



Spatial and temporal (non)binding of audiovisual rhythms in sensorimotor synchronisation

Olivia Morgan Lapenta^{1,2} · Peter E. Keller^{1,3} · Sylvie Nozaradan^{1,4} · Manuel Varlet^{1,5}

Received: 28 January 2022 / Accepted: 6 February 2023 / Published online: 14 February 2023
© The Author(s) 2023

Abstract

Human movement synchronisation with moving objects strongly relies on visual input. However, auditory information also plays an important role, since real environments are intrinsically multimodal. We used electroencephalography (EEG) frequency tagging to investigate the selective neural processing and integration of visual and auditory information during motor tracking and tested the effects of spatial and temporal congruency between audiovisual modalities. EEG was recorded while participants tracked with their index finger a red flickering (rate $f_V = 15$ Hz) dot oscillating horizontally on a screen. The simultaneous auditory stimulus was modulated in pitch (rate $f_A = 32$ Hz) and lateralised between left and right audio channels to induce perception of a periodic displacement of the sound source. Audiovisual congruency was manipulated in terms of space in Experiment 1 (no motion, same direction or opposite direction), and timing in Experiment 2 (no delay, medium delay or large delay). For both experiments, significant EEG responses were elicited at f_V and f_A tagging frequencies. It was also hypothesised that intermodulation products corresponding to the nonlinear integration of visual and auditory stimuli at frequencies $f_V \pm f_A$ would be elicited, due to audiovisual integration, especially in Congruent conditions. However, these components were not observed. Moreover, synchronisation and EEG results were not influenced by congruency manipulations, which invites further exploration of the conditions which may modulate audiovisual processing and the motor tracking of moving objects.

Keywords Frequency tagging · Motor tracking · Multisensory integration · Movement synchronisation · Steady-state evoked potentials

Introduction

The integration of multiple sensory modalities, such as visual and auditory inputs, is critical to promote adequate perception and behavioural responses in dynamic environments, as these signals often arise from common external events or objects (Macaluso and Driver 2005). The combination of co-occurring and co-localised visual and auditory cues can lead to enhanced processing of visual signals (Frassinetti et al. 2002). Importantly, such integration strongly relies on temporal and spatial congruency. Huddlestone and colleagues (2008) revealed that the perception of continuous rotational stimuli observed in unimodal auditory and visual conditions vanishes in bimodal conditions where visual vertical and auditory horizontal stimuli are intercalated, showing that spatiotemporal congruence is needed for effective multi-modal integration and perception.

Co-occurring and co-localised audiovisual (AV) stimuli are not only beneficial for perception, but also in motor

Communicated by Melvyn A. Goodale.

✉ Olivia Morgan Lapenta
oliviamlapenta@gmail.com

- ¹ The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith, Australia
- ² Present Address: Psychological Neuroscience Lab, Center for Investigation in Psychology, University of Minho, Rua da Universidade, 4710-057 Braga, Portugal
- ³ Center for Music in the Brain, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
- ⁴ Institute of Neuroscience, Université Catholique de Louvain, Woluwe-Saint-Lambert, Belgium
- ⁵ School of Psychology, Western Sydney University, Penrith, Australia

coordination tasks, including motor coordination with a moving object (Meyer et al. 2005) or another individual (Noy et al. 2017). In particular, Rosati and colleagues (2012) investigated the role of different types of auditory feedback in motor learning and motor control, as additional sensory information to visual and proprioceptive modalities. They found that auditory feedback simulating the sound of a rolling ball improved performance in visual tracking of a dot moving horizontally on the screen compared to conditions without sound feedback or error-related sound feedback. Thus, adding auditory information enhances perception of task-relevant information and improves sensory-motor coordination only when auditory information is congruent with visual information.

Investigations with neuroimaging techniques have been employed to uncover the brain structures and mechanisms underlying AV integration in motion perception and motor coordination. Alink et al. (2012) identified two areas encoding auditory motion stimuli, specifically the planum temporale as a key area for auditory motion perception and high-level visual cortex, suggesting a convergence of motion signals from both visual and auditory modalities. Von Saldern and Noppeney (2013) likewise found an important role of visual areas (hMT+/V5+) together with the planum temporale in the processing of auditory motion stimuli. They also revealed stronger activation of the putamen in audiovisual conditions, especially when stimuli were degraded and a bimodal advantage on motion discrimination accuracy was observed, which suggests that bimodal integration is favoured under challenging conditions.

Neuroimaging investigations in motor coordination tasks have been largely limited due to movement restrictions and artefacts, leaving the underlying audiovisual integration processes and the effects of spatial and/or temporal congruency unclear. However, growing evidence suggests EEG frequency tagging (i.e. steady-state evoked potentials (SSEPs)) as a promising method for addressing these questions (Gordon et al. 2019). By tagging sources of information in the environment at specific frequencies, their neural processing and integration can be reliably examined even if participants are moving (Nozaradan et al. 2015; Varlet et al. 2020). Neural populations involved in the processing of a tagged stimulus exhibit increased activity at this specific frequency (Gordon et al. 2019). SSEPs therefore offer the possibility to tag multiple stimuli, and disentangle their respective processing in EEG signals based on their tagging frequencies (Norcia et al. 2015; Renton et al. 2018; Varlet et al. 2020). Further, frequency tagging enables the investigation of the neural integration of two different sources of information, as reflected in increased amplitude at the intermodulation frequency components, corresponding to the addition or subtraction of the two fundamental tagging frequencies and their harmonics (Gordon et al. 2019).

However, while neural responses at the intermodulation frequencies have been shown within sensory modalities (e.g. between two visual stimuli) in numerous studies, evidence supporting neural responses at the intermodulation frequencies across sensory modalities is still lacking. Indeed, Giani et al. (2012) compared passive perception under uni- and bimodal conditions, and found responses at the intermodulation frequencies only in the unimodal visual condition but not in the bimodal audiovisual condition. In follow-up experiments using a detection task, the authors investigated if actively attending to either auditory or visual stimuli promote integration, and again they failed to identify amplitude increases at the intermodulation frequencies. Interestingly, in a study addressing the coupling between auditory and motor responses during sensorimotor synchronisation with beat sequences, Nozaradan and colleagues found activity at the intermodulation frequencies corresponding to the sum of the auditory beat frequency and participant's movement frequency while tapping a finger in time with every second beat (Nozaradan et al. 2015). These findings highlight the need for further investigations to better understand the relevance of intermodulation frequencies for understanding integration across sensory modalities.

Here, we investigated the relevance of EEG frequency tagging to better understand the neural processing of visual and auditory information and their integration underpinning effective motor synchronisation with a continuously moving stimulus, and determine how these processes are modulated by spatial and temporal (in)congruency between the two modalities. We conducted two experiments to address these questions in which we examined the motor tracking performance of a visual stimulus oscillating horizontally on a monitor with a complex movement trajectory while manipulating either the temporal or spatial congruency of an auditory stimulus with respect to the moving visual stimulus. Visual and auditory stimuli were tagged at specific frequencies ($f_V = 15$ Hz and $f_A = 32$ Hz, respectively) to examine the underlying neural tracking and integration of visual and auditory information and how they are modulated by temporal and spatial congruence. We hypothesised significant EEG responses at f_V and f_A tagging frequencies that would be greater in the Congruent conditions since temporal congruency has been shown to enhance the processing of multi-sensory inputs at sensory-specific stages of cortical processing (Nozaradan et al. 2012) and considering that spatially and temporally Incongruent conditions are likely to promote divided attention (De Jong et al. 2010). Further, significant amplitude at the intermodulation frequencies ($f_V \pm f_A$ —17 and/or 47 Hz) was expected in EEG recordings as signature of neural integration of visual and auditory information, especially in temporally and spatially Congruent conditions and in visual areas (Alink et al. 2012; von Saldern and Noppeney 2013; Giani et al. 2012).

Materials and methods

Participants

We invited 19 subjects to participate in the spatial sound manipulation experiment (Experiment 1—E1) and 19 to participate in the temporal sound manipulation experiment (Experiment 2—E2). The sample size was chosen based on an a priori power analysis with G-Power (V. 3.1.9.3) to detect medium effect sizes ($f=0.25$) with at least 80% power, in line with previous similar studies (Giani et al. 2012). The experiments were conducted in accordance to the Declaration of Helsinki and approved by the Human Research Ethics Committee at Western Sydney University (ethics reference number: H13092). All participants gave their written informed consent and complied with the following criteria: age between 18 and 50 years, right-handed, normal or corrected-to-normal visual acuity, and no known past or current auditory impairment, psychological or psychiatric disorders, and central nervous system injury. One participant in each experiment was excluded due to EEG recordings with abnormal noise. Therefore, the final sample of E1 involved 18 participants (13 female) aged between 18 and 42 years ($M=25.50$; $SD=6.82$) and of E2 involved 18 participants (11 female) aged between 18 and 42 years ($M=25.66$; $SD=6.83$).

EEG and motion-tracking recordings

Electrophysiological (EEG) data were recorded with a BioSemi ActiveTwo system (BioSemi, The Netherlands) by means of 64 active Ag–AgCl electrodes placed over the scalp according to the International 10/20 system. Additionally, two external Ag–AgCl electromyography (EMG) electrodes were positioned on the corner and bottom of the left eye to record blinks and eye movements. EEG and EMG were recorded at a sampling rate of 2048 Hz, and stored for off-line analyses.

Motion tracking was used to record the horizontal oscillations of participant's right index finger at a sampling rate of 240 Hz with 0.01 mm spatial resolution via a Polhemus LIBERTY motion tracker (Polhemus Ltd., VT, USA). The movement data were used in real time to control the horizontal position of the cursor on the monitor and saved on a PC for off-line analyses of the synchronisation performance.

Stimuli and procedure

Auditory stimuli were pure tones (sinusoidal waveform) with lateralised amplitude modulations in left and right audio channels (panning) presented via insert earphones (ER-1,

Etymotic Research, Elk Grove Village, IL, USA). The auditory stimuli were tagged using continuous (sinusoidal) pitch modulation at 32 Hz between 500 and 1100 Hz. In other words, the carrier frequency of the sound continuously and sinusoidally alternated between 500 and 1100 Hz at 32 Hz. We opted to modulate the pitch because sound amplitude modulation was already used to generate the panning (i.e. lateralised modulation of amplitude in the left and right audio channels) to create the perception of motion. Comfortable intensity at approximately 70 dB kept the same for all participants was used.

The moving target and the participant's cursor were red (RGB: 255,0,0) dots with 3.2 cm of diameter presented on a 24-inch VIEWPixx LCD monitor (VPixx Technologies, Saint-Bruno, Canada) with a 120 Hz refresh rate. Participants were seated approximately 60 cm from the monitor. The target visual stimuli flickered at 15 Hz and moved horizontally across the screen with a complex movement trajectory that changed every trial to make it more difficult for the participant to track. The target horizontal movement corresponded to the sum of three sine waves (1: 0.185 Hz and 14.5 cm; 2: 0.74 Hz and 9.7 cm; and 3: 1.11 Hz and 4.8 cm, of frequency and amplitude, respectively) with new random phase values for each trial. The participant's cursor was positioned 2 cm below the target, had no flickering, and its position was controlled in real time using motion-capture data.

Although visual SSEPs can be recorded in several ranges, usually studies opt for frequencies within 3–20 Hz with most of them using frequencies above 8–10 Hz (Norcia et al. 2015). We chose to present the visual stimuli at 15 Hz to avoid interference in alpha frequencies and lower frequencies due to the movements of the participants and the moving stimuli. 32 Hz for the auditory stimulus was selected to ensure that auditory fundamental and audiovisual intermodulation frequencies were also above movement and alpha frequencies while avoiding overlaps between their harmonics.

Audiovisual congruency was manipulated to generate three conditions for each experiment. For E1, we manipulated spatial congruency; specifically, sound would either (i) move congruently with the visual target by lowering the volume in the ear on the *opposite* side of the moving direction of the visual target while increasing the volume in the ear on the *same* side (Congruent), (ii) move incongruently with the target by lowering the volume in the ear on the *same* side of the moving direction of the visual target while increasing the volume in the ear on the *opposite* side, i.e. in antiphase (Incongruent), or (iii) stay at consistent volume in both ears, with no perception of auditory motion (No-motion). Combined amplitude (volume) across the two ears was the same for the three conditions. For E2, we manipulated time congruency; specifically, sound would either (i) move simultaneously with the visual target (No-delay); (ii)

move with a time delay of 300 ms (300 ms delay); or (iii) move with a time delay of 600 ms (600 ms delay) (Fig. 1A).

Upon arrival, all participants were informed about the procedure and provided written informed consent. After EEG preparation, participants were asked to seat comfortably on a chair with their right hand positioned on a wooden board over their lap. The motion sensor was attached to the participant's right index finger, which was occluded from sight using a custom-built cardboard box (Fig. 1B). Participants then received the verbal and on-screen instruction to do their best to track the horizontal displacements of the visual stimulus with the cursor on the screen that was controlled in real time by the horizontal oscillations of their index finger. Each participant was allocated either in E1 or E2. Initially, they performed practice trials in which one sample trial of each condition was presented, and thus a total of three practice trials. These practice trials allowed participants to familiarise to the use of the tracking sensor and the control of their cursor on the screen. After these practice trials, the experimenter made sure that participants

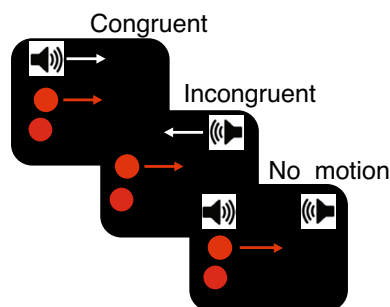
understood the task and were performing accordingly, and the actual experiment began.

Participants were presented with a total of 30 trials with a duration of 60 s, including 10 trials for each audiovisual combination. Ten different complex stimulus trajectories were used, each presented once in a randomised order for each of the three sound conditions of E1 and E2. At the completion of each trial, a neutral screen appeared informing the participant of the trial number, and again displaying the instruction to follow the target with the cursor. Participants set their own pace regarding trial progression, as they pressed the space bar on a keyboard situated beneath the monitor in front of them to continue to the next trial. A break was given halfway through the trials. At the completion of the 33 trials (3 practice + 30 experimental), participants had all sensors and electrodes removed and were then debriefed regarding the experiment. Participants took roughly 40 min to complete the task and the EEG preparation varied between 30 and 50 min. Therefore, the experiment was in total about 80 min long.

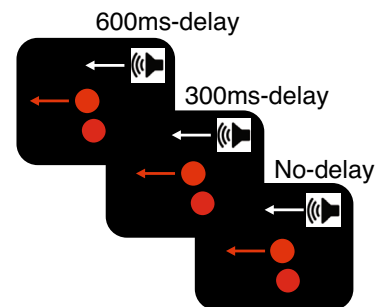
Fig. 1 HYPERLINK "sps:id::fig1lllocator::gr1llmediaobject::0" Experimental design. Panel A Schematic illustration of the different experimental conditions, the arrows illustrate the movement of auditory and visual stimuli, and the upper red dot illustrates the visual stimuli, whereas the bottom one represents the visual feedback of the motion tracker. Panel B The experimental setup

A Experimental trials

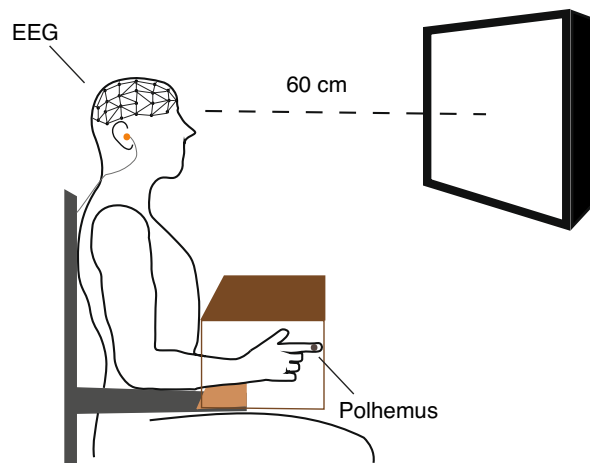
Experiment 1- Spatial Manipulation



Experiment 2 - Temporal Manipulation



B Participant setup



EEG data processing

Data were processed using Fieldtrip toolbox for EEG analysis (Oostenveld et al. 2011) in the MatLab software environment (The MathWorks, Natick, USA). When more than 50% of trials was contaminated with excessive EEG and/or EMG noise, the participant was excluded ($n_{E1} = 1$; $n_{E2} = 1$). For the remaining participants (18 in each experiment), channels with excessive noise were interpolated by the mean of surrounding channels; a maximum of three and of five EEG channels per participant was interpolated in Experiment 1 ($M_{E1} = 1$; $SD_{E1} = 0.88$) and Experiment 2 ($M_{E2} = 1.27$; $SD_{E2} = 1.28$), respectively. Trials were considered bad trials and rejected when visual inspection detected large waves in EEG due to loss of reference or ground channels. Each participant had maximum three rejected trials ($M_{E1} = 0.27$; $SD_{E1} = 0.8$; $M_{E2} = 0.22$; $SD_{E2} = 0.71$) due to poor EEG quality. Independent component analysis was performed using FASTICA algorithm, as implemented in Fieldtrip, to identify and reject eye blink artefacts. The average number of ICA components removed was $M_{E1} = 1.6$ ($SD_{E1} = 0.61$) and $M_{E2} = 1.94$ ($SD_{E2} = 0.85$). A 0.1 Hz high-pass filter was applied and data were notched filtered to remove 50 Hz power contamination and its harmonics. EEG data were re-referenced to the average of all EEG channels. Data were downsampled to 1000 Hz and stored for further analyses.

For each participant and condition, EEG epochs were averaged across trials in the time domain. Across-trial averaging in the time domain is expected to cancel out or at least markedly reduce the contribution of EEG signals that are not phase locked to the stimulation train, and therefore increase signal-to-noise ratio (Mouraux and Iannetti 2008). Subsequently, for each participant and condition, the obtained waveforms were then examined in the frequency domain by using a discrete Fourier transform (Frigo and Johnson 1998), using a Hanning window, yielding a power spectrum ranging from 0 to 100 Hz, since the spectrum was trimmed up to 100 Hz, with a frequency resolution of 0.017 Hz (i.e. 1/60).

To remove unrelated background noise due to spontaneous EEG activity, muscle activity or eye movements, and obtain estimates of the magnitude of the auditory and visual steady-state evoked potentials (SSEPs) and their intermodulation frequencies, data were normalised by performing a baseline subtraction. Noise was removed by subtracting, at each bin of the frequency spectra, the average power measured at ten neighbouring frequency bins excluding adjacent ones (Varlet et al. 2020; Lenc et al. 2018), with bin size determined by the frequency resolution of the Fourier transform. Finally, the neural tracking and integration of the visual and auditory stimuli were estimated by averaging the power of all EEG channels at the frequencies corresponding to visual (15 Hz) and

auditory (32 Hz) tagging frequencies and their intermodulation frequencies (17 and 47 Hz). Topographic maps were generated to visualise the distribution of the EEG responses at the frequencies of interest (Fig. 2).

Statistical analysis

All statistical analyses were carried out using JASP software (JASP Team 2018, <https://jasp-stats.org/>) with $\alpha = 5\%$ for analysis of variance and pairwise comparisons. Further, Bayesian equivalent tests were performed to report statistical evidence using Bayes factors (BFs), BF_{10} for paired sample comparisons and correlational analysis and BF_{incl} for ANOVAs denoting the level of evidence for the alternate hypothesis (non-signed difference), and the inclusion of a specific parameter in a model (ANOVA), respectively.

Behaviour

For each participant and condition, we computed the mean distance between the visual target and the participant's cursor. We then conducted for each experiment a one-way repeated-measures ANOVA considering Condition as a within-subject factor.

SSEPs

For each experiment, we conducted a one-way repeated-measures ANOVA considering Condition as a within-subject factor on the baseline-subtracted power at each frequency of interest, i.e. 15 Hz for visual and 32 Hz for auditory stimuli, and the intermodulation frequencies representing their integration (i.e. 17 and 47 Hz).

Furthermore, to examine the occurrence of significant EEG responses at the four frequencies of interest, we averaged the power spectra of all EEG channels together before baseline subtraction and computed Z-scores at each frequency bin as the difference in power between that frequency bin and the mean of the ten neighbouring frequency bins (excluding the two immediately adjacent frequency bins), divided by the standard deviation of those ten neighbouring bins. Z-scores were computed individually and at the group level (power spectra averaged across participants) for each condition and averaged between conditions. EEG responses at specific frequency bins were considered to be significant when the Z-score value was greater than 3.1 ($p < 0.001$, one-tailed), in line with previous studies that used frequency tagging techniques (Varlet et al. 2020; Jacques et al. 2016; Quek et al. 2018), which indicated signal power significantly greater than the noise background.

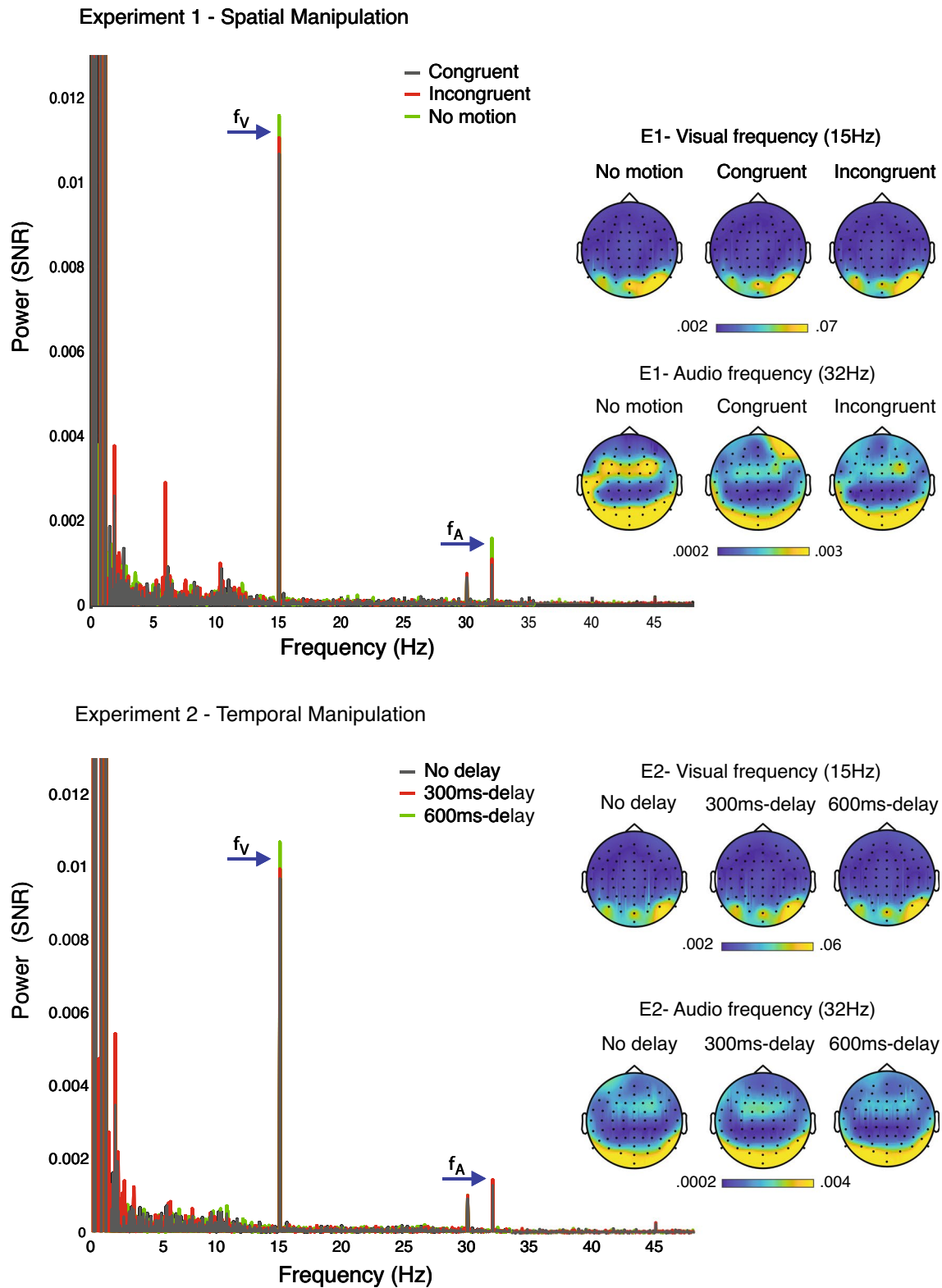


Fig. 2 Power as a function of frequency and topoplots of the target frequencies. Baseline-subtracted power spectra and topographic plots were generated based on the average of all electrodes. The blue arrows indicate the fundamental frequencies of interest for the visual

($f_V = 15$ Hz) and auditory ($f_A = 32$ Hz) stimuli. Upper and lower panels show the averaged data of all participants for Experiment 1 and 2, respectively

Correlation between behavioural and SSEPs data

Finally, we tested whether changes in EEG and behavioural responses across the different conditions were correlated, i.e. whether increase in SSEP power was accompanied by better motion-tracking performance. Therefore, we computed for each experiment EEG and behavioural changes relative to the control audio condition. For both mean distance and SSEP power, we subtracted the control condition (i.e. no motion in E1 and No-delay in E2) from the other conditions (i.e. Congruent and Incongruent in E1, and 300 ms and 600 ms delay in E2) and performed Pearson correlation analyses between the two for each frequency of interest.

Results

Motion-tracking performance

Experiment 1

The rmANOVA performed for the mean distance, with Greenhouse–Geisser correction because sphericity assumption was violated ($\chi^2(2) = 7.573$, $p = 0.023$), revealed no significant effect for condition ($F_{1,17} = 3.458$; $p = 0.061$; $\eta_p^2 = 0.169$, $BF_{Incl} = 1.500$), as depicted in Fig. 3.

Experiment 2

The rmANOVA performed for the mean distance, with Greenhouse–Geisser correction because sphericity assumption was violated ($\chi^2(2) = 13.878$, $p = 0.001$), revealed no significant effect for condition ($F_{1,17} = 2.823$; $p = 0.100$; $\eta_p^2 = 0.142$, $BF_{Incl} = 0.995$), as depicted in Fig. 3.

SSEPs

Experiment 1

Fundamental frequencies Reliable peaks at the fundamental frequencies (i.e. 15 and 32 Hz) based on Z-score data were observed at group level in all conditions. They were also reliably observed at individual level. Significant peaks at 15 Hz for the visual tagging were observed in all conditions and all participants. Significant peaks at 32 Hz for the audio tagging were found in 13, 15 and 15 over 18 participants in the Congruent, Incongruent and No motion conditions, respectively.

The rmANOVA performed on the SNR of EEG power at 15 Hz revealed no significant effect for condition ($F_{1,17} = 0.794$; $p = 0.460$; $\eta_p^2 = 0.045$, $BF_{Incl} = 0.251$). The rmANOVA on the EEG power at 32 Hz with Greenhouse–Geisser correction, because sphericity assumption was violated ($\chi^2(2) = 7.901$, $p = 0.019$), revealed a significant effect of condition ($F_{1,17} = 7.320$; $p = 0.007$; $\eta_p^2 = 0.301$, $BF_{Incl} = 15.704$). Bonferroni corrected pairwise comparisons revealed significant differences between the No motion

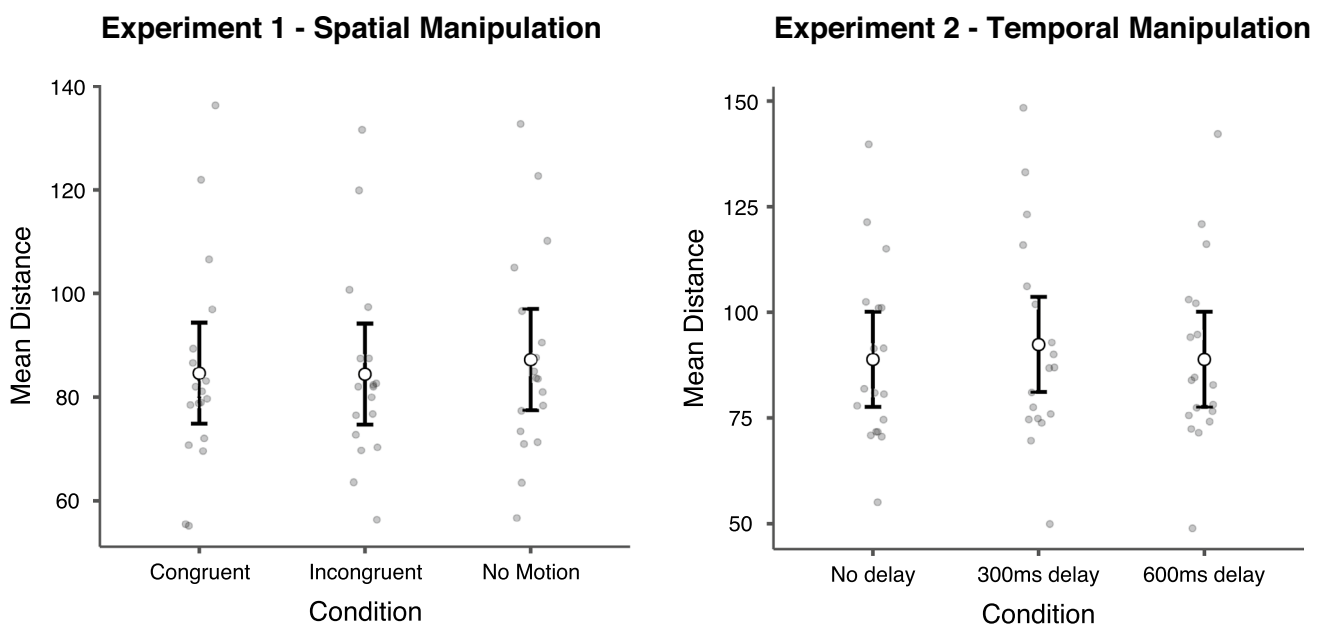


Fig. 3 Motion-tracking results. Graphs depict group and individual mean distance with bars representing the confidence interval between the visual target and the motion tracker cursor in each experimental condition for both experiments

condition when compared to Congruent ($t(17)=2.924$, $p=0.009$, $d=0.689$, $BF_{10}=5.528$) and Incongruent ($t(17)=2.870$, $p=0.011$, $d=0.676$, $BF_{10}=5.028$), but no differences between Congruent and Incongruent conditions ($t(17)=1.171$, $p=0.258$, $d=0.276$, $BF_{10}=0.440$). This result suggests larger EEG response to the audio stimulus in the No motion condition compared to the Congruent and Incongruent conditions. Grand-average power spectra for each condition are presented in Fig. 2 and EEG power at 15 and 32 Hz more specifically with individual participants' data points are presented in Fig. 4.

Intermodulation frequencies Analyses on Z-score data did not show significant peaks at the intermodulation frequen-

cies at the group level in any conditions and at the individual level in any participants and conditions. The rmANOVA revealed no significant effect of condition for 17 Hz ($F_{1,17}=0.197$; $p=0.822$; $\eta_p^2=0.011$, $BF_{Incl}=0.173$) or for 47 Hz ($F_{1,17}=1.371$; $p=0.267$; $\eta_p^2=0.075$, $BF_{Incl}=0.442$).

Analyses (not presented here) conducted on the harmonics of the fundamental and intermodulation frequencies, which can also contain relevant activity despite much smaller magnitude, did not show any significant effects either (all p values > 0.05).

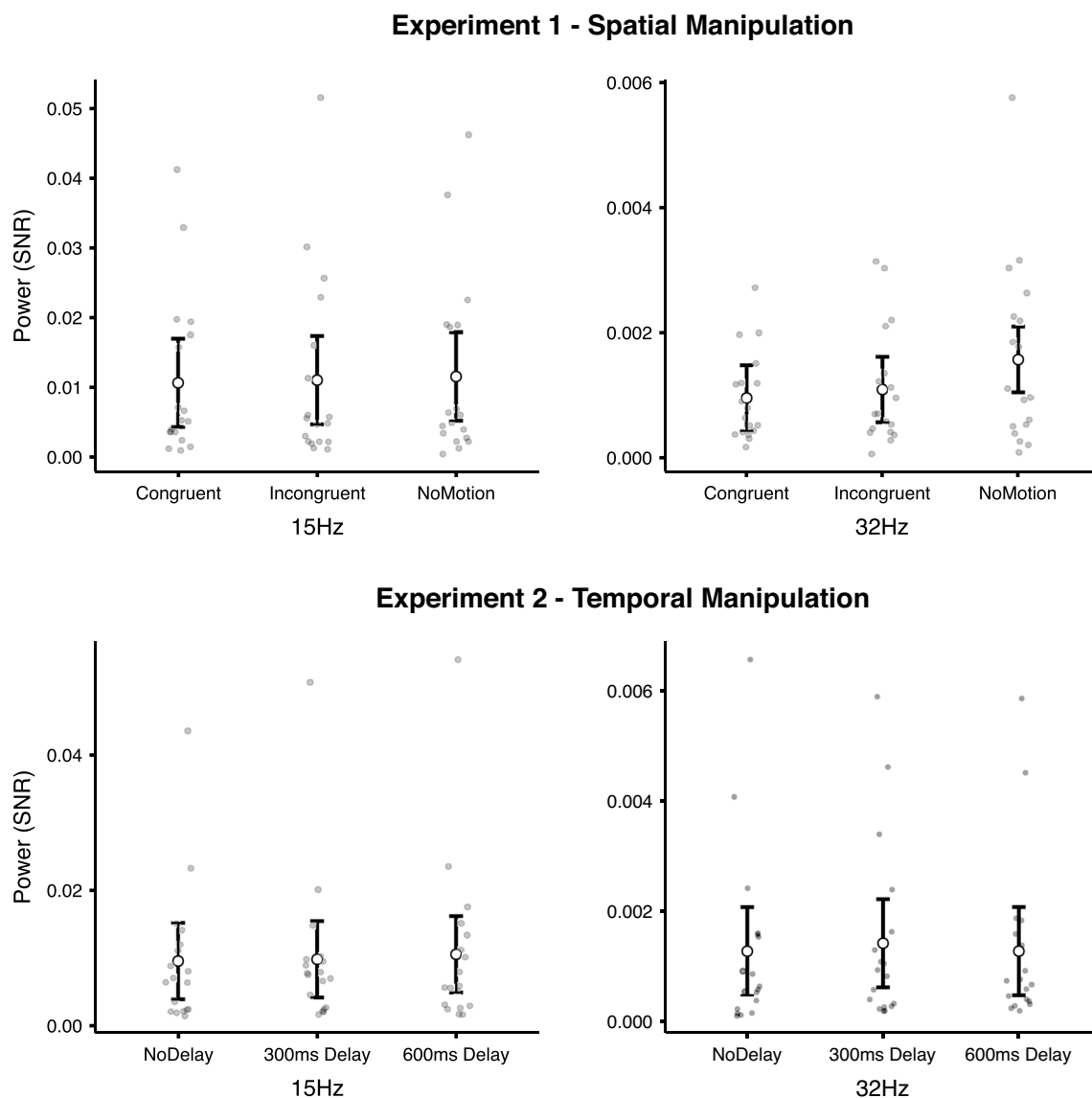


Fig. 4 Target frequencies across conditions. Graphs depict group and individual signal-to-noise ratio of the power spectra generated based on the average of all electrodes with bars representing the confidence interval

Experiment 2

Fundamental frequencies Z-score data indicated reliable peaks at the fundamental frequencies (i.e. 15 and 32 Hz) at the group level in all conditions. At the individual level, significant peaks at 15 Hz for the visual tagging were observed in all conditions and all participants, and significant peaks at 32 Hz for the audio tagging were found in 12, 13 and 13 over 18 participants in No-delay, 300 ms delay and 600 ms-delay conditions, respectively.

The rmANOVA performed on the EEG power at 15 Hz revealed no significant effect for condition ($F_{1,17}=1.222$; $p=0.307$; $\eta_p^2=0.067$, $BF_{\text{Incl}}=0.346$). The rmANOVA on the EEG power at 32 Hz also showed no significant effect of condition ($F_{1,17}=1.463$; $p=0.246$; $\eta_p^2=0.079$, $BF_{\text{Incl}}=0.405$). Grand-average power spectra for each condition are presented in Fig. 2 and EEG power at 15 Hz and 32 Hz, in particular with individual participants' data points are presented in Fig. 4.

Intermodulation frequencies Z-score analyses showed no significant peaks at the intermodulation frequencies at group level in any conditions and at individual level in any participants and conditions. The rmANOVA conducted on the intermodulation frequencies revealed no significant effect of condition for 17 Hz ($F_{1,17}=1.698$; $p=0.198$; $\eta_p^2=0.091$, $BF_{\text{Incl}}=0.482$) or for 47 Hz ($F_{1,17}=0.707$; $p=0.500$; $\eta_p^2=0.040$, $BF_{\text{Incl}}=0.272$).

Correlation between SSEPs and motion-tracking performance

Experiment 1

Correlational analyses on the changes in EEG power and motion-tracking performance relative to the control condition indicated no significant correlations between motion-tracking performance and EEG power at the 15 Hz visual tagging frequency in the Congruent condition ($r=0.007$, $p=0.978$, $BF_{10}=0.291$) and the Incongruent condition ($r=0.137$, $p=0.587$, $BF_{10}=0.334$) or EEG power at the 32 Hz audio tagging frequency in the Congruent condition ($r=-0.151$, $p=0.549$, $BF_{10}=0.344$) and the Incongruent condition ($r=-0.189$, $p=0.453$, $BF_{10}=-0.189$). Correlational analyses indicated no significant correlations between motion-tracking performance and EEG power at the 17 Hz intermodulation frequency in the Congruent condition ($r=-0.264$, $p=0.289$, $BF_{10}=0.491$) and the Incongruent condition ($r=0.050$, $p=0.844$, $BF_{10}=0.297$), and between motion-tracking performance and EEG power at the 47 Hz intermodulation frequency in the Congruent condition ($r=0.184$, $p=0.464$, $BF_{10}=0.374$) and the Incongruent condition ($r=0.198$, $p=0.431$, $BF_{10}=0.389$).

Experiment 2

Correlational analyses indicated no significant correlations between motion-tracking performance and EEG power at the 15 Hz visual tagging frequency in the 300 ms-delay condition ($r=-0.279$, $p=0.262$, $BF_{10}=0.523$) and the 600 ms-delay condition ($r=0.224$, $p=0.371$, $BF_{10}=0.423$) or EEG power at the 32 Hz audio tagging frequency in the 300 ms-delay condition ($r=-0.107$, $p=0.672$, $BF_{10}=0.317$) and the 600 ms-delay condition ($r=-0.138$, $p=0.585$, $BF_{10}=0.335$). Correlational analyses indicated no significant correlations neither between motion-tracking performance and EEG power at the 17 Hz intermodulation frequency in the 300 ms-delay condition ($r=-0.290$, $p=0.243$, $BF_{10}=0.550$) and the 600 ms-delay condition ($r=0.237$, $p=0.237$, $BF_{10}=0.442$), and between motion-tracking performance and EEG power at the 47 Hz intermodulation frequency in the 300 ms-delay condition ($r=0.283$, $p=0.255$, $BF_{10}=0.532$) and the 600 ms-delay condition ($r=-0.076$, $p=0.765$, $BF_{10}=0.304$).

Discussion

In this study, we used EEG frequency tagging to examine the selective neural processing of visual and auditory information and their integration in the tracking of a moving stimulus and determine how spatial and temporal congruency between the two modalities modulate these mechanisms and the motion-tracking performance.

As expected, the results revealed significant EEG activity at both visual and auditory tagging frequencies (15 and 32 Hz, respectively) in both experiments. Activities at these frequencies were reliably observed in the EEG recordings of most participants in all conditions. These findings are in accordance with previous studies showing neural synchronisation to fast periodic visual and auditory stimuli in the form of SSEPs at the stimulation frequencies (Nozaradan et al. 2011, 2012, 2015; Aissani et al. 2011). Further, we observed stronger EEG responses at the visual frequency (f_V) compared to the auditory frequency (f_A). Although a number of factors might have influenced this difference, such as differences in how neural activity related to processing of the visual vs. auditory stimulus project onto the scalp and are captured with EEG or/and intensity of the stimuli, voluntary sustained attention towards the visual stimuli constrained by the visuomotor tracking task might have played a significant role. Participants were instructed to track the visual stimulus irrespective of the auditory information and sustained attention has been shown to increase stimulus-driven electrophysiological activity (Kim et al. 2007, 2017).

Spatial and temporal audiovisual (AV) congruency was expected to result in greater EEG response at the 15 and

32 Hz tagging frequencies, but no differences between Congruent and Incongruent conditions were found in any of the two experiments for any of the two frequencies. For the AV spatial congruency manipulation in Experiment 1, although not directly of interest for the current research, a larger response at the 32 Hz auditory frequency was found for the No motion condition compared to both Congruent and Incongruent conditions, which might be due to the constant intensity resulting in facilitated tracking and/or higher average loudness despite the same combined intensity across the two ears. We expected that incongruent auditory stimuli would generate conflict and elicit divided attention, whereas congruent auditory stimuli would increase attention towards the visual stimuli. Therefore, a smaller EEG response at the visual frequency was expected for the Incongruent compared to Congruent conditions due to divided attention (De Jong et al. 2010), which would also be associated to poorer motor tracking performance. However, we found no such effects in the SSEPs nor in the motor performance, contrary to our expectations and to previous research that showed enhanced SSEPs in temporally Congruent AV conditions compared to unimodal and Incongruent bimodal conditions (Nozaradan et al. 2012). Importantly, previous research showed that unimodal visual and auditory modalities differ on sensorimotor synchronisation (Hove et al. 2013; Varlet et al. 2012). Specifically, tapping synchronisation improved with visual and decreased with auditory continuous pacing signals, whereas discrete auditory signals improved sensorimotor synchronisation (Hove et al. 2013; Varlet et al. 2012). It has been further demonstrated that both the continuity of the pacer and whether the pacing and movement patterns match contribute to visuomotor synchronisation, but that such facilitation is greater influenced in visual continuous pacers in comparison to the discrete ones (Zelic et al. 2018). In turn, a similar setup investigating effects of the match between pacer and movement patterns on auditory–motor synchronisation indicated improved and less variable synchronisation for discrete pacers and continuous movement, and it was argued that due to lower spatial resolution of the auditory sensory modality the benefits of spatial pattern matching on auditory–motor synchronisation might be limited (Zelic et al. 2019). Therefore, it is possible that the absence of clear landmarks on the auditory loudness modulation in this study influenced auditory–motor synchronisation and might explain the null results irrespective of audiovisual congruence. Furthermore, in our experiment, participants were instructed to track the visual stimulus, which likely made them focus on the visual information disregarding the auditory one, which could have contributed to the null findings amongst conditions both on SSEPs and motor performance, and also to the larger peak at the visual 15 Hz frequency.

Also contrary to our initial hypothesis, in both experiments we found no significant EEG activity at the

intermodulation frequencies (17 and 47 Hz), which were examined to investigate multisensory integration or segregation. No significant EEG activity at these frequencies was found in any participant or condition despite reliable responses at the fundamental frequencies. Despite non-significant effects at the behavioural level, activity at the intermodulation frequencies, at least for some of the 38 participants tested across the two experiments, could have been expected considering the large responses observed at both the visual and auditory tagging frequencies. While these results differ from those observed in previous unimodal studies, including in movement synchronisation tasks (Varlet et al. 2020), they corroborate the results of Giani et al. (2012), where intermodulation frequencies were not observed across visual and auditory stimuli in various experimental settings. Our study extends this research by showing that intermodulation frequencies across visual and auditory modalities do not necessarily occur in movement synchronisation tasks either.

Still, considering previous findings that successfully show audiovisual intermodulation frequencies (Regan et al. 1995; Drijvers et al. 2021), it is important to identify similarities and differences in the experimental designs to better understand the parameters that might facilitate audiovisual intermodulation frequency measurement to be achieved, along with further understanding of their functional meaning in multisensory integration. In Regan et al. (1995) audiovisual integration was identified while passively viewing and listening to stimuli presented at 10.304 and 5.054 Hz, respectively. It is noteworthy that participants were not engaged in any particular task whereas in our study participants were instructed to track the visual stimuli. Also, in Regan et al. (1995) the stimuli were stationary and the auditory stimuli were delivered only at the left ear, whereas in our study we presented horizontally moving visual and auditory stimuli and, further, the sound was presented binaurally. Therefore, it may be the case that AV integration is more prominent when participants are not tracking one specific modality from the two presented modalities, i.e. visual or auditory. Furthermore, one could also argue that lower frequencies may be more effective in generating AV integration detectable in intermodulation frequencies, since Regan et al. (1995) used lower tagging frequencies (10.304 and 5.054 Hz vs. 15 and 32 Hz in our study). However, such a simple low-pass filter effect is unlikely, since a recent study found AV integration evidenced by significant response at $f_V - f_A$, using higher frequencies, specifically, f_V of 68 Hz and f_A of 61 Hz (Drijvers et al. 2021). Finally, the main focus of the study of Regan and colleagues (Regan et al. 1995) was to identify the spatial source in the brain supporting AV integration, the processing of f_A and f_V fundamentals, harmonics and intermodulation frequencies. There was no comparison of the different frequencies across conditions as tested in our study.

In turn, Drijvers et al. (2021) paradigm consisted in an active task where participants were presented with action verbs and gestures and had to identify which verb they heard in a forced-choice task. One important aspect to consider is that speech–gesture integration is an automatic process of extraction and combination of AV information to generate the percept of what the interlocutor intends to communicate (Soto-Faraco et al. 2004b). This process is so strong that mismatching facial and auditory information generates audiovisual illusions, such as the McGurk effect (McGurk and MacDonald 1976, see Marques et al. 2016 for a review). It is also noteworthy that in Drijvers et al. (2021) the auditory stimuli were modulated in amplitude, whereas we modulated the pitch, since sound amplitude modulation was used to generate the panning for the perception of sound movement. Finally, another relevant difference is that Drijvers and colleagues (2021) presented the stimuli with noise, a strategy that potentiates integration of both modalities to allow mutual disambiguation of speech and visual stimuli, whereas our stimuli were presented without noise. Therefore, these differential methodological approaches could have accounted for our lack of intermodulation findings.

Importantly, since no difference between congruent and incongruent auditory stimuli was found on behavioural synchronisation performance in the current study, it remains possible that different movement synchronisation tasks could have promoted audiovisual integration and significant activity at the intermodulation frequencies. The complexity of the stimulus trajectories used, and the task difficulty more generally, could have been insufficient to promote the integration and intermodulation frequencies. Considering differences with previous studies, another aspect to account for in future research is that our stimuli were neither ambiguous nor presented with noise. Adding such conditions could possibly foster integration, as the degradation of one modality has been shown to result in greater use of the other modalities (e.g. Drijvers et al. 2021; Parise and Ernst 2017; Stacey et al. 2020; Van der Zwan et al. 2009). Similar experiments could also compare shorter and longer trials, as SSEPs have been shown to be more pronounced for short trials (see Gander et al. 2010). More direct attention manipulations towards visual or auditory stimuli could also be tested. Importantly, the best parameters to promote integration across sensory modalities, and potentially the occurrence of intermodulation frequencies, remain unknown. There are numerous experimental manipulations that will have to be explored before confirming or excluding the relevance of intermodulation frequencies for measuring and understanding bimodal integration.

Moreover, we also found no significant correlation between the EEG power in the main or in the intermodulation frequencies and the behavioural responses in none of the conditions of the two experiments. This further confirms the absence of

link between synchronisation performance and magnitude of EEG activity at the fundamental and intermodulation frequencies. This further supports the potential influence of focused attention on the visual stimuli, which was encouraged by the task instructions, in limiting or preventing the use of auditory information (De Jong et al. 2010; Kim et al. 2007). Conditions and instructions encouraging further focus on auditory information could have been more conducive for the occurrence of significant effects. Furthermore, in our experiment, none of the sensory modalities was degraded so participants could more easily rely solely on the visual information to perform the task (von Saldern and Noppeney 2013) while ignoring the non-target auditory modality. When visual and auditory information are discrepant during spatial localisation, and the visual stimulus is reliable (i.e. not degraded), the visual modality dominates via the “visual capture” of auditory information, and thus there is a distortion of the perceived auditory location (Neuhoff 2004). This is demonstrated in the classic ventriloquism illusion, where sounds seem to be displaced in the direction of a visual stimulus (Alais and Burr 2004). Moreover, both spatial and temporal factors play a significant role in this crossmodal effect, with auditory motion perception being strongly modulated by apparent visual motion (Soto-Faraco et al. 2004a, b). Such interpretation is also in line with Riels and colleagues’ (Riels et al. 2020) findings, where the only AV intermodal effect was observed during auditory target trials resulting in weaker visual SSEPs compared to auditory non-target trials, which argues in favour of selective attention models.

In this sense, the auditory stimuli in our experiment might have played as a distractor that participants attempted to ignore, as even the unimodal spatial detection of sound due to loudness modulation is not necessarily straightforward (Neuhoff 2004). However, we cannot be categorical in such interpretation, as we compared different conditions of AV input signals, but did not look into the processing of each unimodal stimulus separately. Examining the competing modalities combined and in isolation would be necessary to draw further conclusions (Gordon et al. 2019). Nevertheless, we are still scratching the surface when it comes to understanding the extent of application of EEG frequency tagging to detect multimodal integration and crossmodal interactions. The absence of EEG activity at the intermodulation frequencies in all participants and conditions in the current experiment remains rather surprising, and therefore encourages further investigations in multimodal settings in future research.

Conclusion

To conclude, the current study provides no evidence of selective unimodal processing and bimodal integration of visual and auditory information using EEG frequency

tagging in participants manually tracking a moving object. The absence of effects of spatial and temporal congruency manipulations at both behavioural and neural levels in the two experiments suggest that experimental conditions, such as task instructions and difficulty, as well as sensory signal degradations, might need to be further manipulated in future research to determine conditions in which bimodal information is beneficial for the motor tracking of moving objects in the environment. These null results also encourage further investigations in future studies of the relevance of EEG activity at the intermodulation frequencies while using frequency tagging for indexing and understanding multisensory integration processes (also see Adam et al. 2021).

Acknowledgements We thank Ashleigh Clibborn and Ayah Hammoud for their assistance with data collection. This work was supported by a grant from the Australian Research Council (DP170104322, DP220103047). OML is supported by the Portuguese Foundation for Science and Technology and the Portuguese Ministry of Science, Technology and Higher Education, through the national funds, within the scope of the Transitory Disposition of the Decree No. 57/2016, of 29 August, amended by Law No. 57/2017 of 19 July (Ref.: SFRH/BPD/72710/2010)

Author contributions OML—study conception; experiment design; data acquisition, analysis and interpretation; prepared all figures; drafted and revised the manuscript. PEK—study conception; data interpretation; manuscript revision. SN—study conception; data interpretation; manuscript revision. MV—study conception; experiment design; data analysis and interpretation; manuscript revision.

Funding Open access funding provided by FCT/IFCCN (b-on).

Data availability All data are held in a public repository, available at OSF database (URL access: https://osf.io/2jr48/?view_only=17e3f6f57651418c980832e00d818072).

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adam KCS, Chang L, Rangan N, Serences JT (2021) Steady-state visually evoked potentials and feature-based attention: pre-registered null results and a focused review of methodological considerations. *J Cogn Neurosci* 33(4):695–724. https://doi.org/10.1162/jocn_a_01665
- Aissani C, Cottureau B, Dumas G, Paradis AL, Lorenceau J (2011) Magnetoencephalographic signatures of visual form and motion binding. *Brain Res* 1408:27–40. <https://doi.org/10.1016/j.brainres.2011.05.051>
- Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14:257–262. <https://doi.org/10.1016/j.cub.2004.01.029>
- Alink A, Euler F, Kriegeskorte N, Singer W, Kohler A (2012) Auditory motion direction encoding in auditory cortex and high-level visual cortex. *Hum Brain Mapp* 33:969–978. <https://doi.org/10.1002/hbm.21263>
- De Jong R, Toffanin P, Harbers M (2010) Dynamic crossmodal links revealed by steady-state responses in auditory–visual divided attention. *Int J Psychophysiol* 75:3–15. <https://doi.org/10.1016/j.ijpsycho.2009.09.013>
- Drijvers L, Jensen O, Spaak E (2021) Rapid invisible frequency tagging reveals nonlinear integration of auditory and visual information. *Hum Brain Mapp* 42:1138–1152. <https://doi.org/10.1002/hbm.25282>
- Frassinetti F, Bolognini N, Ladavas E (2002) Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res* 147:332–343. <https://doi.org/10.1007/s00221-002-1262-y>
- Frigo M, Johnson SG (1998) FFTW: an adaptive software architecture for the FFT. In: Proceedings of the 1998 IEEE international conference on acoustics, speech and signal processing, vol 3, pp 1381–1384. <https://doi.org/10.1109/ICASSP.1998.681704>
- Giani AS et al (2012) Steady-state responses in MEG demonstrate information integration within but not across the auditory and visual senses. *Neuroimage* 60:1478–1489. <https://doi.org/10.1016/j.neuroimage.2012.01.114>
- Gordon N, Hohwy J, Davidson MJ, van Boxtel JJA, Tsuchiya N (2019) From intermodulation components to visual perception and cognition—a review. *Neuroimage* 199:480–494. <https://doi.org/10.1016/j.neuroimage.2019.06.008>
- Hove MJ, Fairhurst MT, Kotz SA, Keller PE (2013) Synchronizing with auditory and visual rhythms: an fMRI assessment of modality differences and modality appropriateness. *Neuroimage* 67:313–321. <https://doi.org/10.1016/j.neuroimage.2012.11.032>
- Huddleston WE, Lewis JW, Phinney RE, Deyoe EA (2008) Auditory and visual attention-based apparent motion share functional parallels. *Percept Psychophys* 70(7):1207–1216. <https://doi.org/10.3758/PP.70.7.1207>
- Jacques C, Retter TL, Rossion B (2016) A single glance at natural face images generate larger and qualitatively different category-selective spatio-temporal signatures than other ecologically-relevant categories in the human brain. *Neuroimage* 137:21–33. <https://doi.org/10.1016/j.neuroimage.2016.04.045>
- Kim YJ, Grabowecy M, Paller KA, Muthu K, Suzuki S (2007) Attention induces synchronization-based response gain in steady-state visual evoked potentials. *Nat Neurosci* 10:117–125. <https://doi.org/10.1038/nn1821>
- Kim YJ, Tsai JJ, Ojemann J, Verghese P (2017) Attention to multiple objects facilitates their integration in prefrontal and parietal cortex. *J Neurosci* 37:4942–4953. <https://doi.org/10.1523/JNEUROSCI.2370-16.2017>
- Lenc T, Keller PE, Varlet M, Nozaradan S (2018) Neural tracking of the musical beat is enhanced by low-frequency sounds. *PNAS* 115(32):8221–8226. <https://doi.org/10.1073/pnas.1801421115>
- Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neurosci* 28:264–271. <https://doi.org/10.1016/j.tins.2005.03.008>
- Marques LM, Lapenta OM, Costa TL, Boggio PS (2016) Multisensory integration processes underlying speech

- perception as revealed by the McGurk illusion. *Lang Cogn Neurosci* 31(9):1115–1129. <https://doi.org/10.1080/23273798.2016.1190023>
- McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746–748. <https://doi.org/10.1038/264746a0>
- Meyer GF, Wuerger SM, Rohrbein F, Zetzsche C (2005) Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Exp Brain Res* 166:538–547. <https://doi.org/10.1007/s00221-005-2394-7>
- Mouraux A, Iannetti GD (2008) Across-trial averaging of event-related EEG responses and beyond. *Magn Reson Imaging* 26:1041–1054. <https://doi.org/10.1016/j.mri.2008.01.011>
- Neuhoff JG (2004) Interacting perceptual dimensions in Ecological psychoacoustics. In: Neuhoff JG (ed) New York. Academic Press, USA, pp 249–279
- Norcia AM, Appelbaum LG, Ales JM, Cottreau BR, Rossion B (2015) The steady-state visual evoked potential in vision research: a review. *J vis* 15:4. <https://doi.org/10.1167/15.6.4>
- Noy D et al (2017) Audiovisual integration increases the intentional step synchronization of side-by-side walkers. *Hum Mov Sci* 56:71–87. <https://doi.org/10.1016/j.humov.2017.10.007>
- Nozaradan S, Peretz I, Missal M, Mouraux A (2011) Tagging the neuronal entrainment to beat and meter. *J Neurosci* 31:10234–10240. <https://doi.org/10.1523/JNEUROSCI.0411-11.2011>
- Nozaradan S, Peretz I, Mouraux A (2012) Steady-state evoked potentials as an index of multisensory temporal binding. *Neuroimage* 60(1):21–28. <https://doi.org/10.1016/j.neuroimage.2011.11.065>
- Nozaradan S, Zerouali Y, Peretz I, Mouraux A (2015) Capturing with EEG the neural entrainment and coupling underlying sensorimotor synchronization to the beat. *Cereb Cortex* 25:736–747. <https://doi.org/10.1093/cercor/bht261>
- Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Comput Intell Neurosci*. <https://doi.org/10.1155/2011/156869>
- Parise CV, Ernst MO (2017) Noise, multisensory integration, and previous response in perceptual disambiguation. *PLOS Comput Biol* 13(7):e1005546. <https://doi.org/10.1371/journal.pcbi.1005546>
- Quek G, Nemrodov D, Rossion B, Liu-Shuang J (2018) Selective attention to faces in a rapid visual stream: hemispheric differences in enhancement and suppression of category-selective neural activity. *J Cogn Neurosci* 30:393–410. https://doi.org/10.1162/jocn_x_01254
- Regan MP, He P, Regan D (1995) An audio–visual convergence area in the human brain. *Exp Brain Res* 106:485–487. <https://doi.org/10.1007/BF00231071>
- Renton AI, Painter DR, Mattingley JB (2018) Differential deployment of visual attention during interactive approach and avoidance behavior. *Cereb Cortex* 29:2366–2383. <https://doi.org/10.1093/cercor/bhy105>
- Riels KM, Rocha HA, Keil A (2020) No intermodal interference effects of threatening information during concurrent audiovisual stimulation. *Neuropsychologia* 136:107283. <https://doi.org/10.1016/j.neuropsychologia.2019.107283>
- Rosati G, Oscari F, Spagnol S, Avanzini F, Masiero S (2012) Effect of task-related continuous auditory feedback during learning of tracking motion exercises. *J NeuroEng Rehabil* 9:79. <https://doi.org/10.1186/1743-0003-9-79>
- Soto-Faraco S, Lyons J, Gazzaniga M, Spence C, Kingstone A (2004a) The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cogn Brain Res* 14:139–146. [https://doi.org/10.1016/s0926-6410\(02\)00068-x](https://doi.org/10.1016/s0926-6410(02)00068-x)
- Soto-Faraco S, Navarra J, Alsius A (2004b) Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition* 92(3):B13–B23. <https://doi.org/10.1016/j.cognition.2003.10.005>
- Stacey JE, Howard CJ, Mitra S et al (2020) Audio-visual integration in noise: Influence of auditory and visual stimulus degradation on eye movements and perception of the McGurk effect. *Atten Percept Psychophys* 82:3544–3557. <https://doi.org/10.3758/s13414-020-02042-x>
- van der Zwan R, Machatch C, Kozłowski D, Troje NF, Blanke O, Brooks A (2009) Gender bending: auditory cues affect visual judgements of gender in biological motion displays. *Exp Brain Res* 198(2–3):373–382. <https://doi.org/10.1007/s00221-009-1800-y>
- Varlet M, Marin L, Issartel J, Schmidt RC, Bardy BG (2012) Continuity of visual and auditory rhythms influences sensorimotor coordination. *PLoS ONE* 7(9):e44082. <https://doi.org/10.1371/journal.pone.0044082>
- Varlet M, Nozaradan S, Nijhuis P, Keller PE (2020) Neural tracking and integration of ‘self’ and ‘other’ in improvised interpersonal coordination. *Neuroimage* 206:116303. <https://doi.org/10.1016/j.neuroimage.2019.116303>
- von Saldern S, Noppeney U (2013) Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. *J Neurosci* 33(20):8841–8849. <https://doi.org/10.1523/JNEUROSCI.3020-12.2013>
- Zelic G, Varlet M, Wishart J, Kim J, Davis C (2018) The dual influence of pacer continuity and pacer pattern for visuomotor synchronisation. *Neurosci Lett* 683:150–159. <https://doi.org/10.1016/j.neulet.2018.07.044>
- Zelic G et al (2019) The influence of pacer-movement continuity and pattern matching on auditory-motor synchronisation. *Exp Brain Res* 237:2705–2713. <https://doi.org/10.1007/s00221-019-05625-9>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.