Universidade do Minho
Escola de Psicologia

Catarina Carvalho Senra

Action recognition of point-light displays presented with semantically (in)congruent auditory stimuli: Behavioral correlates

outubro de 2021

Action recognition of point-light-displays presented with semantically (in)congruent auditory stimuli : Behavorioal correlates

Catarina Carvalho Senra

UMinho | 2021

Universidade do Minho
Escola de Psicologia

Catarina Carvalho Senra

Action recognition of point-light displays presented with semantically (in)congruent auditory stimuli : Behavioral correlates

Dissertação de Mestrado
Mestrado Interuniversitário em Neuropsicologia Clínica e Experimental

Trabalho efetuado sob a orientação da
Doutora Olivia Morgan Lapenta
e da
Prof. Doutora Adriana da Conceição Soares Sampaio

outubro de 2021

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho

Atribuição-Compartilhalgual
CC BY-SA

https://creativecommons.org/licenses/by-sa/4.0/

Catarina Carvalho Ferreira

Acknowledgments

First of all my thanks to Doctor Olivia Morgan Lapenta, that was essential to this research journey, and that made this master thesis possible. Secondly, to Professor Adriana Sampaio and all the other Professors from the Interuniversity Master for all the knowledge, help and courage that they gave me through this process. Lastly, to my family and my boyfriend that were always with me in this journey and for understanding the moments of absence**.**

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Signature: Catarina Carvalho Ferreira

Action recognition of point-light displays presented with semantically (in)congruent auditory stimuli : Behavioral correlates

Abstract

Humans are experts in identifying and understanding other's movements. The visual information is often enough for a very accurate action identification. However, actions usually have an associated sound and thus the integration between visual and auditory modalities can benefit perception. Considering the improved identification driven by the integration of multiple sensory information, researchers began to investigate how such inputs are integrated in a unified percept. Following such line of investigation, we aimed to evaluate if the presence of congruent/incongruent action sounds would improve/impair recognition of the visual Point-Light Display (PLD) of human actions and if non-biological PLDs accompanied by action sounds would bias participants into a false perception of visual human action. Therefore, participants were presented with several masked human and scrambled PLD videos accompanied by sounds that were either biological or a white noise and asked to judge if the video depicted a human figure or not and in the affirmative case, they should name the action. Results showed a significant enhancement for audiovisual biological congruent when compared to the visual biological paired to auditory noise and unimodal visual stimuli. Similarly, a significantly better performance on action recognition occurred for the audiovisual biological congruent condition when compared to the unimodal visual stimuli. Lastly, considering the scramble stimuli we found a significant bias towards the identification of a human figure for the visual stimuli paired to auditory noise when compared to the unimodal visual scramble condition.These findings suggest that adding coincident sounds to a human action visual display impacts human figure identification and action perception in biological PLDs and that insignificant sounds might lead to confound perception of non-biological PLDs.

Keywords: action recognition; biological motion; multisensory integration; point-light displays; semantic congruence

Table of Contents

Table of Figures

List of Tables

Introduction

One important characteristic of the human being is the ability to identify and understand others' movements with high accuracy (Lapenta et al., 2017). Such ability is also essential for survival to both prey and predator (Thornton, 1998) and further allows effective communication and interaction (Thompson & Parasuraman, 2012).

Humans are so skilled in recognizing and interpreting other`s movements that accurate interpretations occur even in the absence of pictorial information, e.g., hand shadows (Alaerts et al. 2009, 2015) and point-light displays (Ulloa & Pineda 2007; Krakowski et al. 2011; Lapenta et al., 2017).

Although there are different explanations concerning how humans process biological motion, one integrative view is that people's understanding of movements and actions occur throughout their experience and a higher-level processing (Blakemore & Decety, 2001). In this sense, it is suggested that the processing of observed actions occurs in an active manner, thus comprehension is elaborated by the connection between sensorial and motor nodes (Pulvermüller & Fadiga, 2010). Therefore, action recognition occurs by transposing the observed actions into one's own repertoire (Buccino et al., 2014, Blake & Shiffar, 2007; Rizzolatti & Craighero, 2004).

One way to investigate how humans process and recognize actions is by using Point Light Displays (PLD) which are generated by filming human actors wearing dark suits with reflector patches attached to their major joints. The first author to use this technique was Johansson (1973) and it was later adopted in many studies concerned with the neural basis of biological motion perception (e.g., Vonck et al., 2015).

According to Thompson and Parasuraman (2012) there are multiple brain regions involved in serial and parallel processes that account for perceiving and understanding human actions. The middle temporal area (MT) is involved in the detection of motion whereas the inferior parietal lobe (IPL), the left anterior intraparietal sulcus (aIPS) and the inferior frontal gyrus (IGF) seem responsible for interpreting and understanding the goals and intentions of the actions. Finally, the ventral premotor cortex (vPMC), the superior temporal sulcus (STS) and the extrastriate body area are likely linked to the kinematic goals and the integration of action intentions and body form, respectively.

Multiple sensorimotor brain regions have been shown active during action observation and thus this circuitry has been called Action Observation Network (AON) (Blake & Shiffrar, 2007; Caspers et al., 2010; Gazzola et al., 2007). The core regions of this network include the IFG, PMC, IPL, posterior MT, and the pSTS (Blake & Shiffar, 2007; Kilroy et al., 2019). The AON contributes to the understanding of

others' actions by mapping those actions into our own repertoire. Specifically, its activation is believed to reflect an internal simulation of the observed actions, which is corroborated by the recruitment of motor representations, in accordance to the Penfield homunculus, involved in the observed action (Buccino et al., 2001 Lapenta et al., 2013). Therefore, it is involved with the processing of the goals and intentions of others' actions (Blake & Shiffar, 2007; Gazzola et al., 2006; Ortigue et al., 2010).

Behaviourally, relevant human motion information is received through multiple sensory modalities, and our perception is shaped by the co-occurrence and integration of these sensory events (Thomas & Shiffrar, 2013). When we are presented with two stimuli modalities, such as auditory and visual, e.g., by adding action auditory information to a PLD video, the observer can either perceive them as referring to the same unitary audiovisual event or referring to two separate unimodal events. This percept is mainly modulated by temporal and spatial congruence of the two stimuli which hint the observer concerning if the two stimuli are or not arriving from the same source (Spence, 2007).

Integrating multisensory inputs require a coordinate work of a large-scale brain network, including the auditory (van Atteveldt et al., 2007a, 2007b; Zhou et al., 2020) and visual (Macaluso et al., 2004; Zhou et al., 2020) cortices, the frontoparietal dorsal attention network (Binder, 2015) and other regions, such as superior colliculi (Calvert et al., 2001), the insula (Bushara et al., 2001; Lamichhane et al., 2016), the inferior parietal cortex (Adhikari et al., 2013; Dhamala et al., 2007) and the superior temporal sulcus (Marchant et al., 2012; Noesselt et al., 2012; Stevenson et al., 2010; Stevenson et al., 2011). In particular, STS has been demonstrated as an important multisensory hub (Bolognini & Maravita, 2011; Bolognini et al., 2010; Marques et al., 2014; Stevenson et al., 2011) composed by two anatomically distinct subregions being one more sensitive to low-level temporal synchrony, and other responsible for processing the multisensory integration at a higher cognitive level (Zhou et al., 2020).

Several psychophysical studies have shown that meaningful congruence impacts multisensory perception of actions (e.g., Eramudugolla et al., 2011; Thomas & Shiffrar, 2013), human speech (e.g., Grant & Seitz, 2000; Vatakis et al., 2008) and emotions (e.g., Collignon et al., 2008) by promoting faster and more accurate perception.

At the action perception domain several researchers demonstrated that meaningful congruence of multimodal inputs improves perception. Thomas and Shiffrar (2010), developed a study comparing the detection of masked point light walkers in visual (unimodal) and in audiovisual conditions. At the audiovisual conditions the participants heard either with simple tones or footsteps sound temporally and spatially coincident with the movement of the point light walker. They found increased detection sensitivity when visual displays were paired with the audio cues (footsteps) and the detection sensitivity decreased

when the displays were paired with simple tones. Demonstrating the role of experience and the meaning of the auditory input in the action recognition process. Aligned with such results, Van der Zwan and colleagues (2009) also demonstrated that meaningful congruence can influence multisensory processing, but in this case regarding gender information. Specifically, by using unambiguous gender specific auditory information paired to visual PLD of female walking sequences they found that the perceived gender was significantly influenced by the auditory cues when visual gender was ambiguous, therefore the gender-ambiguous PLD when paired with auditory female sequences were judged more frequently as being female walkers than when presented with no auditory stimulus, even when using extreme male and extreme female PLD, thus indicating that such results were not a result of response bias and/or cognitive artefact. Additionally, such effects were not only a product of paring any auditory information. In a second experiment they used gender-neutral auditory cues and those did not significantly influence perceived PLD gender. Furthermore, Arrighi et al., (2009) also measured the perception of audiovisual tap dancing and found that the audio tap information influenced the visual motion information, but only when AV stimuli were synchronous.

These studies combined bring strong psychophysical evidence that meaningful congruence along with visual and temporal synchrony of visual and auditory information influence visual sensitivity and the final perception of human movements.

However, some contradictory findings have also been reported, questioning the weight of temporal congruence (Jack & Thurlow, 1973). In this sense, Thomas and Shiffrar (2013) demonstrated that hearing footsteps sound enhanced visual sensitivity to walking motion, regardless of whether they are temporally synchronous with the PLD. Therefore arguing that temporal synchrony of multiple sensory inputs can be beneficial but doesn't seem crucial regarding human action perception.

In sum, the above mentioned studies suggest that adding sounds to a human action visual display can impact visual sensitivity and movement identification. Therefore, action perception, as many other cognitive functions, benefits from multisensory inputs and their integration. When two congruent sensory modalities are presented the attention directing towards the target stimuli of the task is increased. In addition, the congruence of these auditory stimuli can benefit, for example, the perception and identification of actions.

Still, most studies use simple tones and/or only one type of movement (e.g., point-light walking videos). Further, the stimuli employed are typically presented for a long period of time, which does not allow to evaluate if the AV integration is also beneficial in short and punctual action stimuli. Finally, usually only the auditory stimuli is manipulated and not the visual one. Herein, we aim to fill such gaps by bringing

three novel approaches. Firstly, we paired action sounds with meaningful or not meaningful visual actions, in contrast with the simple tones that are usually used. Secondly, we used short videos (~1 s long), contrary to other studies that use long and repetitive videos. This is extremely important since the impact of auditory stimuli on action recognition hasn't been checked in shorter amounts of time, and this could bring insights on if auditory stimuli can be beneficial even in very punctual and brief movements. Thirdly, we manipulated both auditory and visual stimuli, i.e., we proposed a pool of visual action movements congruently or incongruently paired to a pool of action sounds. which could clarify the weight of each modality and if participants tend to rely more on one of the modalities and further, we used both meaningful and non-meaningful sounds with non-meaningful visual stimuli, in order to see if auditory action sounds can bias participants inducing human perception in scrambled point light motion, i.e., not depicting a human action.

Hitherto, we aim to bring novel insights with our methodological approach to further explore how auditory stimuli impact the recognition of human actions. Specifically, we sought to evaluate if the presence of congruent/incongruent action sounds would improve/generate wrong (respectively) recognition of visual PLD actions and further, if the non-biological PLD accompanied by action sounds would generate a fake perception of human action. Therefore, we hope to add evidence on how the audiovisual integration affects behavioral responses of the broad domain of action perception, thus unveiling human perception when the audiovisual in(congruency) is integrated during action recognition, which frequently happens in our daily lives.

Methods

Participants

Participants were recruited through the credits platform system provided by the School of Psychology of the University of Minho. Inclusion criteria were: males and females, with 18 to 40 years old, with normal or corrected-to-normal vision; no history of neurological or psychiatric disorders. All participants gave their informed consent previous to participation in this experiment and were informed of their right to withdraw their participation at any time and that their data would be anonymized.

This study was elaborated in accordance with the declaration of Helsinki and was approved by the local Ethics Committee to the Investigation on Social and Human Science of the University of Minho, Braga, Portugal (CEICSH 069/2020). Initially, a total of 119 participants were recruited online. Subjects were debriefed about the aims of the study before starting the survey. Twenty subjects were excluded for

completing less than 75% of the experiment. Thus the final sample consisted of 99 participants aged between 18 and 38 years old (Mean = 20.88 ; SD= 3.70) residing in Portugal, 93 right-handed and 6 left-handed. None of the participants reported having psychiatric disorders. Concerning medication intake, 2 participants reported using allergy-related medication, and 1 reported the use of anxiolytics.

Conception of the stimuli

Visual Stimuli

This investigation used the database created by Lapenta and colleagues (2017), which is composed of short videos of one cycle of Point-Light Scrambled and Human movements, performed by athletes at natural velocity. From their pool of 15 recommended human actions we have selected 7 to compose our visual biological (Vbio) condition. The criteria of selection was based on the action itself having a sound related to the movement; for example, 'stretching' that does not have a specific action-related sound was not included. The selected movements were: walk, jump rope, jump, march, soccer kick, lateral step and jumping jack. In consonance we had also selected 7 from their recommended 14 scrambled (i.e., non-biological) movements to compose the visual scramble (Vscr) condition.

Because humans have a great ability to recognize human motion (Cutting et al., 1998) even when they are briefly presented as our stimuli (~1 s) we add a visual noise, i.e., a mask of moving point lights along with the actual stimuli) which is typically applied to increase difficulty of the task when using PLD (e.g., Hiris et al., 2005; Lu & Liu, 2006). For creating this mask, we add extra moving dots to visual stimuli, specifically more twenty two points, this generates noise to the visual stimuli thus making them more difficult to be identified by the participant. The creation of the 22th point mask and the video editing to add the scramble masking was done on Blender software 2.91.2 (Blender Foundation and Institute, Community, 2018). The Blender software comprises an open and free 2D and 3D creation that supports the totality of 3D pipeline-modeling, rigging, composition, rendering, motion tracking and video editing feature (Community, 2018).

Auditory Stimuli

The auditory action stimuli were selected from different available online databases and the white noise sound used was provided by the Perception, Interaction and Usability Domain Lab from the Center of Computer graphics located in Guimarães, Portugal.

The sounds were equalized for perceived loudness following the Cambridge loudness model (Moore & Glasberg, 1997; also applied by Varlet et al., 2020) using Matlab (The MathWorks Inc., Natick,

MA, USA). This Matlab code was written for a procedure for the computation of loudness of steady sounds, called ANSI S3.4-2007 (ANSI, 2007), the model of Moore and Glasberg (2007), ending by being proposed as standardize method for calculating loudness, the ISO 532- 2 standard (ISO, 2016) .

Before adding the sounds to the videos, an online study with 48 participants (aged between 18 and 42 years old) was performed in order to evaluate if the sounds were perceived as the action they represent. The task, built at Qualtrics (Qualtrics, Povro, UT) online platform, consisted in the presentation of each sound twice and in random order. After hearing each sound participants had to choose an answer between two options of actions that the sound could be representative of. All sounds were correctly recognized, i.e., as the expected action, at a rate above chance (Table 1). Additionally, we computed each participant's mean score for each sound (considering 1 for correct and 0 for incorrect answers) and performed a one sample t-test for each sound to evaluate at the group level if the rate of accurate response significantly differed from zero. Furthermore, Bayesian equivalent tests were performed to report statistical evidence using Bayes Factor (BF_{10}) denoting the level of evidence of the alternative hypothesis. As can be seen in Table 1, the t-test analysis for all sounds were statistically different from zero and further, suggested as a more reliable alternative (e.g., Van der Linden et al., 2018) than the null hypothesis. According to Jeffreys (1961) guidelines, our obtained values (greater than 100) indicate a decisive evidence for H1, meaning that the responses from our sample in each sound was different from 0. Thus, considering that zero and one would represent wrong and right sound identification, we conclude that all of our selected sounds were perceived as representing the action that they intended to.

Table 1.

One sample T-test, Bayes Factor (BF_{10}) values for each sound, and Means and standard deviations of correct answers for each sound

Sounds	T statistic	df	p	Mean (SD)	BF_{10}
Jump Rope	54.87	47.0	<0.001	0.969 (0.17)	4.364[40]
Soccer Kick	67.19	47.0	<0.001	0.979 (0.20)	4.287[44]
Jump	20.34	47.0	<0.001	0.792 (0.41)	5.702[21]
Jumping Jack	11.21	47.0	<0.001	0.594 (0.49)	9.832[11]
March	9.88	47.0	<0.001	0.563 (0.50)	1.715[10]
Walk	12.38	47.0	<0.001	0.625 (0.49)	2.906[13]
Lateral Step	12.35	47.0	<0.001	0.708 (0.46)	2.700[13]

Audiovisual stimuli combination

The auditory stimuli was combined with the visual stimuli in video using Adobe Premiere Pro (Adobe Systems, 2017). All videos were exported in mp4 format with a frame size of 800x600 pixels, a frame rate of 29.97 frames/s.

Importantly, for the temporal synchronicity of the audiovisual stimuli, we based the sound positioning on their natural occurrence at audio visual biological congruent condition (AVbioCong). In order to do that, videos of similar real actions were checked to assure that our combination of auditory and visual stimuli were done on an ecological valid temporal binding window. Following, for each corresponding counterpart of the other conditions, i.e., auditory biological and visual scramble condition (AbioVscr); white noise auditory stimuli with the visual scramble condition (AnoiseVscr); and audio visual biological incongruent (AVbioInc), the auditory stimuli were presented at the same time points as in the AVbioCong condition. Therefore, the sound in all different conditions was presented at compatible timings.

Finally, once we had the stimuli ready we set up the experimental design and programmed the task on Qualtrics (Qualtrics, Povro, UT).

General Procedure

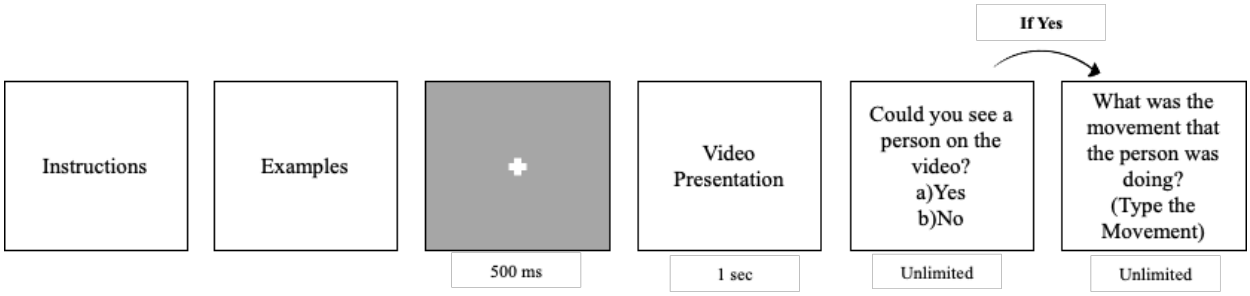
Participants were asked to complete an online survey, which took approximately 30 minutes. First, they read and virtually signed the informed consent, following they provided answers regarding dominant hand, psychiatric history, medical conditions and medication. Then, participants performed a judgement task composed by the following conditions:

- Visual Scramble (**Vscr**): In this condition we present only the scramble videos (without the sound);
- Visual Biological (**Vbio**): In this condition we present only the biological videos (without the sound);
- Visual Biological with Audio Congruent (**VbioAbioCong**): In this condition we present biological videos together with correspondent biological sounds;
- Visual Biological with Audio Incongruent (**VbioAbioInc**): In this condition we present biological videos together with a different sound also biological but not correspondent to the video;
- Visual Biological with Audio Noise (**VbioAnoise**): In this condition we present biological videos together with a white noise sound;
- Visual Scramble with Audio biological (**VscrAbio**): In this condition we present scramble videos together with biological sounds (the ones that we use on VbioAbioCong condition);
- Visual Scramble with Audio Noise (**VscrAnoise**): In this condition we present scramble videos together with a white noise sound.

Each biological and scrambled combination was presented twice in each category. Therefore, the task was composed of a total of 98 trials. Each trial started with a fixation cross in the center of the screen for 500 ms followed by one video presentation (approximate duration of 1 s). As soon as the video ended, participants were asked if they could see a person in the video (using a forced-choice of yes or no). When participants replied yes, a follow up question appeared asking what was the action that the person on the video was doing, for which they should respond by typing the perceived movement. After replying, another trial started. This repeated until all videos from all conditions were presented (please see Figure 1 for an overview of the online task).

Figure 1.

Overview of the online task



Note. This diagram represents an overview of how the task is organized and its presented, from the instructions and examples until the questions that are done regarding the video.

After the presentation and evaluation of all the videos the participants were asked to answer some questions about the task (e.g., the number of points on the video add difficulty to your response?).

Data Analysis

Analyses were performed with Jamovi software (*The jamovi project*, version 1.6.23), considering alpha=5%.

For the biological movement stimuli, two analyses were performed in order to evaluate differential perception according to the uni- and bimodal AV conditions.

Firstly, we evaluated participants' judgement regarding the Human identification. Therefore, we computed each individual percentage of correct responses, i.e., when participants answered seeing a person in the biological videos of each condition. Following a repeated measures ANOVA was conducted to compare the percentage of correct answers of human movement identification considering condition (VbioAbioCong vs VbioAnoise vs VbioAbioInc vs Vbio) as within-subjects factor.

Secondly, we evaluated the identification of Actions in the video. Similarly for Human analysis, we computed each individual percentage of correct responses, i.e., when participants perceived the correct visually presented action. A rmANOVA was conducted considering conditions (VbioAbioCong vs VbioAnoise vs VbioAbioInc vs Vbio) as within-subjects factor.

Regarding the Scramble movement stimuli, we also evaluated participants' judgement regarding Human identification. Therefore, we computed each individual percentage of correct responses, in this case, when they reported not seeing a person in the visually presented stimuli. Following, a rmANOVA was conducted considering condition (Vscrvs VscrAbio vs VscrNoise) as within-subjects factor.

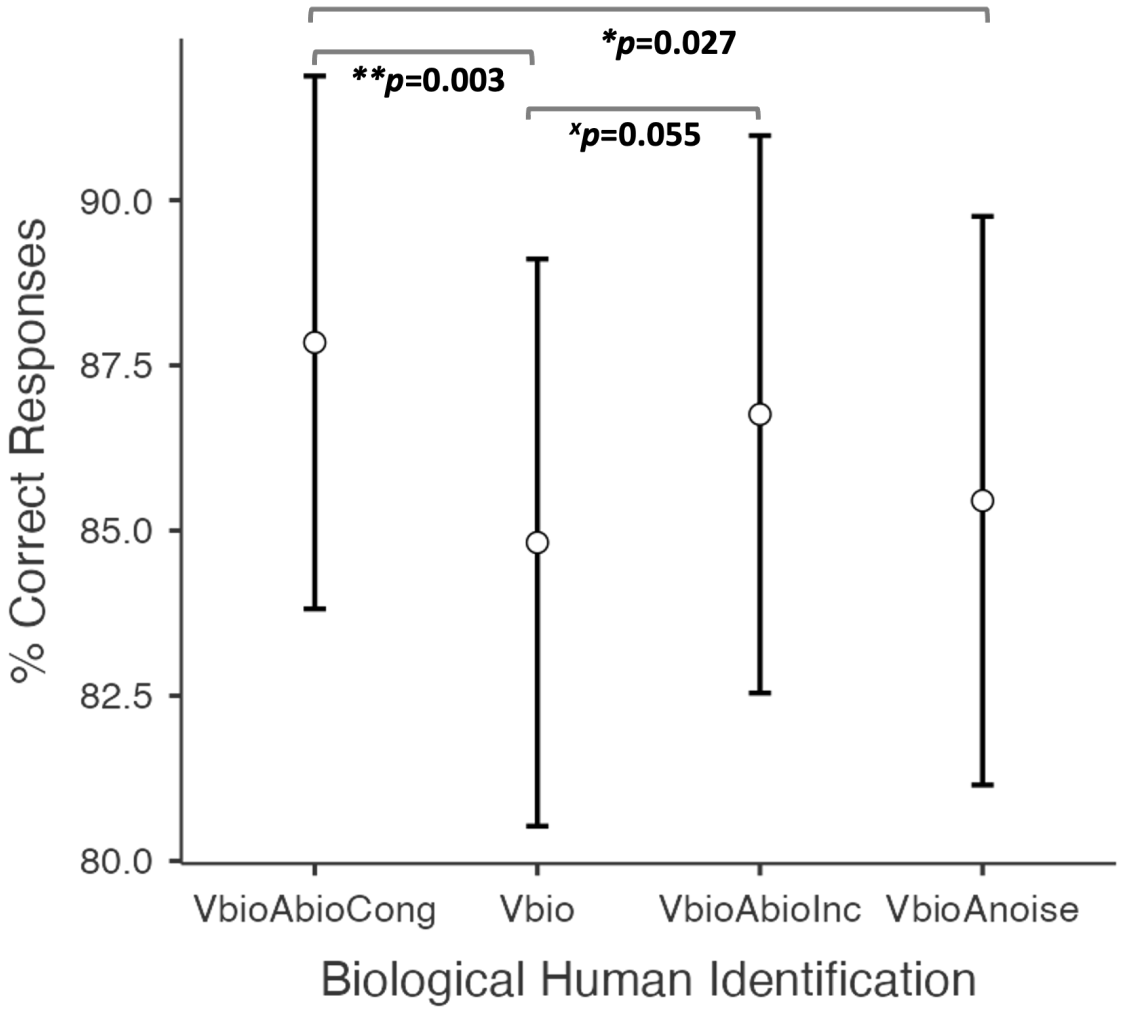
Results

Biological Human identification

The rmANOVA for perceiving a human figure presented a significant effect of condition ($F_{3, 294}$ = 5.67; $p < .001$; η^2_p = 0.055; BF_{Incl} = 18.1). Bonferroni post hoc comparisons showed that such effect was due to a significant difference between the percentage of correct answers in the VbioAbioCong condition when compared to VbioAnoise (p = 0.027; BF_{10} = 5.703) and Vbio (p = 0.003; BF_{10} = 44.705) (Figure 2). Specifically, when AV biological stimuli are congruently presented, participants identify more the human figure when compared to the VbioAnoise, and Vbio (Table 2). The comparison between Vbio and VbioAbioInc (p = 0.055; BF_{10} = 3.081), showed a trend for a significant effect, supported by the bayesian analysis that showed substantial evidence for the alternative hypothesis. The comparisons between VbioAbioCong and VbioAbioInc (p = 1.00; BF_{10} = 0.270), VbioAbioInc and VbioAnoise (p = 0.450; BF_{10} = 0.525), Vbio and VbioAnoise (p = 1.00; BF_{10} = 0.143) were not significantly different. The percentage of correct answers can be seen in Table 2.

Figure 2.

Correct responses for Biological Human identification



Note: VbioAbioCong condition resulted in higher human identification when compared to Vbio and VbioAnoise conditions.

* $p < .05$; ** $p < .01$; x p = trend for a significant effect

Table 2.

Mean and Standard deviation of the percentage (%) of correct responses on Biological Human Identification

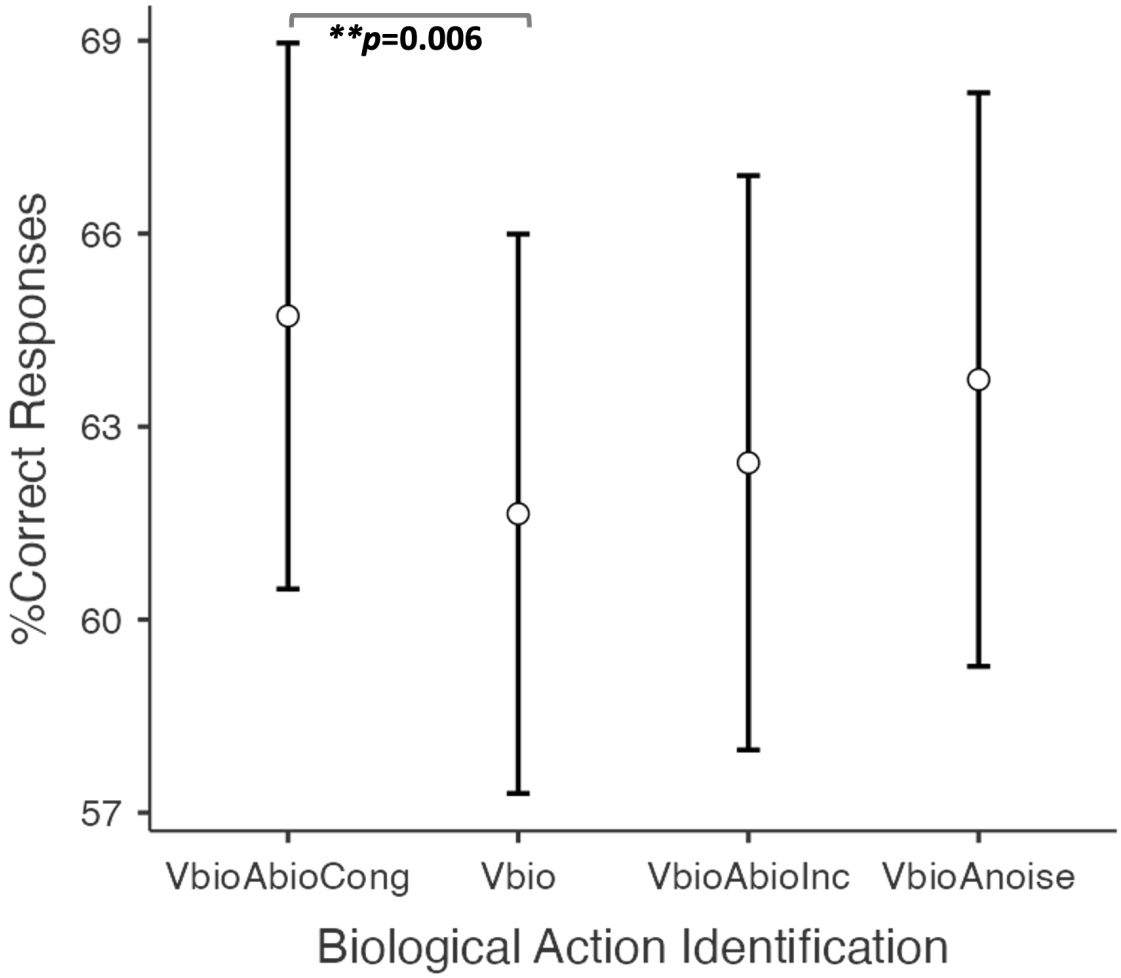
Condition	Mean (SD) %
VbioAbioCong	87.8(20.2)
Vbio	84.8(21.5)
VbioAbioInc	86.8(21.1)
VbioAnoise	85.5(21.6)

Biological Action identification

The rmANOVA for identification of actions yielded a significant effect of condition ($F_{3,294}$=3.86 ; p =0.010; η^2_p=0.038;BF_{Incl}= 1.76). Bonferroni post hoc comparisons showed that such effect was due to a significant difference between the percentage of correct answers in the VbioAbioCong condition when compared to Vbio (p = 0.006; BF_{10}= 22.78) (Figure 3). Specifically, when VbioAbioCong stimuli are presented, participants identify more correct actions when compared to Vbio (Table 3). The comparisons between Vbio and VbioAbioInc (p = 1.00; BF_{10}= 0.157), Vbio and VbioAnoise (p = 0.201;BF_{10}= 1.013), VbioAbioCong and VbioAbioInc (p = 0.124; BF_{10}= 1.526), VbioAbioCong and VbioAnoise (p = 1.00; BF_{10}= 0.172), VbioAbioInc and VbioAnoise (p = 1.00; BF_{10}= 0.231), were not significantly different. The percentage of correct answers can be seen in Table 5.

Figure 3.

Correct responses for Biological Action identification



Note: VbioAbioCong condition resulted in promoted more correct action identification when compared to Vbio condition.

$^{**}p < .01$

Table 3.

Mean and Standard deviation of the percentage (%) of correct responses on Biological Action Identification

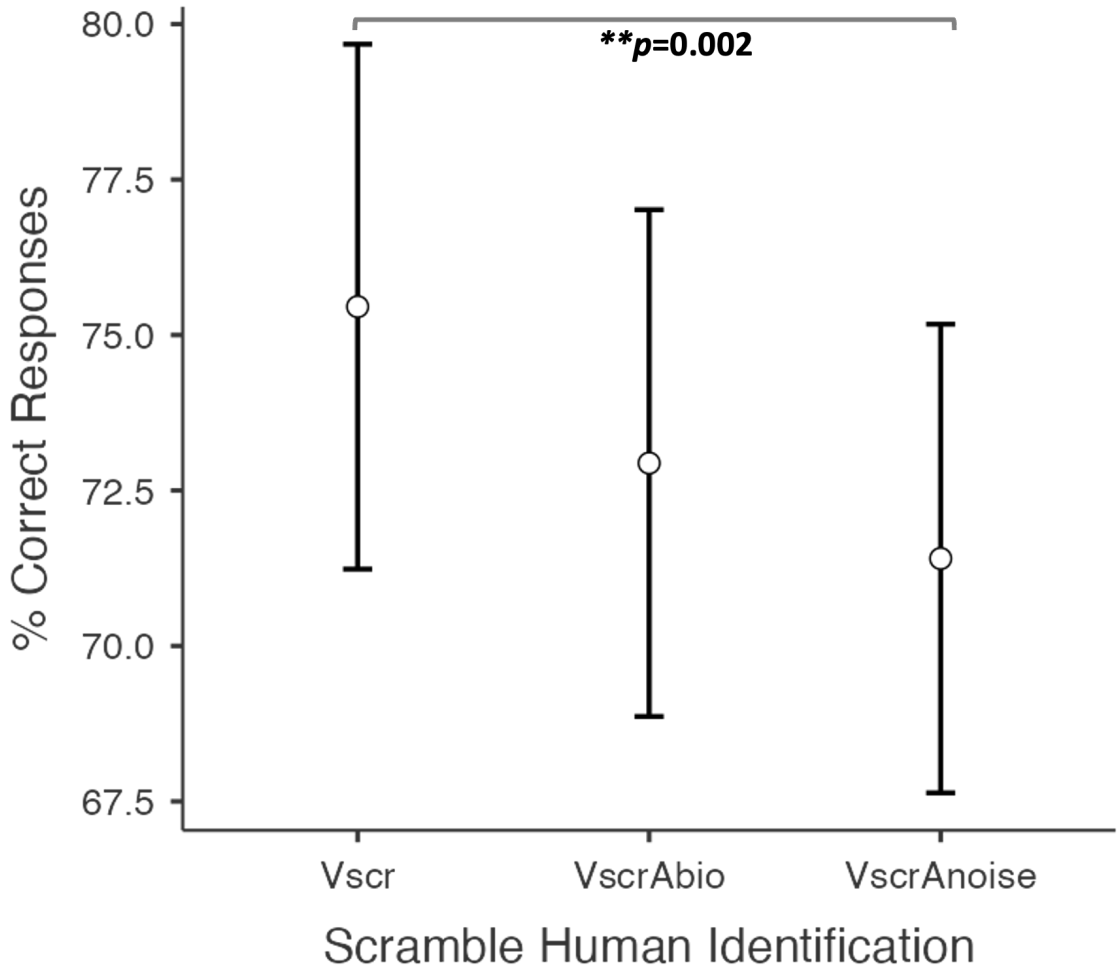
Condition	Mean (SD) %
VbioAbioCong	64.7(21.3)
Vbio	61.6(21.8)
VbioAbioInc	62.4(22.4)
VbioAnoise	63.7(22.4)

Scramble Human identification

The rmANOVA for perceiving a human figure yielded a significant effect of condition $F_{2,196}$=6.68 ; p =0.002; η^2_p=0.064;BF_{Incl}= 13.8). Bonferroni post hoc comparisons showed that such effect was due to a significant difference between the percentage of correct answers in the Vscr condition when compared to VscrNoise (p = 0.002; BF_{10}= 33.43) (Figure 4). Specifically, in the VscrAnoise conditions participants perceived more the human figure in the stimuli that actually did not depict it. The comparison between Vscr and VscrAbio (p = 0.070; BF_{10}= 1.377), showed a trend that we assumed to be not significant as the bayesian analysis showed anedoctal evidence for the null hypothesis. The comparison between VscrAbio and VscrNoise (p = 0.515; BF_{10}= 0.277) was also not significant. The percentage of correct answers can be seen in Table 4.

Figure 4.

Correct responses for Scramble Human identification



Note: VscrAnoise condition resulted in less accurate response, i.e., bias for more human identification, when compared to the unimodal Vscr condition.

** $p < .01$

Table 4.

Mean and Standard deviation of the percentage (%) of correct responses on Scramble Human Identification

Condition	Mean (SD) %
Vscr	75.5(21.2)
VscrAbio	72.9(20.4)
VscrAnoise	71.4(18.9)

Discussion

The environment simultaneously stimulates several of our senses, this is also true during human-human interactions (Thornton, 1998). For example, often one must integrate action sounds or disentangled sounds from other sources when observing and interpreting other´s movements. Considering the relevance of the auditory cues in human action discrimination, the central aim of this study was to explore how action and non-action sounds might impact the recognition of visual representation of human shape and actions. Specifically, we sought to evaluate if the presence of congruent/incongruent action sounds would improve/impair recognition of the visual PLD of human actions and further, if non-biological PLD accompanied by action sounds would bias participants into a false perception of visual human action.

The main findings of this study were: i. a significant enhancement of human identification for AV biological and congruent stimuli when compared to visual biological paired to auditory noise and to unimodal visual stimuli; ii. a significant bias towards the identification of the human figure on visual scramble displays paired to auditory noise when compared to the unimodal scrambled condition; iii. a better performance of action recognition for AV biological congruent when compared to the unimodal visual stimuli.

The improvement in biological human identification at AV biological congruent stimuli when compared to unimodal visual stimuli and visual biological stimuli paired to auditory noise suggests that congruent AV multisensory integration occurred and benefited perception. Such results are aligned with

previous findings showing that visual sensitivity to human figure is enhanced by the presence of temporally coincident sounds (e.g., Thomas & Shiffrar, 2010; 2013) and extends such findings by demonstrating that semantic congruence also plays a role.

However, in contrast to the initial hypothesis, there was no difference between the AV biological congruent and incongruent stimuli. It can be argued that any action sound suffice to improve human identification because of its human nature, as the semantic conflict present in the biological incongruent condition may have less robust effect. Still, the incongruent condition also didn't differ from the VbioAnoise or Vbio, but there was an important trend with substantial evidence for it to differ from VBio. Therefore, it is arguable that there might be subtle differences in the processing of each condition that our task is not sensitive to capture.

This is in line with the findings of Laurienti et al., (2004) where semantically congruent combinations of color circles and color words vocalizations but not their incongruent combinations resulted in enhancement on the cross modal discrimination task. This task required the participants to press a red or blue button according to the stimuli that was being presented (red or blue circle or the vocalization of the word blue or red). The congruent conditions occurred when the color of the circle matched the color word vocalization. On the other hand the incongruent condition consisted of a green circle paired with the blue or red vocalization, or a blue or red circle paired with the vocalization of the word green.

The present results do not allow for a more complex evaluation regarding the role of AV semantic congruence of actions. Further studies with similar paradigms coupling physiological measures of brain activity are needed to clarify the level of integration of AV modalities regarding both their nature (i.e., biological vs. non-biological) and semantics (congruent vs. incongruent multimodal biological stimuli). In particular, the superior temporal sulcus (STS) shows a particular activation to audiovisual stimuli when compared to the unimodal (auditory or visual) stimuli (Bruce et al., 1981; Hein et al., 2007). Therefore, the STS could be a target area to investigate using our audiovisual paradigm, in order to find if our congruent and incongruent audiovisual conditions activate this area in the same way, or if the congruence plays a role in the activation of this area. Additionally, it would be important to see if the medial prefrontal cortex (dACC/pre-SMA) and more caudal regions of left dorsolateral prefrontal cortex would be activated by our AV incongruent stimuli, since a study from Mayer et al., (2017) found that these specific prefrontal areas were activated while resolving audiovisual stimulus conflict.

In turn, the null effect when comparing VbioAnoise and Vbio was expected, and is in accordance with the modality appropriateness hypothesis (Welch & Warren, 1980). This hypothesis proposes that in

the case of bimodal stimulation, the sensory system with the higher acuity with respect to the most critical aspects of a task plays a predominant role on how multisensory inputs are integrated (Talsma et al., 2010; Welch & Warren, 1980). In the VbioAnoise condition the sensory system with higher acuity was the visual one, while the auditory cue was a non-informative white noise. Consequently, the responses given in the VbioAnoise and Vbio were similar.

The biological human identification findings corroborate our initial hypothesis on the effect of multimodal meaningful stimuli providing more cues and, therefore, increasing human visual detection when compared to unimodal or non-meaningful pairings. In turn, the semantic conflict in the incongruent condition may have a less robust effect.

Regarding the scramble stimuli, our study also adds a novelty to the previous literature by evaluating the interference of sounds in human figure identification where in reality, there is not a human figure being displayed. Specifically, we showed that the presence of auditory noise induced more human figure perception when compared to Vscr condition. Previous research showed similar findings, demonstrating that the white noise presented in the pace of step-like motion, but not having the spectral characteristics of footsteps, activates the pSTS/pMTG region (Callan et al., 2017), a region known to be involved in biological motion processing (e.g., Blake & Shiffrar, 2007). The articulation of these previous findings with our results suggests that white-noise sounds combined to visual ambiguous or non-action visual stimuli somehow augment human motion processing and thus human figure identification, however this argument needs to be further explored, as we presented the white noise at crucial time-points of the real action sound, but it was a punctual, i.e., not paced, auditory cue.

Contrary to what we expected to see in scramble conditions, the presence of a biological sound with a scramble visual stimuli did not bias the participants to see more human figures. Saygin et al., (2008) found that biological sounds enhanced visual sensitivity in the scramble stimuli and inverted point light walkers, but such effects are less robust when compared to the multisensory judgments of the upright point light walkers. Considering that there was a trend between VscrAbio and VscrNoise that did not reach statistical significance, the presence of biological sound might not have been enough for participants to attribute the sound of the action that is being heard to the scramble stimuli. This might be due to the shortness of our sound stimuli and it is also possible that the visual stimuli itself interfered with the actual sound recognition. Alternatively, it could be due to the small effect size and maybe a larger sample size could yield different and/or clearer results. It is worthy to mention that the Bayesian analysis on this comparison showed anecdotal evidence (Jeffrey, 1961) for the alternative hypothesis.

Regarding the action identification, only the combination of AV meaningful and congruent sounds improved performance when compared to the unimodal modality. These results are in line with the findings on biological human identification, which demonstrated that there is an enhancement with synchronous multimodal information over the unimodal information during the analysis of human action. Yet again, the AV congruent condition did not differ from any other AV pairings (i.e., VbioAbioInc and VbioAnoise). These results are partly aligned with the literature demonstrating that at action perception level, meaningful congruent inputs improve perception (e.g., Arrighi et al., 2009; Eramudugolla et al., 2011; Thomas & Shiffrar, 2013) and such effects are likely mediated by AV multisensory and attentional mechanisms. In this direction Alink et al., (2008) investigated how the human brain integrates auditory and visual motion showing a relationship between auditory and visual brain activations that suggests that audiovisual motion capture relies on attentional shifts from the auditory to visual modality. One can hypothesize that there is an attentional directing towards visual stimuli due to the presence of sound. Therefore our results might be a combination of attentional and multisensory integration effects and that could be the reasons why we only see a significant difference between unimodal and AbioVbioCong conditions.

Moreover, it is likely that incongruent action sounds and the white noise sounds act as a confusion factors. In the AVbioInc and VbioAnoise the audiovisual information is either incongruent or ambiguous and, consequently will imply in divided attention as defended by many authors (e.g., Nakayama and Joseph, 1998; Thompson & Parasuraman, 2012; Thornton et al., 2002). In this line of thought Alink et al. (2018) tested how the brain integrates motion information when this information is different between sensory modalities, as it is the case of our AVbioInc and VbioAnoise conditions. The authors found that when the audiovisual information is not congruent, there is a decrease in the activation of the auditory motion complex (AMC). On the other hand, the visual motion area hMT/V5+ increased the levels of activation. Consequently, we can also argue that the reason why we didn't see differences between our AV conditions could be that the participants mainly focused on the visual modality in an attempt to disentangle the discrepant visual and auditory information. Considering the findings from Alink and collaborators (2018), we can argue that the absence of effects between our AV pairings and the AVbioCong condition could have been related to a major activation in the visual areas compared to the auditory areas, however our experimental design do not allow for a concrete interpretation due to the lack of physiological measures that could provide information on specific brain area activations in the different conditions.

Altogether, we propose that further neuroimaging studies using our audiovisual paradigm could elucidate if AVbioInc and VbioAnoise conditions lead to the activation of the visual motion areas such as hMT/V5+ and decrease the auditory motion complex (AMC) recruitment (Alink et al., 2008). In addition, it would be interesting to evaluate if the attentional level varies and the frontoparietal dorsal attention network (Binder, 2015) is differently activated when AVbioCong and AVbioInc stimuli are presented. Ultimately, it would be of extreme relevance to explore if attentional areas such as frontoparietal dorsal attention network (Binder, 2015) and areas more related with the multisensory integration as the superior temporal sulcus (STS) (e.g., Marques et al., 2014) are being activated at the same time or if there is a specific order of the activation, when we present the stimuli from AV biological congruent condition.

It would also be interesting to explore the paradigm coupled to electroencephalography (EEG) or magnetoencephalography, which would allow comparisons of congruence detection. For example, these techniques would allow evaluating if AVbioInc when compared to AVbioCong condition would elicit a larger N4 component, since this component emerges in responses of violations of expectancies (Luck, 2005). N4 has been reported with larger amplitudes during the observation of audiovisual incongruent when compared to congruent presentation of musicians playing violin or clarinet notes paired with sounds of either instrument (Mado Proverbio et al., 2014).

In sum, neurophysiological data collection during our audiovisual paradigm could shed light on the brain systems involved in action identification, and whether they include regions linked to action observation network (AON) (e.g., which has been shown to be active during the action observation, and encompasses multiple series of sensorimotor brain regions, namely the IFG, PMC, IPL, posterior MT and the pSTS; e.g., Blake & Shiffrar, 2007; Caspers et al., 2010; Gazzola et al., 2007; Neal & Kilner, 2010), attribution of actions (e.g., ventral premotor cortex (vPMC) and the superior temporal sulcus (STS), frontoparietal dorsal attention network (Binder, 2015), modality-specific auditory and visual areas (e.g., Macaluso et al., 2004; Zhou et al., 2020), or motor components of mirror-neuron system (Rizzolatti & Craighero, 2004).

The better understanding of AV congruence of human action could have important implications for populations with visual impairments. In fact, some investigations show that populations with perceptual deficits, such as in hemianopia and hemineglect patients, demonstrating that the presence of audiovisual stimulation of the affected hemifield can improve perception of the visual events in the blind hemifield (Frassinetti et al., 2005). Thus it would be interesting to compare our unimodal visual biological and multimodal biological conditions in these patients with perceptual deficits, in order to investigate if there is a specifically perception enhancement on action recognition and human figure identification.

Moreover, it would be interesting to collect data of individuals with Autism Spectrum Disorder (ASD) using our task, since the Mirror-Neuron System (MNS) and action decoding is known to be compromised in such individuals (Lapenta & Boggio, 2014). The MNS plays a critical role in superior functions such as learning through imitation, action comprehension and language development (Lapenta & Boggio, 2014; Rizzolati et al., 2001). Previous research showed a differential Mu suppression (i.e. lower Mu rhythm desynchronization) in autistic population, suggesting that there is dysfunction in the observation/execution system (Bernier et al. , 2013; Bernier et al., 2007), since the Mu rhythm is thought to be an indirect indicator of mirror neuron activity (Perkins et al. 2010). Importantly, Mu desynchronization has been demonstrated also when human actions are depicted in PLD compared to scrambled PLD (Ulloa & Pineda, 2007). Still such studies with ASD used only visual stimuli and therefore it would be interesting to explore the Mu rhythm in this population comparing unimodal and AV action stimuli, such as the case of our paradigm that contains simple, short, direct daily actions in PLD paired with congruent and incongruent sounds.

Conclusion

The central aim of this study was to explore how action and non-action sounds might impact the recognition of human shape and specific actions. Based on our results it can be concluded that visual sensitivity to human figures is increased by the presence of temporally and semantically congruent action sounds, demonstrating that semantic congruence plays a significant role in human figure identification; non-action sounds, in this case white noise, bias the participants to see more human figures on the non-biological PLDs, and action recognition is improved when visual biological stimuli is combined with action congruent sounds, compared to unimodal modality.

The enhanced visual perception of biological actions in the presence of auditory semantically congruent sounds brings important insights on multisensory audio-visuomotor integration that can have practical implications at interpersonal level. Specifically, the presence of sound optimizes perception due to multisensory integration mechanisms that combine multiple sources of complementary information allowing a better understanding other person's behavior and intentions, thus making interpersonal interaction less ambiguous. Our study adds to the previous literature on how audiovisual integration affects behavioral responses at human and action perception domain, and further that auditory stimuli impacts recognition of human actions and human figure identification when it's congruent with the presented visual action.

It is important to highlight that our study coupled the psychological domains such as biological motion perception and multisensory integration. Such approach integrates perception-action systems providing further demonstration of the flexibility and versatility of the human sensory systems in optimizing its performance at human and action perception.

Considering the methodological approach, the study had the innovative aspects of collecting two information types namely, human identification and action identification. Also, we provided multiple controls, specifically, unimodal vs bimodal stimuli; multiple controls for AV semantic congruence (biological congruent vs. biological incongruent vs. white noise sound) and visual control stimuli (biological vs. non-biological). Additionally, the validation of auditory and visual stimuli was carefully conducted. These factors provide an important differential to enable different controls to a better interpretation of the significant effects presented.

The study is limited by the absence of neurophysiological measures and the fact that the task was performed online. Data interpretation could be clearer if neurophysiological measures, such EEG and or functional magnetic resonance imaging (fMRI), were correlated to the behavioral performance. Specifically, neurophysiological measures could contribute to a clearer interpretation regarding: i. the semantic conflict in the incongruent condition, at human figure identification level; ii. the lack of significance between audiovisual biological congruent condition and the other AV pairings (i.e., VbioAbioInc and VbioAnoise), at action identification level. In turn, online data acquisition did not allow to standardize the volume of the sound to all participants nor to assure that task was performed without external distractors.

Finally, follow-up studies are suggested, throughout the discussion, in order to better understand how meaningful related sounds enhance visual sensitivity to human figure and improve the action recognition, at a neurophysiological level. Extending our investigation by combining techniques to assess the underlying neural network and also how AV integration plays a role in deficitary population, could bring important contributions. In particular, it would be ultmost interesting to investigate these effects in populations with visual, motor or auditory impairments, such as hemianopia and hemineglect patients. Additionally, further studies could use our audiovisualmotor task on populations with a dysfunction in the observation/execution system, in particular the Autism Spectrum Disorder (ASD) population.

References

Adhikari, B. M., Goshorn, E. S., Lamichhane, B., & Dhamala, M. (2013). Temporal-order judgment of audiovisual events involves network activity between parietal and prefrontal cortices. *Brain Connectivity, 3*, 536–545. https://doi.org/10.1089/brain.2013.0163

Adobe Systems. (2017). Adobe Premiere Pro (Version S6 2017 64x) [Windows].

Alaerts, K., Nayar, K., Kelly, C., Raithel, J., Milham, M. P., & Di Martino, A. (2015). Age-related changes in intrinsic function of the superior temporal sulcus in autism spectrum disorders. *Social Cognitive and Affective Neuroscience, 10*(10), 1413–1423.doi:10.1093/scan/nsv029

Alaerts, K., Van Aggelpoel, T., Swinnen, S. P., & Wenderoth, N. (2009). Observing shadow motions: Resonant activity within the observer's motor system? *Neuroscience Letters, 461*(3), 240–244.doi:10.1016/j.neulet.2009.06.055

Alink, A., Singer, W., & Muckli, L. (2008). Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience, 28(11), 2690–2697.

ANSI (2007). ANSI S3.4-2007. Procedure for the computation of loudness of steady sounds (American National Standards Institute, New York).

Arrighi, R., Marini, F., & Burr, D. (2009). Meaningful auditory information enhances perception of visual biological motion. *Journal of vision*, *9*(4), 1–7. https://doi.org/10.1167/9.4.25

Bernier, R., Aaronson, B., & McPartland, J. (2013). The role of imitation in the observed heterogeneity in EEG mu rhythm in autism and typical development. *Brain and cognition*, *82*(1), 69–75. https://doi.org/10.1016/j.bandc.2013.02.008

Bernier, R., Dawson, G., Webb, S., & Murias, M. (2007). EEG mu rhythm and imitation impairments in individuals with autism spectrum disorder. *Brain and cognition*, *64*(3), 228–237. https://doi.org/10.1016/j.bandc.2007.03.004

Binder, M. (2015). Neural correlates of audiovisual temporal processing–comparison of temporal order and simultaneity judgments. *Neuroscience, 300*, 432–447. https://doi.org/10.1016/j.neuroscience.2015.05.011

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

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience, 2*(8), 561–567. https://doi.org/10.1038/35086023

Bolognini, N., & Maravita, A. (2011). Uncovering multisensory processing through non-invasive brain stimulation. *Frontiers in Psychology*, *2*(MAR), 1–10. https://doi.org/10.3389/fpsyg.2011.00046

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental brain research, 160*(3), 273–282. https://doi.org/10.1007/s00221-004-2005-z

Bolognini, N., Papagno, C., Moroni, D., & Maravita, A. (2010). Tactile temporal processing in the auditory cortex. *Journal of Cognitive Neuroscience*, *22*(6), 1201–1211. https://doi.org/10.1162/jocn.2009.21267

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of neurophysiology*, *46*(2), 369–384. https://doi.org/10.1152/jn.1981.46.2.369

Buccino G. (2014). Action observation treatment: a novel tool in neurorehabilitation. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 369(1644), 20130185. https://doi.org/10.1098/rstb.2013.0185

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., & Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *The European journal of neuroscience*, *13*(2), 400–404.

Bushara, K. O., Grafman, J., Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *The Journal of Neuroscience, 21*, 300–304. https://doi.org/10.1523/JNEUROSCI.21-01-00300.2001

Callan, A., Callan, D., & Ando, H. (2017). The Importance of Spatiotemporal Information in Biological Motion Perception: White Noise Presented with a Step-like Motion Activates the Biological Motion Area. *Journal of cognitive neuroscience*, *29*(2), 277–285. https://doi.org/10.1162/jocn_a_01046

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage, 14*, 427–438.

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE Meta-Analysis of Action Observation and Imitation in the Human Brain. *Neuroimage, 50*, 1148–1167. doi: 10.1016/j.neuroimage.2009.12.112

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain research*, *1242*, 126–135. https://doi.org/10.1016/j.brainres.2008.04.023

Community, B. O. (2018). Blender - a 3D modeling and rendering package. Stichting Blender Foundation, Amsterdam. Retrieved from http://www.blender.org

Cutting, J. E., Moore, C., & Morrison, R. (1988). Masking the motions of human gait. *Perception & Psychophysics, 44*(4), 339–347. https://doi.org/10.3758/BF03210415 detection of masked speech and non-speech sounds. *Brain and Cognition, 75*, 60–66.

Dhamala, M., Assisi, C. G., Jirsa, V. K., Steinberg, F. L., & Kelso, J. A. S. (2007). Multisensory integration for timing engages different brain networks. *Neuroimage, 34*, 764–773. https://doi.org/10.1016/j.neuroimage.2006.07.044

Eramudugolla, R., Henderson, R., & Mattingley, J. B. (2011). Effects of audio-visual integration on the detection of masked speech and non-speech sounds. *Brain and cognition, 75*(1), 60–66. https://doi.org/10.1016/j.bandc.2010.09.005

Frassinetti, F., Bolognini, N., Bottari, D., Bonora, A., & Làdavas, E. (2005). Audiovisual integration in patients with visual deficit. *Journal of cognitive neuroscience, 17*(9), 1442–1452. https://doi.org/10.1162/0898929054985446

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the Somatotopic Auditory Mirror System in Humans. *Current Biology, 16*, 1824–1829. https://doi.org/10.1016/j.cub.2006.07.072

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The Anthropomorphic Brain: The Mirror Neuron System Responds to Human and Robotic Actions. *Neuroimage, 35*, 1674–1684. https://doi.org/10.1016/j.neuroimage.2007.02.003

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America, 108*(3), 1197–1208.

Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *27*(30), 7881–7887. https://doi.org/10.1523/JNEUROSCI.1740-07.2007

Hiris, E., Humphrey, D., & Stout, A. (2005). Temporal properties in masking biological motion. *Perception & psychophysics, 67*(3), 435–443. https://doi.org/10.3758/bf03193322

ISO (2016). ISO/DIS 532-2. Methods for calculating loudness — Part 2: Moore-Glasberg method (International Organization for standardization, Geneva)

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the ‘‘ventriloquism’’ effect. *Perceptual and Motor Skills, 37*, 967–979.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*, 195–204.

Kilroy, E., Cermak, S. A., & Aziz-Zadeh, L. (2019). A review of functional and structural neurobiology of the action observation network in autism spectrum disorder and developmental coordination disorder. *Brain Sciences*, *9*(4). https://doi.org/10.3390/brainsci9040075

Krakowski, A. I., Ross, L. A., Snyder, A. C., Sehatpour, P., Kelly, S. P., & Foxe, J. J. (2011). The neurophysiology of human biological motion processing: A high-density electrical mapping study. *NeuroImage, 56*(1), 373–383. https://doi.org/10.1016/j.neuroimage.2011.01.058

Lamichhane, B., Adhikari, B. M., & Dhamala, M. (2016). Salience Network Activity in Perceptual Decisions. *Brain Connectivity, 6*, 558–571. https://doi.org/10.1089/brain.2015.0392

Lapenta, O. M., & Boggio, P. S. (2014). Motor network activation during human action observation and imagery: Mu rhythm EEG evidence on typical and atypical neurodevelopment. *Research in Autism Spectrum Disorders*, *8*(7), 759-766. https://doi.org/10.1016/j.rasd.2014.03.019

Lapenta, O. M., Minati, L., Fregni, F., & Boggio, P. S. (2013). Je pense donc je fais: transcranial direct current stimulation modulates brain oscillations associated with motor imagery and movement observation. *Frontiers in human neuroscience*, *7*, 256. https://doi.org/10.3389/fnhum.2013.00256

Lapenta, O. M., Xavier, A. P., Côrrea, S. C., & Boggio, P. S. (2017). Human biological and nonbiological point-light movements: Creation and validation of the dataset. *Behavior Research Methods, 49*(6), 2083–2092. https://doi.org/10.3758/s13428-016-0843-9

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental brain research, 158*(4), 405–414. https://doi.org/10.1007s00221-004-1913-2/

Lu, H., & Liu, Z. (2006). Computing dynamic classification images from correlation maps. *Journal of vision, 6*(4), 475–483. https://doi.org/10.1167/6.4.12

Luck, S. J. (2005). *An introduction to event-related potential technique*. The MIT press. Cambridge, MA.

Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage, 21*, 725–732. https://doi.org/10.1016/j.neuroimage.2003.09.049

Mado Proverbio, A., Calbi, M., Manfredi, M., & Zani, A. (2014). Audio-visuomotor processing in the Musician's brain: an ERP study on professional violinists and clarinetists. *Scientific Reports, 4*, 1-10. https://doi.org/10.1038/srep05866

Marchant, J. L., Ruff, C. C., & Driver, J. (2012). Audiovisual synchrony enhances BOLD responses in a brain network including multisensory STS while also enhancing target-detection performance for both modalities. *Human Brain Mapping, 33*, 1212–1224. https://doi.org/10.1002/hbm.21278

Marques, L. M., Lapenta, O. M., Merabet, L. B., Bolognini, N., & Boggio, P. S. (2014). Tuning and disrupting the brain-modulating the McGurk illusion with electrical stimulation. *Frontiers in Human Neuroscience*, *8*(AUG), 1–9. https://doi.org/10.3389/fnhum.2014.00533

Mayer, A. R., Ryman, S. G., Hanlon, F. M., Dodd, A. B., & Ling, J. M. (2017). Look Hear! The Prefrontal Cortex is Stratified by Modality of Sensory Input During Multisensory Cognitive Control. *Cerebral cortex (New York, N.Y. : 1991), 27*(5), 2831–2840. https://doi.org/10.1093/cercor/bhw131

Moore, B. C. J., & Glasberg, B. R. (1997). A model of loudness perception applied to cochlear hearing loss. *Auditory Neuroscience, 3*, 289–311.

Moore, B. C., & Glasberg, B. R. (2007). Modeling binaural loudness. *The Journal of the Acoustical Society of America*, *121*(3), 1604–1612. https://doi.org/10.1121/1.2431331

Nakayama K, Joseph, J. S. (1998). Attention, pattern recognition and popout in visual search. In: Parasuraman, R. *The Attentive Brain* (pp. 279-298). MIT Press, Cambridge.

Neal, A., & Kilner, J. M. (2010). What is simulated in the action observation network when we observe actions?. *The European journal of neuroscience*, *32*(10), 1765–1770. https://doi.org/10.1111/j.1460-9568.2010.07435.x

Noesselt, T., Bergmann, D., Heinze, H. J., Münte, T., & Spence, C. (2012). Coding of multisensory temporal patterns in human superior temporal sulcus. *Frontires Integrative Neuroscience, 6*. https://doi.org/10.3389/fnint.2012.00064

Ortigue, S., Sinigaglia, C., Rizzolatti, G., & Grafton, S., T. (2010). Understanding Actions of Others: The Electrodynamics of the Left and Right Hemispheres. A High-Density EEG Neuroimaging Study. *PLoS ONE, 5*, e12160. https://doi.org/10.1371/journal.pone.0012160

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. Nature *Reviews Neuroscience, 11*(5), 351–360. https://doi.org/10.1038/nrn2811

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Saygin, A. P., Driver, J., & de Sa, V. R. (2008). In the footsteps of biological motion and multisensory perception: judgments of audiovisual temporal relations are enhanced for upright walkers. *Psychological science*, *19*(5), 469–475. https://doi.org/10.1111/j.1467-9280.2008.02111.x

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology, 28*(2), 61–70. https://doi.org/10.1250/ast.28.6

Stevenson, R. A., Altieri, N. A., Kim, S., Pisoni, D. B., & James, T. W. (2010). Neural processing of asynchronous audiovisual speech perception. *Neuroimage, 49*, 3308–3318. https://doi.org/10.1016/j.neuroimage.2009.12.001

Stevenson, R. A., VanDerKlok, R. M., Pisoni, D. B., & James, T. W. (2011). Discrete neural substrates underlie complementary audiovisual speech integration processes. *Neuroimage,55*, 1339–1345. https://doi.org/10.1016/j.neuroimage.2010.12.063

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in cognitive sciences, 14*(9), 400–410. https://doi.org/10.1016/j.tics.2010.06.008

The jamovi project (2021). *jamovi*. (Version 1.6) [Computer Software]. Retrieved from https://www.jamovi.org.

The survey for this paper was generated using Qualtrics software, Version 2021 of Qualtrics. Copyright © 2021 Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA. https://www.qualtrics.com

Thomas, J. P., & Shiffrar, M. (2010). I can see you better if I can hear you coming: action-consistent sounds facilitate the visual detection of human gait. *Journal of vision*, *10*(12), 14. https://doi.org/10.1167/10.12.14

Thomas, J. P., & Shiffrar, M. (2013). Meaningful sounds enhance visual sensitivity to human gait regardless of synchrony. *Journal of Vision, 13*(14), 1–13. https://doi.org/10.1167/13.14.8

Thompson, J., & Parasuraman, R. (2012). Attention, biological motion, and action recognition. *NeuroImage*, *59*(1), 4–13. https://doi.org/10.1016/j.neuroimage.2011.05.044

Thornton, I. M. (1998). The visual perception of human locomotion. *Cognitive Neuropsychology, 15*(6–8), 535–552. https://doi.org/10.1080/026432998381014

Thornton, I. M., Rensink, R. A., & Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception*, *31*(7), 837–853. https://doi.org/10.1068/p3072

Ulloa, E. R., & Pineda, J. A. (2007). Recognition of point-light biological motion: Mu rhythms and mirror neuron activity. *Behavioural Brain Research, 183*(2), 188–194. https://doi.org/10.1016/j.bbr.2007.06.007

van Atteveldt, N. M., Formisano, E., Blomert, L., Goebel, R. (2007a). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex, 17*, 962–974. https://doi.org/10.1093/cercor/bhl007

van Atteveldt, N. M., Formisano, E., Goebel, R., & Blomert, L. (2007b). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage, 36*, 1345–1360. https://doi.org/10.1016/j.neuroimage.2007.03.06

Van der Linden, L., Van de Putte, E., Woumans, E., Duyck, W., & Szmalec, A. (2018). Does Extreme Language Control Training Improve Cognitive Control? A Comparison of Professional Interpreters, L2 Teachers and Monolinguals. *Frontiers in psychology*, *9*, 1998. https://doi.org/10.3389/fpsyg.2018.01998

van der Zwan, R., Machatch, C., Kozlowski, D., Troje, N. F., Blanke, O., & Brooks, A. (2009). Gender bending: auditory cues affect visual judgements of gender in biological motion displays. *Experimental brain research*, *198*(2-3), 373–382. https://doi.org/10.1007/s00221-009-1800-y

Varlet, M., Williams, R., & Keller, P.E. (2020). Effects of pitch and tempo of auditory rhythms on spontaneous movement entrainment and stabilisation. *Psychological Research, 84*, 568–584. https://doi. org/10.1007/s00426-018-1074-8

Vatakis, A., Ghazanfar, A. A., & Spence, C. (2008). Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *Journal of vision*, *8*(9), 1–11. https://doi.org/10.1167/8.9.14

Vonck, S., Swinnen, S. P., Wenderoth, N., & Alaerts, K. (2015). Effects of transcranial direct current stimulation on the recognition of bodily emotions from point-light displays. *Frontiers in Human Neuroscience*, *9*(AUGUST), 1–8. https://doi.org/10.3389/fnhum.2015.00438

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological bulletin, 88*(3), 638–667.

Zhou, H. Yu, Cheung, E. F. C., & Chan, R. C. K. (2020). Audiovisual temporal integration: Cognitive processing, neural mechanisms, developmental trajectory and potential interventions. *Neuropsychologia*, *140*, 107396. https://doi.org/10.1016/j.neuropsychologia.2020.107396

Appendix

Universidade do Minho

Conselho de Ética

Comissão de Ética para a Investigação em Ciências Sociais e Humanas

Identificação do documento: CEICSH 069/2020

Relatores: Emanuel Pedro Viana Barbas Albuquerque e Marlene Alexandra Veloso Matos

Titulo do projeto: *Action recognition of point-light displays presented with semantically (in)congruent auditory stimuli: behavioral and electrophysiological correlates*

Equipa de Investigação: Catarina Carvalho Senra, Mestrado Interuniversitário em Neuropsicologia Clínica e Experimental, Escola de Psicologia, Universidade do Minho/Universidade de Lisboa/Universidade de Coimbra; Olivia Morgan Lapenta (Orientadora), Centro de Investigação em Psicologia, Escola de Psicologia da Universidade do Minho

PARECER

A Comissão de Ética para a Investigação em Ciências Sociais e Humanas (CEICSH) analisou o processo relativo ao projeto de investigação acima identificado, intitulado *Action recognition of point-light displays presented with semantically (in)congruent auditory stimuli: behavioral and electrophysiological correlates.*

Os documentos apresentados revelam que o projeto obedece aos requisitos exigidos para as boas práticas na investigação com humanos, em conformidade com as normas nacionais e internacionais que regulam a investigação em Ciências Sociais e Humanas.

Face ao exposto, a Comissão de Ética para a Investigação em Ciências Sociais e Humanas (CEICSH) nada tem a opor à realização do projeto, emitindo o seu parecer favorável, que foi aprovado por unanimidade pelos seus membros.

Braga, 6 de agosto de 2020.

O Presidente da CEICSH

(Acílio Estanqueiro Rocha)