan, R., Heller, N. A., Morrison, D. G., Pritham, U. A., ... & Hayes, M. J. (2014). Development of auditory event- related potentials in infants prenatally exposed to methadone. *Developmental psychobiology*, *56*(5), 1119-1128.

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*, 345.

Paulmann, S., & Pell, M. D. (2010). Contextual influences of emotional speech prosody on face processing: How much is enough?. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 230-242.

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological psychology*, *111*, 14-25

Peykarjou, S., Wissner, J., & Pauen, S. (2017). Categorical erp repetition effects for human and furniture items in 7- month- old infants. *Infant and Child Development*, *26*(5), e2016.

Peykarjou, S., Wissner, J., & Pauen, S. (2020). Audio-visual priming in 7-month-old infants: An ERP study. *Infant Behavior and Development*, *58*, 101411.

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: movement influences infant rhythm perception. *Science*, *308*(5727), 1430-1430.

Picton, T. E., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson Jr, R., ... & Taylor, M. J. (2000). Guidelines for using human event- related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127-152.

Pizzamiglio, L., Aprile, T., Spitoni, G., Pitzalis, S., Bates, E., D'amico, S., & Di Russo, F. (2005). Separate neural systems for processing action-or non-action-related sounds. *Neuroimage*, *24*(3), 852-861.

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition*, *98*(1), 1-11.

Pourtois, G., De Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport*, *11*(6), 1329-1333.

Quadrelli, E., Geangu, E., & Turati, C. (2019). Human action sounds elicit sensorimotor activation early in life. *Cortex*, *117*, 323-335.

Quadrelli, E., & Turati, C. (2016). Origins and development of mirroring mechanisms: A neuroconstructivist framework. *British Journal of Developmental Psychology*, *34*(1), 6-23.

Quinn, P. C., Westerlund, A., & Nelson, C. A. (2006). Neural markers of categorization in 6-month-old infants. *Psychological Science*, *17*(1), 59-66.

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42.

Reid, V. M., Hoehl, S., Grigutsch, M., Groendahl, A., Parise, E., & Striano, T. (2009). The neural correlates of infant and adult goal prediction: evidence for semantic processing systems. *Developmental Psychology*, *45*(3), 620.

Righi, G., Westerlund, A., Congdon, E. L., Troller-Renfree, S., & Nelson, C. A. (2014). Infants' experience-dependent processing of male and female faces: Insights from eye tracking and event-related potentials. *Developmental cognitive neuroscience*, *8*, 144-152.

Robinson, C. W., & Sloutsky, V. M. (2010). Effects of multimodal presentation and stimulus familiarity on auditory and visual processing. *Journal of Experimental Child Psychology*, *107*(3), 351-358.

Saby, J. N., Meltzoff, A. N., & Marshall, P. J. (2015). Neural body maps in human infants: Somatotopic responses to tactile stimulation in 7-month-olds. *NeuroImage*, *118*, 74-78.

Saenz, M., & Langers, D. R. (2014). Tonotopic mapping of human auditory cortex. *Hearing research*, *307*, 42-52.

Saville, D. J. (1990). Multiple Comparison Procedures: The Practical Solution. *The American Statistician,* 44:174-180.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in cognitive sciences*, *10*(1), 24-30.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*(2), 228-233.

Schröger, E., Näätänen, R., & Paavilainen, P. (1992). Event-related potentials reveal how non-attended complex sound patterns are represented by the human brain. *Neuroscience letters*, *146*(2), 183-186.

Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences*, *110*(1), 70-75.Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, *105*(2), 809-813.

Siddappa, A. M., Georgieff, M. K., Wewerka, S., Worwa, C., Nelson, C. A., & Deregnier, R. A. (2004). Iron deficiency alters auditory recognition memory in newborn infants of diabetic mothers. *Pediatric research*, *55*(6), 1034-1041.

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, *105*(2), 809-813.

Skipper, L. M., Ross, L. A., & Olson, I. R. (2011). Sensory and semantic category subdivisions within the anterior temporal lobes. *Neuropsychologia*, *49*(12), 3419-3429.

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in cognitive sciences*, *22*(4), 325-336

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558-1568

Slater, A. (Ed.). (1998). *Perceptual development: Visual, auditory, and speech perception in infancy*. Psychology Press.

Spruyt, A., Hermans, D., De Houwer, J., Vandromme, H., & Eelen, P. (2007). On the nature of the affective priming effect: Effects of stimulus onset asynchrony and congruency proportion in naming and evaluative categorization. *Memory & cognition*, *35*(1), 95-106.

Steinbeis, N., & Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *Journal of cognitive neuroscience*, *23*(3), 604-621.

Thierry, G., Vihman, M., & Roberts, M. (2003). Familiar words capture the attention of 11-month-olds in less than 250 ms. *Neuroreport*, *14*(18), 2307-2310.

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(2), 394.

van Boxtel, J. J., & Lu, H. (2013). A biological motion toolbox for reading, displaying, and manipulating motion capture data in research settings. *Journal of vision*, *13*(12), 7-7.

van Elk, M., van Schie, H. T., Hunnius, S., Vesper, C., & Bekkering, H. (2008). You'll never crawl alone: neurophysiological evidence for experience-dependent motor resonance in infancy. *Neuroimage*, *43*(4), 808-814.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences*, *102*(4), 1181-1186

von Koss Torkildsen, J., Syversen, G., Simonsen, H. G., Moen, I., & Lindgren, M. (2007). Brain responses to lexical-semantic priming in children at-risk for dyslexia. *Brain and Language*, *102*(3), 243-261.

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 625-629.

Vogel, M., Monesson, A., & Scott, L. S. (2012). Building biases in infancy: The influence of race on face and voice emotion matching. *Developmental science*, *15*(3), 359-372.

Vroomen, J., Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics,* 72, 871-884. https://doi.org/10.3758/APP.72.4.871

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, *19*(12), 1964-1973.

Zinchenko, A., Obermeier, C., Kanske, P., Schröger, E., and Kotz, S. A. (2017). Positive emotion impedes emotional but not cognitive conflict processing. Cognitive, Affective, and Behavioral Neuroscience, 17, 665–677. doi: 10.3758/s13415-017-0504-1

Walker-Andrews, A. S., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior & Development*. 6(4), 491–498.

Webster, P. J., Skipper-Kallal, L. M., Frum, C. A., Still, H. N., Ward, B. D., & Lewis, J. W. (2017). Divergent human cortical regions for processing distinct acoustic-semantic categories of natural sounds: Animal action sounds vs. vocalizations. *Frontiers in neuroscience*, *10*, 579.

Wilding, E. L., Doyle, M. C., & Rugg, M. D. (1995). Recognition memory with and without retrieval of context: An event-related potential study. *Neuropsychologia*, *33*(6), 743-767.

Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huotilainen, M., ... & Sussman, E. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences*, *100*(20), 11812-11815.

Woods, D. L., & Clayworth, C. C. (1985). Click spatial position influences middle latency auditory evoked potentials (MAEPs) in humans. *Electroencephalography and Clinical Neurophysiology*, *60*(2), 122-129.

Yeh, P. W., Geangu, E., & Reid, V. (2016). Coherent emotional perception from body expressions and the voice. *Neuropsychologia*, *91*, 99-108.

Figure Captions

Figure 1. Example of a trial structure (A) and the schematic illustration of the prime-target stimuli combinations (B).

Figure 2. The ERPs for auditory targets representing walking actions at frontal, central, and parietal ROIs (left, midline, and right sites). The clusters of electrodes corresponding to each ROI were: frontal (left: electrode 19, 20, 23, 24, 27, 28; right: electrode 3, 4, 117, 118, 123, 124; midline: electrode 16, 11); central (left: electrode 29, 30, 35, 36, 37, 41, 42; right: electrode 87, 93, 103, 104, 105, 110, 111; midline: electrode REF, 55); and parietal (left: electrode 53, 54, 60, 61, 67; right: electrode 77, 78, 79, 85, 86; midline: electrode 62, 72). The congruency of the trials is established relative to the target stimuli: solid lines represent trials where auditory target is preceded by a visual prime depicting the same human action (W_{alk}-W_{alk}), dashed lines represent trials where auditory target is incongruent with visual prime (C_{lap}-W_{alk}). The prime-target congruency effect was statistically tested through the Prime x Target interaction. The *SEM* for each data point in the ERP time series is represented as the shaded area.

Figure 3. The ERPs for auditory targets representing clapping actions at frontal, central, and parietal ROIs (left, midline, and right sites). The clusters of electrodes corresponding to each ROI were: frontal (left: electrode 19, 20, 23, 24, 27, 28; right: electrode 3, 4, 117, 118, 123, 124; midline: electrode 16, 11); central (left: electrode 29, 30, 35, 36, 37, 41, 42; right: electrode 87, 93, 103, 104, 105, 110, 111; midline: electrode REF, 55); and parietal (left: electrode 53, 54, 60, 61, 67; right: electrode 77, 78, 79, 85, 86; midline: electrode 62, 72). The congruency of the trials is established relative to the target stimuli: solid lines represent trials where auditory target is preceded by a visual prime depicting the same human action (C_{lap} - C_{lap}), dashed lines represent trials where auditory target is incongruent with visual prime (W_{alk} - C_{lap}). The prime-target congruency effect was statistically tested through the Prime x Target interaction. The *SEM* for each data point in the ERP time series is represented as the shaded area.

Figure 4. Average N100 (50 - 110 ms) amplitudes at frontal locations. **(A)** at all frontal locations, walking sounds elicited increase negative voltage when primed by walking PLDs (M = -.478 μV; SD = .923 μV) compared to when primed by clapping PLDs (M = 1.904 μV; SD = .760 μV); **(B)** walking sounds primed by clapping PLDs also elicited reduced negative amplitude (M = 1.904 μV; SD =.760 μV) compared to the clapping sounds primed by clapping PLDs (M = -.768 μV; SD = 1.018 μV, $t(19) = 2.651$, $p = .008$, $d = 0.59$). **(C)** The bar chart summarizes A and B. The *SEM* for each data point in the ERP time series is represented as the shaded area.

Figure 5. Average LPC (450-800 ms) amplitudes at the central locations. For the left hemisphere, irrespective of the prime, walking sounds elicited increased LPC amplitude (M = 10.95 μV, SD = 4.64 μV) compared to the clapping sounds (M = 39.14 μV, SD = 4.74 μV). The *SEM* for each data point in the ERP time series is represented as the shaded area. Please note that the negative is plotted down.

Figure 6. Average LNC (400-850 ms) amplitudes for the parietal electrodes. Auditory targets primed by clapping PLDs elicited significantly increased negative voltage at left (M_{left} = -.92 μV, SD = 6.87 μV) and midline electrodes ($M_{midline}$ = -8.98 μV, SD = 9.29 μV) compared to targets primed by walking PLDs (M_{left} = 2.07 μV, SD = 6.91 μV; $M_{midline}$ = -3.449 μV, SD = 9.757 μV). For both types of primes, the auditory targets elicited more negative voltage at the midline electrode cluster ($M_{PrimeClap}$ = -8.98 μV, SD = 9.29 μV; $M_{PrimeWalk}$ = -3.45 μV, SD = 9.76 μV) compared to the electrode clusters over the left ($M_{PrimeClap}$ = -.92 μV, SD = 6.87 μV; $M_{PrimeWalk}$ = 2.073 μV, SD = 6.91 μV) and right ($M_{PrimeClap}$ = -1.22 μV, SD = 7.34 μV; $M_{PrimeWalk}$ = 1.08 μV, SD = 7.99 μV) hemispheres. The *SEM* for each data point in the ERP time series is represented as the shaded area. The negative is plotted down.

Figure 1

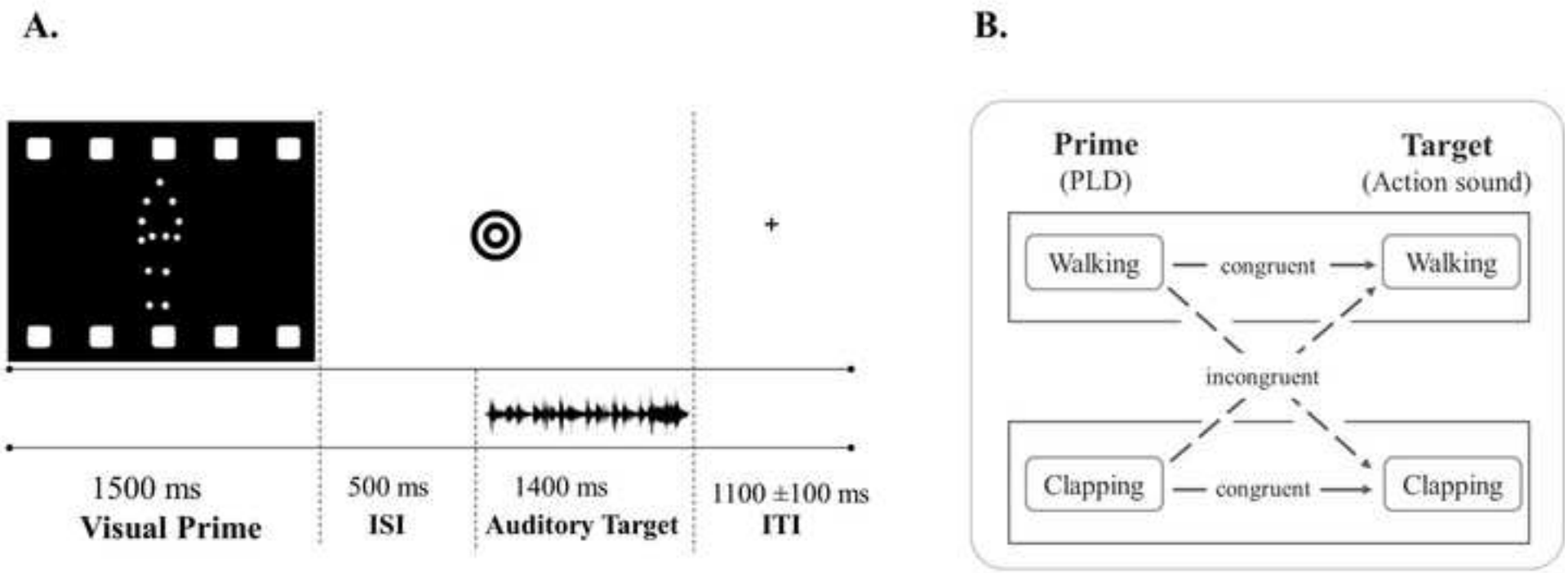

Figure 2

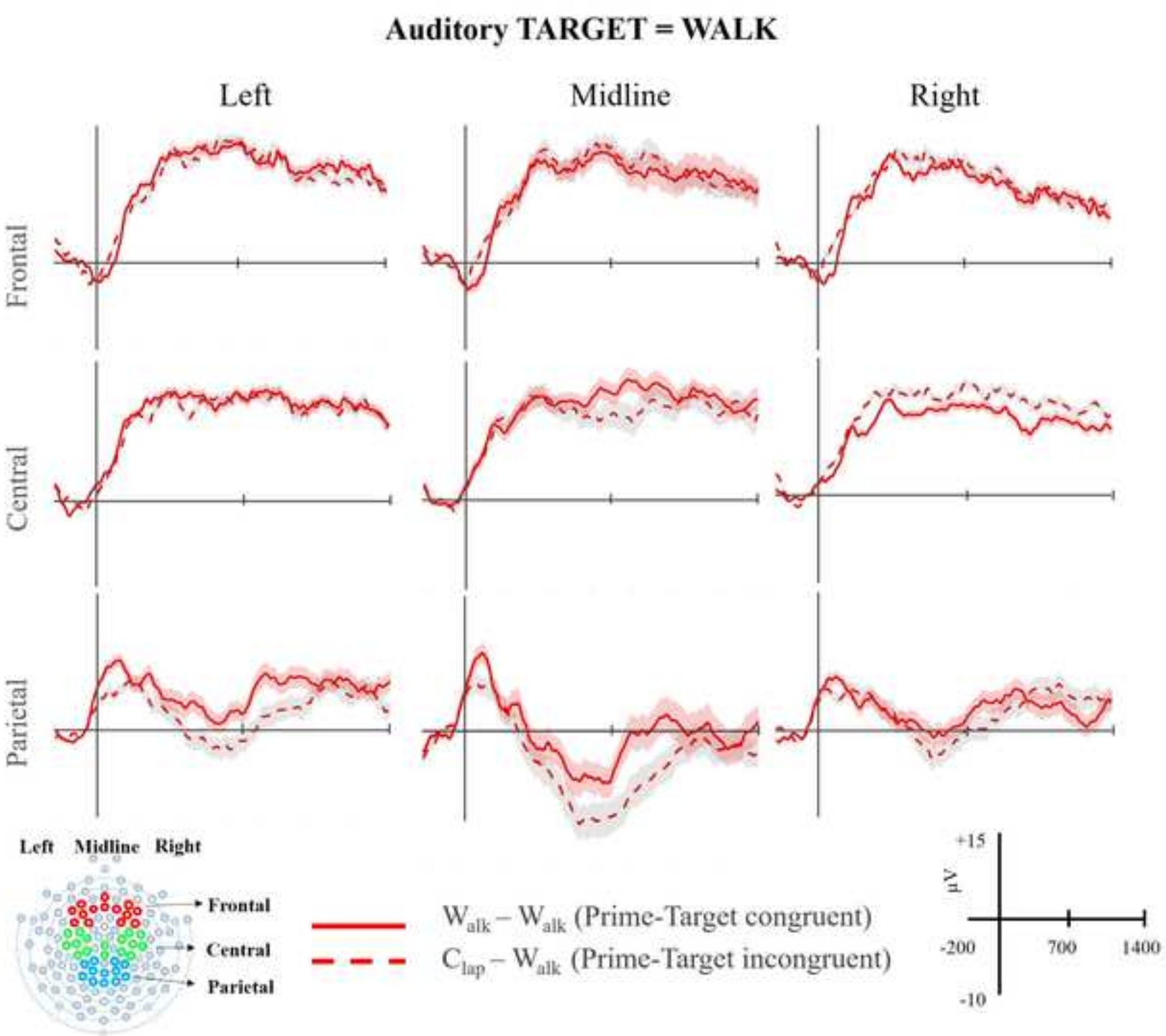

Figure 3

Click here to access/download;Figure;Figure3.tif

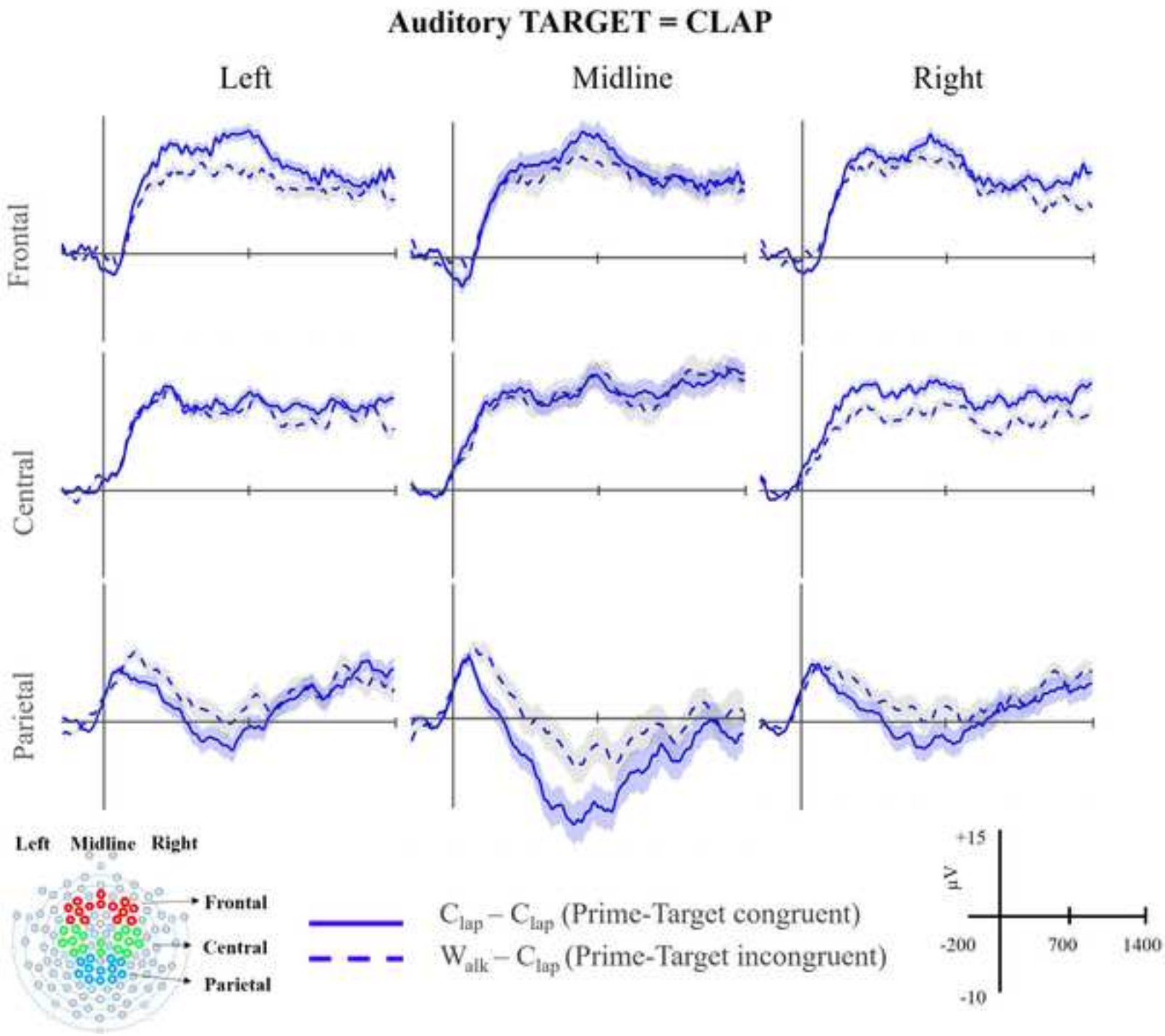

Figure 4

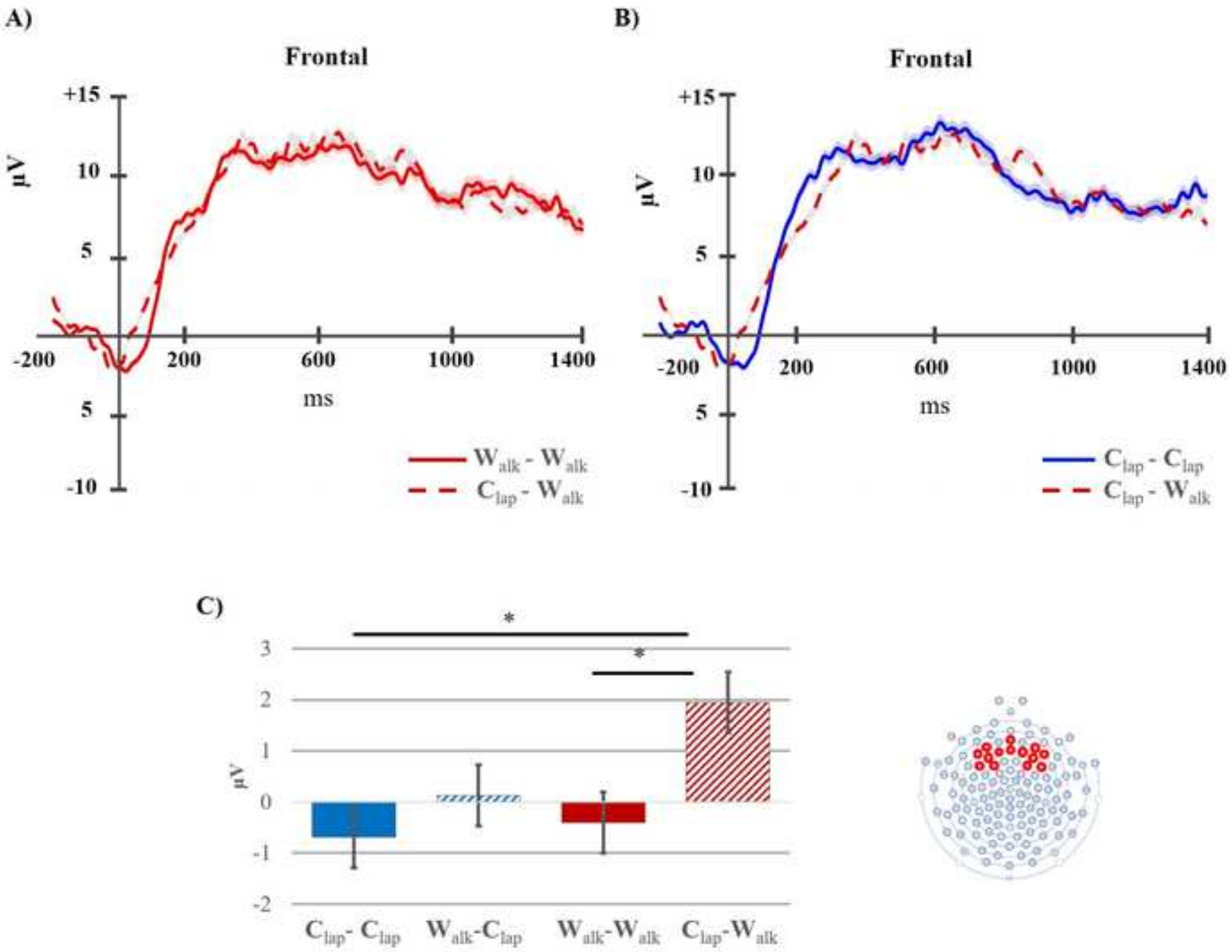

Figure 5

Click here to access/download;Figure;Figure5.tif

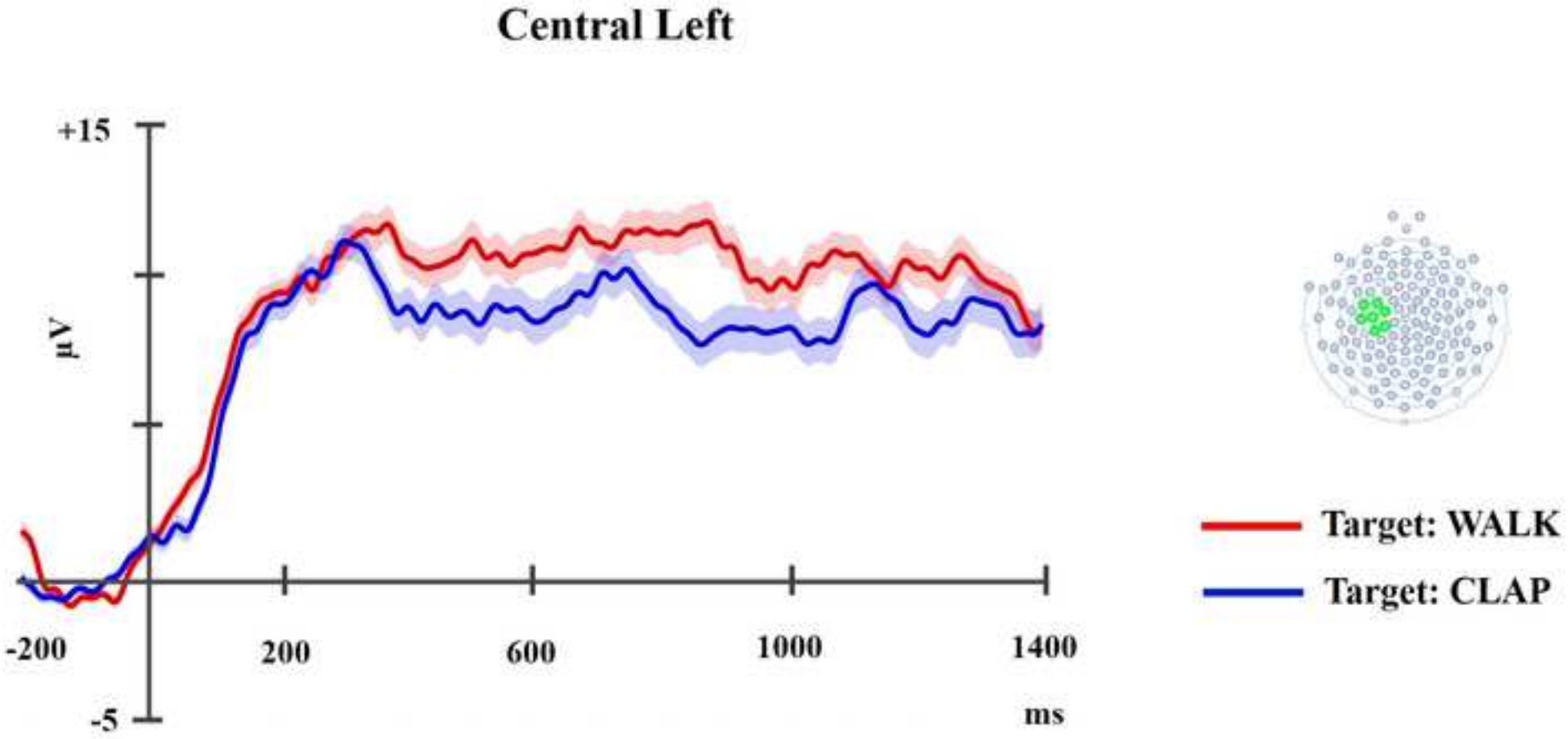

Figure 6