

JYX



This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Proverbio, Alice Mado; Tacchini, Marta; Jiang, Kaijun

Title: What do you have in mind? : ERP markers of visual and auditory imagery

Year: 2023

Version: Published version

Copyright: © 2023 The Authors. Published by Elsevier Inc.

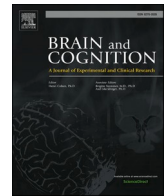
Rights: CC BY-NC-ND 4.0

Rights url: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the original version:

Proverbio, A. M., Tacchini, M., & Jiang, K. (2023). What do you have in mind? : ERP markers of visual and auditory imagery. *Brain and Cognition*, 166, Article 105954.

<https://doi.org/10.1016/j.bandc.2023.105954>



What do you have in mind? ERP markers of visual and auditory imagery

Alice Mado Proverbio^{a,*}, Marta Tacchini^a, Kaijun Jiang^{a,b}

^a Cognitive Electrophysiology lab, Dept. of Psychology, University of Milano-Bicocca, Italy

^b Department of Psychology, University of Jyväskylä, Finland

ARTICLE INFO

Keywords:

EEG/ERPs
BCI
Imagery
Mind reading

ABSTRACT

This study aimed to investigate the psychophysiological markers of imagery processes through EEG/ERP recordings. Visual and auditory stimuli representing 10 different semantic categories were shown to 30 healthy participants. After a given interval and prompted by a light signal, participants were asked to activate a mental image corresponding to the semantic category for recording synchronized electrical potentials. Unprecedented electrophysiological markers of imagination were recorded in the absence of sensory stimulation. The following peaks were identified at specific scalp sites and latencies, during imagination of infants (centroparietal positivity, CPP, and late CPP), human faces (anterior negativity, AN), animals (anterior positivity, AP), music (P300-like), speech (N400-like), affective vocalizations (P2-like) and sensory (visual vs auditory) modality (PN300). Overall, perception and imagery conditions shared some common electro/cortical markers, but during imagery the category-dependent modulation of ERPs was long latency and more anterior, with respect to the perceptual condition. These ERP markers might be precious tools for BCI systems (pattern recognition, classification, or A.I. algorithms) applied to patients affected by consciousness disorders (e.g., in a vegetative or comatose state) or locked-in-patients (e.g., spinal or SLA patients).

1. Introduction

This study aimed to investigate the psychophysiological correlates of perceptual and imagery processes through recording event-related brain potentials (ERPs). In fact, specific electrical markers of perceived and imagined sensory items can be helpful for BCI systems to reconstruct mental representations in patients paralyzed or unable to communicate with the outer world. A Brain-Computer Interface (BCI) usually categorizes brain signals through Machine Learning or classification algorithms and translates them into inputs towards a computer (e.g., Ancau et al., 2022; Hinvest et al., 2022; Xu et al., 2020). In a previous psychophysiological study (Proverbio et al., 2022), we identified the ERP markers of perceptual processing of 10 distinct visual and auditory categories. In addition, we used machine-learning systems to automatically classify brain signals as a function of their semantic category of stimulation (Leoni et al., 2020, 2021). Here, ERPs collected during an imagery condition were closely compared with those recorded during the real perception of the same stimuli to identify the neural markers of mental images. It would be very useful to be able to find evidences of a similarity between perception and imagination, either of response pattern or electrode area, or of latency, even if delayed, in order to have

clues as to the categorical specificity of ERP signals of imagination, to be used in BCI paradigms of mind reading. Abundant neuroscientific evidence suggests a considerable overlap in neural processing during perception and imagery (Pearson & Kosslyn, 2015; Pearson, 2019; Albers et al., 2013; Farah, 1988). The neural representation of imagined and perceived stimuli would be similar in the visual, parietal, and frontal cortex (Dijkstra et al., 2017; Dijkstra et al., 2019). Recent multi-voxel pattern classification studies have shown that in the early visual cortex, patterns of brain activity generated during mental imagery are similar to those generated during perception. Sensory visual features (e.g., stimulus orientation or spatial frequency) would be encoded in detail during mental imagery (Naselaris et al., 2015). Kosslyn et al. (1999) originally showed that magnetic pulses delivered to the medial occipital lobe (V1) vs other areas, impaired visual mental imagery and visual perception to a comparable extent. Indeed, according to a recent model of visual mental imagery (Dijkstra et al., 2019), early visual areas would play a crucial role in imagination. This model predicts a substantial overlap in neural processing during perception and imagery, with an implication of the ventral cortical visual stream and the early visual areas, which would provide the sensory and spatial representational content of visual mental imagery. Clear evidence of a V1 activation has

* Corresponding author at: Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, U6, 20126 Milan, Italy.

E-mail address: mado.proverbio@unimib.it (A.M. Proverbio).

<https://doi.org/10.1016/j.bandc.2023.105954>

Received 21 October 2022; Received in revised form 6 January 2023; Accepted 7 January 2023

Available online 17 January 2023

0278-2626/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

been offered, especially if the imagination requires a detailed stimulus processing (Le Bihan et al., 1993; Klein et al., 2000; Kosslyn et al., 1999; 2001). Secondary visual areas, such as bilateral occipital (Zago et al., 2010), the temporal lobes (Bartolomeo, 2002), the middle and inferior temporal gyrus (Moro et al., 2008) were also found active during the imagery of object shape, size and colour, letter shape, and animal parts. The left fusiform gyrus would be involved in visual mental imagery of orthographic material (Thorudottir et al., 2020; Ralph et al., 2017). Finally, fronto-parietal networks seem to play an important role in imagery for words and pictures (Yomogida et al., 2004). Spagna et al. (2021) propose that a core imagery network could be a small region in the left fusiform gyrus (FG4 area), labelled as *Fusiform Imagery Node* (FIN). FIN would receive semantic and conceptual information from more anterior regions of the left temporal lobe (Ralph et al., 2017), while medial temporal lobe, together with the posterior cingulate, and the fronto-parietal attention networks, would contribute to vivid visual mental imagery (Fulford et al., 2018). While brain forward connections would convey information from the outside world, backward connections would have a role in the formation of mental images in the absence of external bottom-up inputs (Ganis & Schendan, 2008). In this regard, Dentico et al. (2014) conducted a high-density electroencephalography study to measure cortical directed connectivity during visual perception and visual mental imagery and found a reversal of neural signal flow in parieto-occipital cortices during visual imagery as compared to perception. Again, Dijkstra et al. (2017) used fMRI and dynamic causal modelling to characterize the effective connectivity during visual perception and imagery and found that during perception there was an increase in both bottom-up and top-down coupling relative to baseline. According to Zanto et al. (2011), (see also Pearson, 2019) top-down connections from the inferior frontal gyrus are important for selective attention during encoding, as well as for the maintenance of visual information during the delay period. Interestingly, recent evidence linked prefrontal cortex volume, but not V1 anatomy, with imagery vividness (Bergmann et al., 2016). In fact, Dijkstra et al. (2017b) showed that the neural overlap between imagery and perception in the entire visual system would correlate with experienced imagery vividness. Tackling the problem from a different point of view, Fulford et al. (2018) found that low-vividness imaginers activated a more widespread set of brain regions (especially in regions of the fusiform gyrus, posterior cingulate, and parahippocampal gyri) while visualising, than the high-vividness group. In summary, it seems that visual imagery would involve sensory areas normally involved in the processing of the specific features involved, but also more anterior fronto-temporal regions controlling and guiding attentional and memory processes.

As for the auditory modality, the evidences lead to a similar framework. Halpern et al. (2004) found that secondary auditory cortex was significantly active during timbre imagery and perception. Primary auditory cortex was activated during timbre perception but was not activated during timbre imagery. Other PET studies investigating imagery of music melody found an involvement of the right superior temporal gyrus, of the posterior parietal cortex (Zatorre et al., 2010) as well as the right frontal lobe and the supplementary motor area (Halpern & Zatorre, 1999). Other studies have investigated the ability to imagine music key (minor, vs major: Meyer et al., 2007), melodies (Halpern, 1988a), and tempo (Halpern, 1988b). They found an involvement of the superior temporal cortex, cingulate cortex and prefrontal cortex. These areas are known to be quite active during music perception (e.g., Proverbio et al., 2020). Finally, imagery for speech would involve the left inferior frontal gyrus (Broca area) at the basis of speech production (Bookheimer, 2002; Shergill et al., 2001).

Because of this obvious commonality between the neural bases of perception and imagination, it is entirely conceivable that the electrical activity generated in the two cases would result, at least in part, in a similar morphology of ERP components on the scalp. Unfortunately, not many EEG/ERP studies have directly compared perception and imagination of simple stimuli by recording brain potentials, and not many ERP

evidences are available. Most of the electrophysiological studies on imagery concern motor imagery (e.g., Lu and Yin, 2015; Wang et al., 2021; Xu et al., 2020), or music imagery (Marion et al., 2021). Considering auditory imagery, Janata (2001) found that imagined notes (pitch) evoked similar evoked responses to the perceived notes. In another EEG study (Wu et al., 2011) it was found that the amplitude of the imagery-related late positive complex (LPC) increased with loudness of the imagined sound, similarly to the perceptual response. More recently, Zhou et al. (2019; 2020) compared readers' ERPs when imagining native and non-native speech and found that imagining a native speaker's voice when reading silently elicited similar ERP effects (N400 and P600 responses to linguistic violations) as normal silent reading, except that with smaller amplitudes. For the visual modality, Page et al. (2011) recorded evoked potentials during perception and imagery of colour patterns vs B/W checkerboards and correlated the morphology of evoked signals, finding a similarity between imagery and perception, as a function of stimulus type. However, they did not quantify amplitude or latencies of ERP components, but only considered the shape of waveforms. Again, Llorella et al. (2020) applied classification systems to evoked potentials recorded during imagination of a tree, a dog, an airplane and a house but reported partially satisfying results (60 % success rate in the classification). Finally, Lanfranco et al. (2021) applied multivariate decoding analysis to compare N170 ERP responses to hypnotically hallucinated, imagined and perceived faces, and found that classifier was able to distinguish between the hallucinated and imagined condition over the right occipito/temporal area.

Overall, the available knowledge about specific ERP indicants of class-specific imagery is quite poor, and not much related to specific ERP components. In this pilot study of exploratory nature, we expected that imagery-related ERP components, after a delay due to the timing of voluntarily generating mental images (estimated in about 300–400 ms), would share some of the morphologic properties with sensory-evoked components. The latter would be more anteriorly distributed (Lee et al., 2012; Dentico et al., 2014), being the posterior brain not actually stimulated by external stimuli (Dijkstra et al., 2018, 2019; Marion et al., 2021; Xie et al., 2020). Studies on timing oscillation in human visual imagery (Crovitz et al., 1971) have shown that it takes 500 ms to imagine the letters of the alphabet, which is easier than imagining a laugh or a musical fragment. In Kosslyn's series of studies participants took about 1800–2000 ms to compare in size two imagined animals (Kosslyn, 1980), or to generate a mental image on a pattern grid (Lewis et al., 2011). Based on this knowledge, the experimental procedure was devised so that participants were given 2 full seconds for activating an image after the go signal, that were followed by an ISI interval for resetting mental representation before the next trial.

2. Methods

2.1. Participants

Thirty healthy right-handed students (15 males and 15 females) voluntarily took part in this study as unpaid volunteers and earned academic credits for their participation. Experimental sample size was based on a Power Analysis performed using G*Power software (ANOVA repeated measures; $\alpha = 0.05$, critical $F = 3.354$; number of trials = 40). All the subjects were right-handed, as determined by the Italian version of the *Edinburgh Inventory Questionnaire* (Salmasso and Longoni, 1985). The experiments were conducted with the written and informed consent of each participant. All the participants had a normal or correct-to-normal vision with no history of neurological or psychiatric diseases or drug abuse. The experiment was conducted in accordance with international ethical standards (Declaration of Helsinki, 1964), and with the approval of the local Ethical Committee. The project, entitled "Neurobiological bases of mental reconstruction of visual and auditory stimuli" was approved by the Research Assessment Committee of the Department of Psychology (CRIP) for minimal risk projects, under the

aegis of the Ethical committee of University of Milano-Bicocca, on February 3rd, 2020, protocol n: RM-2020-242). The data of ten participants were excluded after averaging and artifact rejection due to the excessive EEG/EOG artifacts. The final sample comprised twenty participants (7 males, 13 females), aging on average 23.9 years ($SD = 3.34$).

2.2. Stimuli

The stimuli were the same used in Proverbio et al. (2022) as well as in the machine learning study by Leoni et al. (2020, 2021). The whole picture set comprised 280 images belonging to 7 categories (infant, adult and animal faces, bodies, checkerboards, words and objects; 40 images per category). The auditory stimuli were 120 auditory files lasting 1.5 sec and belonging to 3 different categories (40 speech items, 40 music items, and 40 affective vocalization items). Stimuli were matched for sensory (e.g., luminance, size), perceptual (i.e., sex and age of faces, sex and age of voices) and linguistic properties (e.g., word frequency, font). Pictures were also matched for shape similarity, colour hue, and position within categories). All the images were in colour on white background and were presented at the centre of the screen. Linguistic auditory stimuli were 40 Italian words, while emotional vocalizations were taken from a previously validated database (used in Proverbio et al., 2020). The intensity of auditory stimuli ranged between 20 and 30 dB; stimuli were normalized and levelled in intensity/volume. Auditory files were played through a set of Sennheiser headphones.

2.3. Procedure

Participants comfortably sat inside an anechoic and faradized cubicle at 114 cm of distance from a HR VGA colour monitor, which was located outside the cabin. Participants were asked to keep their gaze on the fixation point located at the centre of the screen and to avoid any ocular or body movement. Visual stimuli were presented in random order at the centre of the screen in 8 different runs lasting about 3 min each; auditory stimuli were presented in random order in 4 different runs lasting about 2 min and 30 s each. Both sessions comprised a perception and an imagery condition (see Fig. 1). The experimental session was preceded by a training phase.

Participants were followed by a grey screen acting as inter stimulus interval (ISI). The grey screen was meant to cancel possible retinal after-images related to the previous stimulation. A bright yellow frame at the edge of the screen followed, prompting the imagery condition for EEG synchronization. Stimulus duration was 1500 ms, while (ISI) randomly varied between 500 ± 100 ms. The yellow framework lasted for 2000 ms and was followed by an inter-trial interval (ITI) randomly varying from 900 ± 100 ms. Participants were instructed to activate a mental image of the stimulus category just perceived, as vividly and clearly as possible, with their eyes open. The actual instructions were: «You will be presented with images in the center of the screen, or sounds in the headphones, followed by a screen with a yellow frame. Your task will be as follows: When the yellow frame appears, IMAGINE the type of image or sound you have just seen or heard, reconstructing it in your mind as vividly as possible». In order to maintain high levels of attention toward stimulation, subjects were warned that at the end of stimulation they would have to answer questions about the content of the stimuli observed. The questionnaire, administered at the end of the experimental session, comprised 10 questions (1 for each stimulus category), in which participants were asked assess the imagination easiness or difficulty by means of a 5-point Likert scale, where “1” corresponded to very difficult and “5” corresponded to very easy. The question (declined for the various stimuli) was: «Thinking back to the exercise of mentally reconstructing the images you saw during the experiment, what was it like to imagine ...?».

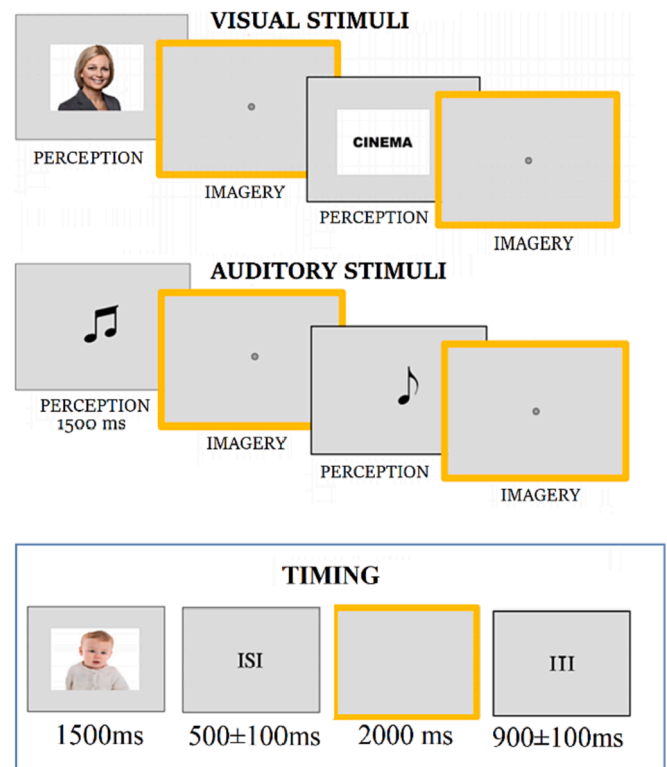


Fig. 1. Time sketch of the experimental procedure. The yellow frame prompted the imagery condition, while fixation was paid to the fixation centre (a little circle). (Top row) Visual stimulus were randomly presented alternating perceptual and imagery conditions. (Middle row) Auditory stimuli were also randomly presented alternating a perceptual and an imagery condition. (Lower row). The timing indicates the temporal sequence of the experimental procedure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.4. EEG recording and analyses

The EEG was continuously recorded from 128 scalp sites at a sampling rate of 512 Hz and according to the 10/5 system (Oostenveld and Praamstra, 2001). Horizontal and vertical eye movements were recorded. Averaged ears served as the reference lead. The EEG and electro-oculogram (EOG) were amplified with a half-amplitude band pass of 0.016–70 Hz. Electrode impedance was maintained below 5 k Ω . The EEG was recorded and analyzed using EEProbe recording software (ANT, Enschede, The Netherlands). Stimuli presentation and triggering was performed using Eevoke Software for audiovisual presentation (ANT, Enschede, The Netherlands).

The EEG epochs were synchronized with the onset of stimuli presentation, i.e., the actual stimulus for the perceptual condition, and the yellow frame for the imagery condition. A computerized artifact rejection criterion was applied before averaging to discard epochs in which eye movements, blinks, excessive muscle potentials or amplifier blocking occurred. The artifact rejection criterion was a peak-to-peak amplitude exceeding 50 μV , and the rejection rate was $\sim 5\%$. ERPs were averaged off-line from -100 ms before to 1500 ms after stimulus onset for the perception condition and from -100 ms to 2000 ms for the imagery condition. ERP components were identified and measured with reference to the average baseline voltage over the interval of -100 to 0 ms at the sites and latencies at which they reached their maximum amplitudes. The electrode clusters and time windows for measuring and quantifying ERP components of interest were based both on previous literature and on the determination of when and where (on scalp surface) they reached their maximum values. The ERP time windows were centered on the peak (positive or negative) of maximum amplitude

when a prominent inflection point in the electrical signals was identifiable, and on wider time intervals for longer monotonic potentials. ERP waves were filtered offline with a bandpass filter of 0.016/30 Hz for illustration purposes.

The mean area amplitude values of the ERP components of interest were subjected to repeated-measure ANOVAs whose factors of variability were stimulus category (depending on the stimuli of interest), electrode (depending on the component of interest) and, where possible, hemisphere (left, right). Tukey (HSD) post-hoc comparisons ($p < 0.01$) were used for contrasting means. The Greenhouse-Geisser correction was also applied to compensate for possible violations of the sphericity assumption associated with factors that had more than two levels. In this case, the degrees of freedom accordingly modified are reported together with the epsilon (ϵ) and the corrected probability level.

The individual scores obtained from the imagery questionnaire were subjected to repeated measures ANOVA. The electrophysiological analyses relative to the imagination condition are detailed below:

Living stimuli (adult faces, animal faces, and infant faces). The mean area amplitude of N2 response was recorded from posterior/temporal (P7, P8) and occipito/temporal sites (PPO9h, PPO10h) in the 280–380 ms temporal window.

Non-living stimuli (words, checkerboards, and objects). The mean area amplitude of 3 late-latency deflections to non-living stimuli during imagination was recorded, respectively: from temporo/parietal sites (P7, P8) between 700 and 900 ms, from occipito-temporal sites (PPO9h, PPO10h) between 900 and 1100 ms and from frontal sites (F3, F4) between 900 and 1100 ms.

Auditory stimuli (emotional vocalization, music, and words). The mean area amplitude of P2-like response was recorded from central and parietal sites (C1, C2, P3, P4) in the 400–600 ms temporal window. The mean area amplitude of the P300-like response was recorded from frontal and fronto-central sites (AF3, AF4, FFC1h, FFC2h) in the 700–900 ms temporal window. The mean area amplitude of the N400-like response was recorded from anterior and fronto-central sites (AF3, AF4, FFC1h, FFC2h) in the 1000–1200 ms temporal window.

Auditory vs Visual stimuli. The mean area amplitude of the PN response was recorded from frontal and central sites (Fz, Cz) in the 200–400 ms temporal window in response to auditory and visual stimuli.

ERP data were also statistically compared to the data collected during the perception condition (see also Proverbio et al., 2022). The ANOVA factors were the same of the imagery analysis with the addition of the between-group factor condition (imagery vs perception). The following ERP components were quantified and analysed:

Living stimuli (adult faces, animal faces, and infant faces). The mean area amplitude of N2 response was recorded from posterior/temporal and occipito/temporal sites (P7, P8, PPO9h, PPO10h) in the 280–380 ms temporal window for both the perception and imagination condition.

The mean area amplitude of the centro/parietal positivity (CPP) response was recorded from centroparietal sites (Cpz, Pz) in the 400–600 ms temporal window, for both the perception and imagination condition. The late CPP (LCPP) response was recorded from the same sites in the 600–900 ms temporal window. The mean amplitude of the anterior negativity response was recorded from anterior frontal areas (AFp3h, AFp4h, AFz, Fpz, Fz) in the 200–400 ms time window for the perception condition, and in the 500–700 ms temporal window for the imagery condition. The mean area amplitude of the anterior positivity was recorded from fronto-central sites (AFz, Fz) in the 600–800 ms time window for the perception condition, and in the 1000–1200 ms temporal window for the imagery condition.

Auditory stimuli (emotional vocalization, music, and words). The mean area amplitude of the P300 response was recorded from fronto-central and central sites (FFC1h, FFC2h, C1, C2) in the 400–500 ms temporal window for the perception condition, and from anterior and fronto-central sites (AF3, AF4, FFC1h, FFC2h) in the 700–900 ms temporal window for the imagery condition. The mean area amplitude of the

N400 response was recorded from parietal and centro/parietal sites (CCP1h, CCP2h, P3, P4) in the 450–650 ms temporal window for the perception condition, and from frontal and fronto-central sites (AF3, AF4, FFC1h, FFC2h) in the 1000–1200 ms temporal window for the imagery condition.

Auditory vs Visual stimuli. The mean area amplitude of the PN300 response was recorded from frontal and central sites (Fz, Cz) in the 200–400 ms temporal window in response to auditory and visual stimuli for both the perception and imagination condition.

3. Results

Behavioural data. The repeated measures ANOVA performed on the individual imageability scores of visual and auditory stimuli showed the significance of sensory modality ($F(1,17) = 6.06, p < 0.05; \epsilon = 1$), indicating that auditory stimuli (4.07, $SD = 0.58$) were rated as easier to imagine than visual stimuli (3.77, $SD = 0.49$). For all stimuli (visual: $t(17) = 6.66, p < 0.001$, auditory: $t(17) = 7.88, p < 0.001; \epsilon = 1$) scores were significantly higher than the scale midpoint, indicating that subjects evaluated the imagination task not particularly difficult. The ANOVA also showed the significance of category factor ($F(9,153) = 3.56, p < 0.01$). Post-hoc comparisons showed that participants considered imagining dressed bodies (3, $SD = 1.08$) significantly more difficult than imagining animal's faces (4, $SD = 0.77$), checkerboards (3.9, $SD = 1.08$) familiar objects (4.06, $SD = 0.72$), words (4.11, $SD = 0.96$), music (4.06, $SD = 0.80$), speech (4.1, $SD = 0.76$), emotional vocalizations (4.06, $SD = 0.87$). This might reflect indeterminacy and increased indecision related to the specific clothing to be imagined.

3.1. Electrophysiological data

Imagery vs perception. Category-related modulations of N170, N2, P2 and P300 responses were not found in the imagination condition. ERPs recorded during imagery showed an overlap of early potentials across classes (reflecting perceptual processing of the bright frame) as can be appreciated in Fig. 2. Since the bright frame, prompting imagery, was identical for all conditions, regardless of the visual or auditory category to be imagined, any difference in the amplitudes of the ERPs should be related to category-dependent imaginative processes, whereas a small and constant part of the sensory signal (e.g., within 50–250 ms) would reflect the VEPs evoked by the probe (visible in Fig. 2).

A repeated measure ANOVA was performed on the mean area amplitude of N2 response recorded from posterior/temporal and occipito/temporal sites (P7, P8, PPO9h, PPO10h) in the 280–380 ms temporal window in the imagery condition. The results showed the significance of Electrode factor, with larger N2 responses over posterior temporal site ($F(1, 29) = 4.9108, p = 0.034; \epsilon = 1$). However, Category

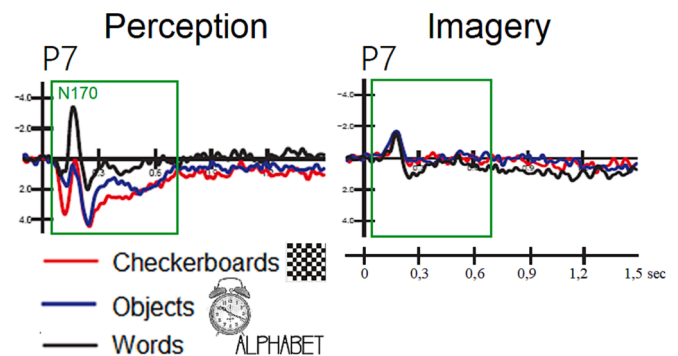


Fig. 2. Grand average ERPs recorded at left occipito/temporal sites during the visual perception and imagination of non-living objects. A lack of ERP modulation can be observed for the N170, N2, P2, and P300 components during imaging, due to the lack of sensory stimulation.

factor was not significant ($F(2, 58) = 1.9, p = 0.145; \epsilon = 0.95$), thus indicating that imagery-related brain responses were not modulated by stimulus category (adult faces, animal faces, infant faces) earlier than 400 ms post-stimulus.

A further ANOVA was performed on N2 responses recorded from occipito/temporal sites (P7, P8, PPO9h, PPO10h) in the 280–380 ms temporal window in the imagery vs perception condition. The effect of category was not significant but only in interaction with condition ($F(2,58) = 5.46, p < 0.006; \epsilon = 0.85, \text{adj. } p < 0.01$). Tukey post-hoc comparisons showed that, while N2 was not modulated at all by stimulus category during imagery, it was strongly modulated during perception. Post-hoc comparisons showed only for perception a significant difference ($p < 0.001$) between infant (4.70 μV , SD = 0.47) and other faces (adult = 3.94 μV , SD = 0.40; animal = 3.58 μV , SD = 0.46).

Later ERP waves (presumably reflecting the activation of mental images) were instead strongly affected by stimulus category, similarly to the perceptual condition, but predominantly at anterior sites (see Fig. 3).

Visual stimuli: Living. The ANOVA performed on the centro/parietal positivity (CPP) (400–600 ms) showed the significance of condition ($F(1,19) = 12.53, p < 0.005; \epsilon = 1$), with smaller amplitudes during the imagery (-1.05 μV , SE = 0.36) than perception condition (0.99 μV , SE = 0.47). A significant effect of category was also found ($F(1,19) = 10.22, p < 0.005; \epsilon = 1$), with much larger CPP responses to infant faces (0.51 μV , SE = 0.35) than human bodies (-0.57 μV , SE = 0.35), in both conditions (imagery and perception) as can be observed in Fig. 4.

The ANOVA performed on the amplitude values of LCPP (600–900 ms) revealed the significance of condition ($F(1,19) = 41.63, p < 0.001; \epsilon = 1$) with larger potentials during the perceptual (1.73 μV , SE = 0.28) than imagery condition (-0.57 μV , SE = 0.32). There was also a significant effect of category ($F(1,19) = 20.99, p < 0.05; \epsilon = 1$) with much larger LCPP responses to infant faces (0.94 μV , SE = 0.30) than body stimuli (0.22 μV , SE = 0.29), regardless of condition.

The ANOVA performed on the Anterior Negativity revealed the significance of condition ($F(1,19) = 21.62, p < 0.001; \epsilon = 1$) with larger responses during the perception (-4.64 μV , SE = 0.77) than imagery

condition (-1.60 μV , SE = 0.33). A significant effect of category was also found ($F(3,57) = 11.37, p < 0.001; \epsilon = 0.75, \text{adj. } p = 0.001$). Post-hoc comparisons showed that AN to human faces (especially infant faces ($p < 0.0001 = -4.05 \mu\text{V}$, SE = 0.67, but also adult faces (-3.73 μV , SE = 0.59) was much larger than to bodies (-1.85 μV , SE = 0.38) and animal faces (-2.84 μV , SE = 0.53).

The ANOVA performed on the amplitude of the anterior positivity revealed a significant effect of condition ($F(1,19) = 4.51, p < 0.05; \epsilon = 1$) with larger amplitudes during the perception (0.28 μV , SE = 0.53) than imagery condition (-1.18 μV , SE = 0.51). The ANOVA also showed the further significance of stimulus category ($F(1,19) = 15.86, p < 0.001; \epsilon = 1$) with much larger anterior positivities to animal (0.18 μV , SE = 0.36) than adult faces (-1.07 μV , SE = 0.47).

Auditory stimuli. ERPs to auditory stimuli are depicted in Fig. 5. The ANOVA performed on fronto/central P300 response revealed a significant effect of category ($F(2,34) = 14.17, p < 0.001; \epsilon = 1$), with larger ($p < 0.0001$) responses to music (2.17 μV , SE = 0.42) compared to emotional vocalizations (0.14 μV , SE = 0.47) and words (0.31 μV , SE = 0.45). This finding can also be appreciated by looking at topographical maps of Fig. 6 (left). The interaction between condition (imagery, perception) and category proved significant ($F(2,34) = 7.72, p < 0.005; \epsilon = 1$). Post-hoc comparisons showed highly significant differences ($p < 0.0001$) between P300 to music and voices in both conditions (Perception: music = 3.37 μV , SE = 0.83; emotional vocalization = 0.18 μV , SE = 0.84; words = 0.58 μV , SE = 0.77. Imagery: music = 0.97 μV , SE = 0.38; vocalizations = 0.09 μV , SE = 0.33; words = 0.03 μV , SE = 0.51).

The ANOVA performed on the amplitude values of centro/parietal N400 response revealed a significant effect of condition (imagery, perception) ($F(1,17) = 22.84, p < 0.001; \epsilon = 1$), with larger negativities during the perception (-0.54 μV , SE = 0.28) than imagery condition (1.18 μV , SE = 0.31). A significant effect of category, independent on condition, was also found ($F(2,34) = 11.13, p < 0.0001; \epsilon = 0.91, \text{adj. } p = 0.0003$). Relative post-hoc comparisons indicated that N400 to speech (-0.40 μV , SE = 0.34) and emotional vocalizations (0.21 μV , SE = 0.30) was significantly larger than that to music (1.13 μV , SE = 0.25), as

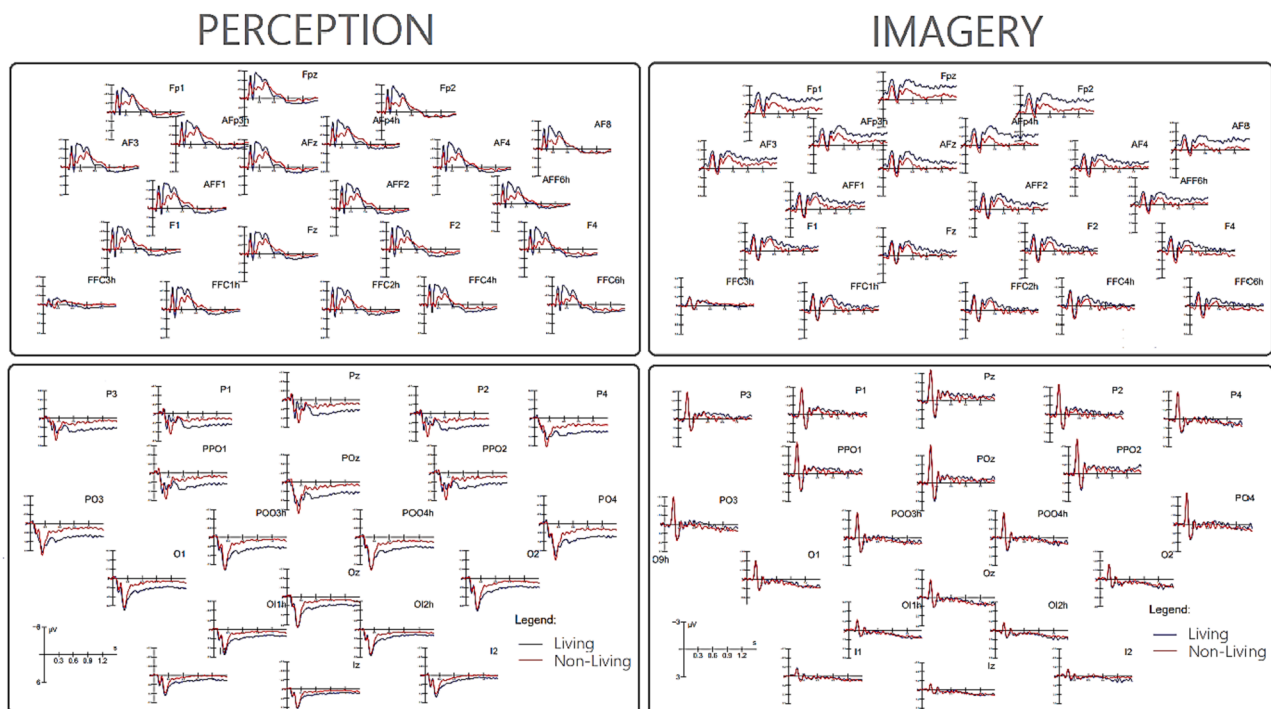


Fig. 3. Grand average ERPs recorded at anterior (top) and posterior (bottom) sites in the perceptual (left) and imaginative (right) conditions, during processing of living and non-living visual stimuli. It can be noted that, during imagination, the category-dependent modulation of ERPs was long latency and more anterior, with respect to the perceptual condition.

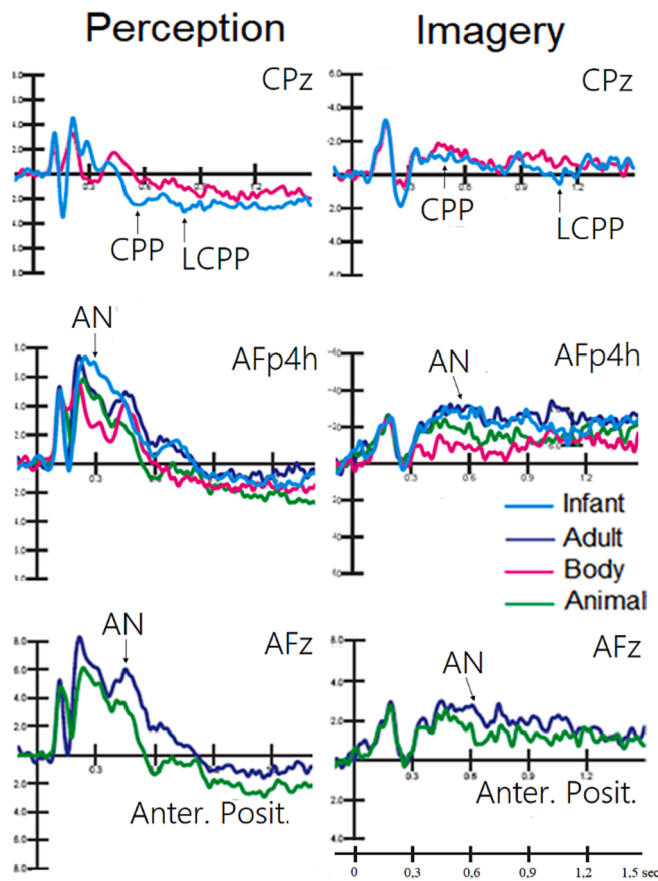


Fig. 4. Visual ERPs. Grand average ERP waveforms showing late-latency category-related modulations for visual stimuli (living). The left column shows ERPs related to stimulus perception, while the right column shows ERPs related to stimulus imagination. There were several similarities in the morphology and amplitude modulation of ERPs, with some time lag for imagination.

shown in waveforms of Fig. 5. Moreover, the effect of hemisphere proved significant ($F(1,17) = 5.84, p < 0.05; \epsilon = 1$), with larger N400s over left ($0.23 \mu\text{V}, SE = 0.22$) than right hemisphere ($0.41 \mu\text{V}, SE = 0.25$).

Auditory vs Visual stimuli. The ANOVA performed on the amplitude values of PN300 deflection showed a significant effect of category ($F(1,19) = 65.37, p < 0.001; \epsilon = 1$). A greater negativity was recorded in response to visual ($-1.51 \mu\text{V}, SE = 0.38$) than auditory stimuli ($2.87 \mu\text{V}, SE = 0.37$). The interaction between condition (imagery, perception) and stimuli was significant ($F(1,19) = 21.89, p < 0.001; \epsilon = 1$), but post-hoc comparisons showed strong differences across modalities ($p < 0.0001$) in both conditions. The visual perceptual condition elicited slightly stronger signals than the auditory conditions, with respect to imagery (Perception: visual = $-2.80 \mu\text{V}, SE = 0.61$; auditory = $3.97 \mu\text{V}, SE = 0.61$. Imagery: Visual = $-0.23 \mu\text{V}, SE = 0.30$; auditory = $1.77 \mu\text{V}, SE = 0.38$. Perception - Imagery: Visual = 2.57, Auditory = 2.2). Topographical maps of Fig. 6 (right) show the similar scalp distribution, across perceptual and imagery conditions, of PN300 as a function of sensory modality.

ERP responses specific to Imagery: The specific ERP responses recorded during imagination are visible in Fig. 5 (right column).

Visual stimuli: Non-living. The ANOVA performed on the amplitude values of the late latency responses recorded over the temporo/parietal sites (P7, P8) in the 700–900 ms time window, over the occipito/temporal sites (PPO9h, PPO10h) in the 900–1100 ms time window, and over the frontal sites (F3, F4) in the 900–1100 ms time window, did not show any significant effect of category. ERPs then did not show a semantic

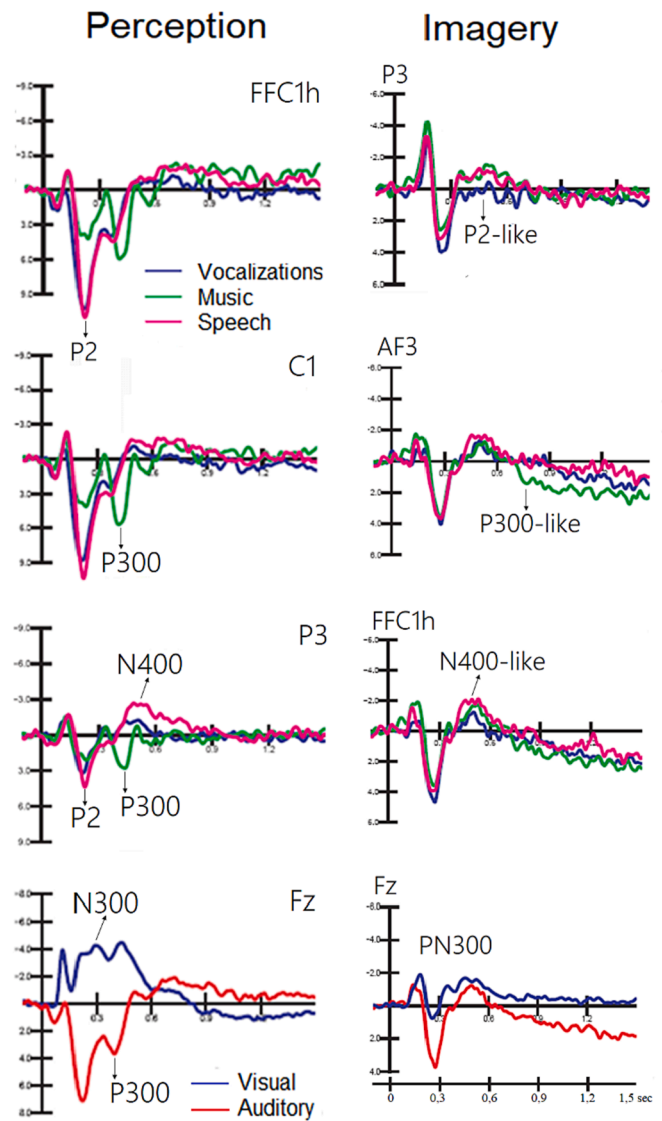


Fig. 5. Auditory ERPs. Grand average ERP waveforms showing late-latency category-related modulations for auditory stimuli. The left column shows ERPs related to stimulus perception, while the right column shows ERPs related to stimulus imagination. There were several similarities in the morphology and amplitude modulation of ERPs, with some time lag for imagination.

specificity for either words, checkerboards or objects in the imagination condition.

Auditory. The ANOVA performed on the P2-like (400–600 ms) amplitude values revealed a significant effect of category ($F(2,38) = 5.72, p < 0.001; \epsilon = 0.955; \text{adj. } p = 0.0076$), with larger responses to emotional vocalizations ($0.07 \mu\text{V}, SE = 0.26$) than music ($-0.92 \mu\text{V}, SE = 0.36$) and speech ($-0.83 \mu\text{V}, SE = 0.37$), as can be appreciate in Fig. 5. A significant effect of hemisphere was found ($F(1,19) = 7.82, p < 0.05$), showing larger responses over the right ($-0.45 \mu\text{V}, SE = 0.29$) than the left hemisphere ($-0.67 \mu\text{V}, SE = 0.27$). The ANOVA performed on the P300-like amplitude values (700–900 ms) revealed a significant interaction of category X electrode ($F(2,38) = 3.37, p < 0.001; \epsilon = 0.95; \text{adj. } p = 0.0047$). P300 to music ($1.05 \mu\text{V}, SE = 0.30$) was significantly more positive than P300 to vocalizations ($0.12 \mu\text{V}, SE = 0.30$) and words ($-0.07 \mu\text{V}, SE = 0.44$), at all sites, and regardless of condition, but especially at anterior sites (AF3, AF4), as shown in Fig. 5. The ANOVA performed on the N400-like (1000–1200 ms) amplitude values revealed a significant effect of category ($F(2,38) = 3.45, p = 0.04; \epsilon = 0.99; \text{adj.}$

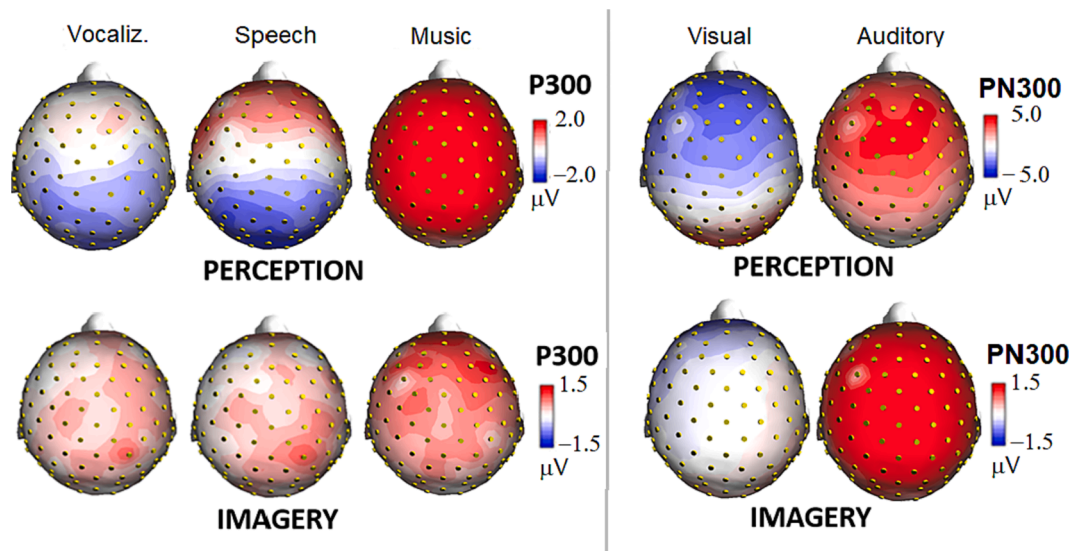


Fig. 6. Isocolour topographical maps of grand average surface voltages recorded in the P300 and PN300 latency ranges for the perception (top) and imagery (bottom) conditions. The P300 to auditory stimuli was modulated by stimulus category, being much larger for musical stimuli, both imagined and perceived (left). The right side of the figure illustrates the effect of sensory modality (visual vs auditory) on PN300 polarity: the effect was quite similar across conditions.

$p = 0.041$), with larger responses to words ($0.80 \mu\text{V}$, $\text{SE} = 0.44$) than music ($1.86 \mu\text{V}$, $\text{SE} = 0.30$), but not compared to vocalizations ($1.24 \mu\text{V}$, $\text{SE} = 0.37$), except that at anterior frontal sites where the difference with vocalization was significant ($p < 0.001$).

4. Discussion

The study aimed to identify reliable electrophysiological markers of the imagination of visual and auditory stimuli belonging to 10 semantic categories. The goal was to provide valuable tools for the development of BCI systems for patients and healthy users. For example, several studies have effectively used EEG data (functional and effective connectivity, wavelet and frequency) for decoding multiple object categories from electroencephalographic signals through multiclass classifiers and machine learning systems (Taghizadeh-Sarabi et al., 2015; Daliri et al., 2013; Behroozi et al., 2016). Overall, the EEG/ERP approach to BCI seems rather promising, not least because it does not suffer from the disadvantages of neuroimaging techniques. The major disadvantages of fMRI-BCI are its high cost, non-portability, complexity of development and usage, the relatively low-temporal resolution and the fact that BOLD is an indirect measure of neuronal signaling. In fact the majority of BCI systems are based on electroencephalogram (EEG) frequency analysis (e.g., for motor imagery) or P300 detection (Leoni et al., 2022). Overall, the data showed that brain signals were always of higher amplitude during perception than imagination, for all stimulus types and sensory modalities. This finding fully agrees with previous neuroimaging evidence describing imagination as a kind of weaker and noisier perceptual experience (Pearson et al., 2015), whereas the latter is more vivid and detailed (Dijkstra et al., 2017). Indeed, a correlation has been found between the vividness of imagery and the similarity of BOLD responses for perception and imagery in early visual cortex (Lee et al., 2012). Overall, ERP deflections related to imagery and sensitive to stimulus category were delayed by approximately 200–400 ms compared to the perception data, probably because they reflected access to mental representation, image reconstruction, and working memory, processes that take longer than purely sensory ones. Indeed, while it is known that visual ERPs are affected by facial information since P1 latency level (e.g., Rossion and Caharel, 2011), here imagery-related ERPs were not even affected by face category (infant, adult, animal) at N2 level (280–380 ms), which usually reflects advanced object/face processing along the ventral stream, because of the lack of incoming sensory

inputs.

Dijkstra et al. (2018) study on the dynamics of visual imagination had also shown a different timing and scalp distribution of brain activity: compared to perception, imagination processes were detected significantly later, as more time is needed to generate internal representations based on purely top-down processes. They measured MEG signals while volunteers perceived or imagined faces vs houses and after having trained classifiers on perceptual data, analyzed imagery data. During imagery, category information could be decoded significantly from 540 ms after cue onset, with a peak at 1073, but a similarity with perception was also found around 300 ms. The author concluded that the generation of a visual representation from a cue takes longer than the activation via bottom-up sensory input. Up to now, not much detail have accumulated in the literature on the ERP components elicited by imagination, not least because of the difficulty of the experimental paradigm, which also makes it problematic to apply classification systems on imagery-related EEG signal for BCI. Therefore, the present data offer an unprecedented set of physiological markers indexing the imagining of biologically relevant visual and auditory stimuli. Table 1 offers a comparative description of these markers, and includes descriptions of the electrode measurement sites, the time windows, and the manifested functional properties.

Imagination of living stimuli (adult and infant faces, animals, and bodies). ERPs to imagined living entities were not modulated by stimulus category earlier than 300–400 ms in posterior brain areas. This differs from perceptual ERPs, which are associated with a larger N170 response to human than animal faces (Farzmañhi et al., 2021), a larger anterior N2 to infant than adult faces (Proverbio et al., 2011), a larger parietal P2 response to bodies than faces, and a larger anterior P300 to animals than faces and bodies (Wu et al., 2006). Similar to the perceptual condition, however, imagining infant faces elicited greater centroparietal positivity (CPP, 400–600 ms) in response to infant faces than bodies, as did later LCPP deflection. This commonality indicates a similarity in the mental representation of conceptual knowledge, whether sense-driven or internally generated (Pearson & Kosslyn, 2015; Albers et al., 2013; Zhou et al., 2019; 2020; Page et al. 2011). Table 1 provides a summary of the ERP markers of imagery identified in this study.

In the same vein, visual perception and imagination of human faces (child and adult ones) were associated with greater anterior negativity than bodies. However, the time window was different: for the perception condition the effect was earlier (200–400 ms) than for the imagination

Table 1

ERP markers of visual and auditory imagery identified during the imagery condition, according to statistical analyses performed across stimulus categories. For each ERP component of interest, the latency in ms, the recording sites/areas on the scalp and their functional properties are detailed.

Summary					
Category	Peak	Latency (in ms)	Scalp Area	Electrodes	Functional Properties
<i>Visual Imagery</i>					
Infant faces	CPP	400-600	centroparietal	Cz, CPz	Larger to infant faces than bodies
Infant faces	LCPP	600-900	centroparietal	Cz, CPz	Larger to infant faces than bodies
Human faces	AN	500-700	anterior frontal	AFp3h, AFp4h, AFz, Fpz, Fz	Larger to human than animal faces, and bodies
Animals	AP	1000-1200	midline frontal	Afz, Fz	Larger to animal than human faces
<i>Auditory Imagery</i>					
Vocalizations	P2-like	400-600	centroparietal	C1, C2, P3, P4	Larger to vocalizations than speech, and music
Music	P300-like	700-900	anterior frontal	AF3, AF4, FFC1h, FFC2h	Larger to music than human voice
Speech	N400-like	1000-1200	anterior frontal	AF3, AF4, FFC1h, FFC2h	Larger to speech than music, and vocalizations
Sensory modality	PN300	200-400	midline fronto/central	Fz, Cz	Negative for visual, positive for auditory stimuli

condition (500–700 ms), as predicted in the literature (e.g., Marion et al. 2021). Finally, animal faces elicited a higher frontal P300 (600–800 ms for the perception condition; 1000–1200 ms for the imagination condition) than adult faces for both the perception and imagination conditions. This finding is in line with other findings in the literature that indicate a greater P300 amplitude in response to animals than to objects (e.g., Proverbio et al., 2007). In another ERP study, it was found that images of animals with instructions to imagine the corresponding sounds (imagery) elicited larger P300 components than images of animals without any imagery instructions (Wu et al., 2006), but there was no category contrast with other stimuli.

Imagination of non-living stimuli (checkerboards, words, objects). ERP signals recorded during imagining words, objects, and checkerboards did not differ statistically in amplitude as a function of category in any of the time windows considered. This might suggest that imagined but not biologically relevant entities are associated with weaker or less distinctive ERP markers, which cannot be detected with the current technique. Previous EEG/ERP studies on imagination have indeed mainly involved images of actions (Su et al., 2020), music (Marion et al., 2021), faces (Kaufmann et al., 2013; Zhang et al., 2012) or speech (Zhou et al., 2019). It is possible that non-living entities generate less vivid mental images, which is related at the neural level to weaker signals and less focused activation (Fulford et al., 2018).

Auditory imagination. The data relative to auditory imagination showed consistent commonalities with perception. This pattern of results fits with previous neuroimaging literature showing a large overlap in neural processing during perception and imagery. Similar to perception (cf. Proverbio et al., 2022), a P2-type response (400–600 ms) was identified whose amplitude was greater for emotional vocalisations than for music or speech, and with a similar right hemispheric lateralisation. Interestingly, in auditory perception P2 is the earliest response, being sensitive to the emotional content of speech stimuli and vocalisations (Paulmann & Kotz, 2008; Paulmann, Seifert, & Kotz, 2010). The data also showed a P300-like anterior response to music (700–900 ms) that can be compared to the P300 fronto/central component (400–500 ms) found in the perception condition. This suggests that musical imagination shares some commonalities with musical perception (Dijkstra et al., 2019). Indeed, Regev et al. (2021) found that specific contents of conscious experience are similarly encoded during imagery and music perception, as reflected in the activity of auditory cortices. These findings are in line with previous ERP literature reporting larger frontal P300 responses in response to pleasant versus unpleasant music (Kayashima et al., 2017). In addition, a frontocentral N400-type response was identified in this study, which was larger for speech than for vocalisations and music, and with a left topographic distribution. This response was very similar to the left lateralised N400 response (450–650 ms) found in the perceptual condition (Proverbio et al., 2022). A large literature on ERPs predicts that the N400 component of ERPs is strongly associated with semantic and word processing (e.g., Kutas and

Federmeier, 2011) and reflects the processing load in word comprehension (Zhang et al., 2021). The asymmetry of the left hemisphere for language processing and word reading in particular is supported by an extensive neuroimaging and electromagnetic literature (for a review, see Richlan 2020), as well as clinical data. Given these similarities, it can be hypothesised that the N400-like frontocentral response recorded in this study, during imagining spoken words, acts as a robust marker of this mental content in BCI systems. It should be noted that in other studies on imagination (Zhou et al., 2019; 2020) the N400 to imagined words behaved similarly to the N400 to perceived words during listening/imagining native and non-native speech.

ERPs also showed unambiguous and clear indices of the sensory modality of the stimulus: visual stimuli as a whole elicited much greater negativity at fronto-central sites (200–400 ms) than auditory stimuli. The same pattern, albeit weaker, was recorded in the imagery condition, with the visual stimuli generating a large negative deflection and the auditory stimuli generating a positive wave. Overall, perception and imagination were associated with partially common electro/cortical markers (Marion et al., 2021), especially for biologically relevant stimuli (Xie et al., 2020; Dijkstra et al., 2019; Pearson, 2019; Regev et al., 2021). During imagination, ERP deflections were generally delayed by about 300 ms relative to perception, which supports very recent EEG/ERP evidence (Marion et al. 2021) that testifies to the cognitive level of conscious representation of different stimulus categories. Category-specific modulation of ERPs was predominantly anterior during imagination, further supporting this hypothesis (Lee et al., 2012; Dentico et al., 2014; Dijkstra et al., 2017; Maier et al., 2020). According to the available literature, neural representations of imagined and perceived stimuli are similar in visual, parietal and frontal cortex and appear to be based on similar top-down connectivity. Since imagination lacks bottom-up information, shared representations between imagination and perception would specifically emerge in the alpha frequency band, which carries feedback information (Dijkstra et al. 2019; Xie et al., 2020). The ERP markers of category-specific processing described above have been identified through expert neuroscience-based supervised analysis, but this information can be extremely useful for establishing spatial or temporal constraints for automatic classification systems (e.g., Azinfar et al., 2013, Yan et al., 2021; Ash and Benson, 2018; Jebari, 2013; Pires et al., 2011; Shan et al., 2018; Su et al., 2020). Indeed, the analysis of ERP and brain signals associated with images are valuable tools for artificial intelligence systems dedicated to reconstructing mental representations related to different categories of visual and auditory stimuli (Jin et al., 2012; Lee et al., 2019; Power et al., 2010).

In conclusion, the aim of this study was to investigate the brain mechanisms underlying the generation of visual and auditory mental imagery with respect to actual perception. Not only was a strong similarity found between the perception and imagery conditions (whose ERPs were distributed more anteriorly and delayed in time by approximately 300–400 ms), but 8 different ERP markers, defined by their

polarity, electrode sites and measurement latency intervals, were provided to possibly identify (in the future) mental content through BCI systems.

5. Study limitations

A possible limitation of this investigation is the presumed inter-individual variability in the capacity to form mental images. In principle, since imagery capacities differ between individuals (Cui et al., 2007; Bergmann et al., 2016), some participants may have needed more time to form the mental image than required by the strictly synchronised EEG paradigm. Furthermore, it was not possible to establish whether and how carefully the participants were always focused on imagination, which is a common problem in this kind of experimental paradigms. Further research should determine the degree of inter-individual variability in the view of BCI application on individual patients.

Another possible limitation of the investigation is that the same set of data was used for analysis and for identifying the ERP peaks, which has been described as “double-dipping” problem in neuroscience (Kriegeskorte, et al., 2009). This problem is commonly observed when a large amount of data is recorded, of which only a small fraction is analyzed in detail and presented in a publication (as in most fMRI studies). However, unlike fMRI data, here the whole time series of electrical recordings from –100 to 1500 ms and for almost all electrode sites are shown, so that the reader can see where ERP signals show a category-related modulation, and where they do not. Of course, a selection of the most representative channels for the various experimental conditions was necessary and this might still be problematic. Since there is not much knowledge about imagery ERP peaks, this issue is of difficult solution. Hopefully, future studies will be able to dip from the present data for identifying the ERP peaks, thus reducing the double dipping problem.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data are available at this link: “Data for: ERP markers of visual and auditory imagery: a ‘mind reading’ approach for BCI systems”, Mendeley Data, V1, doi: 10.17632/h7dm83vgvn.1.

Acknowledgements

The authors are very grateful to Alessandra Brusa and Carola Orsenigo for their help with EEG recordings.

Funding

This project, entitled: “Reading mental representations through EEG signals”, was funded by a grant from University of Milano Bicocca (ATE – Fondo di Ateneo N° 31159-2019-ATE-0064).

Author contributions

AMP conceived and planned the experiment. MT carried out the EEG recordings, performed statistical analyses and contributed to data interpretation. KJ created the stimuli and the experimental setup. AMP interpreted the data and took the lead in writing the manuscript. All authors provided critical feedback and helped shaping the research and writing the manuscript.

Data Accessibility

Anonymized data and details about preprocessing/analyses are available to colleagues through platform at this link: “Data for: ERP markers of visual and auditory imagery: a ‘mind reading’ approach for BCI systems”, Mendeley Data, V1, doi: 10.17632/h7dm83vgvn.1.

References

- Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & De Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, 23(15), 1427–1431.
- Ancau, D. M., Ancau, M., & Ancau, M. (2022). Deep-learning online EEG decoding brain-computer interface using error-related potentials recorded with a consumer-grade headset. *Biomedical Physics & Engineering Express*. <https://doi.org/10.1088/2057-1976/ac4c28>
- Ash, C., & Benson, P. J. (2018). Decoding brain-computer interfaces. *Science*, 360, 615–617.
- Jebari, K. (2013). Brain machine interface and human enhancement: An ethical review. *Neuroethics*, 6, 617–625.
- Azinfar, L., Amiri, S., Rabbi, A., & Fazel-Rezai, R. (2013). A review of P300, SSVEP, and hybrid p300/ssvep brain-computer interface systems. In *Brain-Computer Interface Systems - Recent Progress and Future Prospects*.
- Bartolomeo, P. (2002). The relationship between visual perception and visual mental imagery: A reappraisal of the neuropsychological evidence. *Cortex*, 38(3), 357–378.
- Bergmann, J., Genc, E., Kohler, A., Singer, W., & Pearson, J. (2016). Smaller primary visual cortex is associated with stronger, but less precise mental imagery. *Cerebral Cortex* 26(9), 3838–3850.
- Behroozi, M., Daliri, M. R., & Shekarchi, B. (2016). EEG phase patterns reflect the representation of semantic categories of objects. *Medical & Biological Engineering & Computing*, 54(1), 205–221.
- Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, 25(1), 151–188.
- Cui, X., Jeter, C. B., Yang, D., Montague, P. R., & Eagleman, D. M. (2007). Vividness of mental imagery: individual variability can be measured objectively. *Vision research*, 47(4), 474–478.
- Crovitz, H. F., Rosof, D., & Schiffman, H. (1971). Timing oscillation in human visual imagery. *Psychonomic Science*, 24(2), 87–88.
- Dentico, D., Cheung, B. L., Chang, J.-Y., Guokas, J., Boly, M., Tononi, G., & Van Veen, B. (2014). Reversal of cortical information flow during visual imagery as compared to visual perception. *NeuroImage*, 100(237–243), 37.
- Daliri, M. R., Taghizadeh, M., & Niksirat, K. S. (2013). EEG Signature of Object Categorization from Event-related Potentials. *Journal of Medical Signals and Sensors*, 3(1), 37–44.
- Dijkstra, N., Mostert, P., Lange, F.P., Bosch, S., & van Gerven, M.A. (2018). Differential temporal dynamics during visual imagery and perception. May 29;7:e33904.
- Dijkstra, N., Bosch, S. E., & van Gerven, M. A. (2019). Shared neural mechanisms of visual perception and imagery. *Trends in Cognitive Sciences*, 23(5), 423–434.
- Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. A., & Friston, K. (2017). Distinct top-down and bottom-up brain connectivity during visual perception and imagery. *Scientific Reports*, 7(1), 1–9.
- Dijkstra, N., Bosch, S. E., & van Gerven, M. A. (2017). Vividness of visual imagery depends on the neural overlap with perception in visual areas. *Journal of Neuroscience*, 37(5), 1367–1373.
- Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. Farzmaahdi A, Fallah F, Rajimehr R, Ebrahimpour R. (2021). Task-dependent neural representations of visual object categories. *European Journal of Neuroscience* 54(7), 6445-6462.
- Fulford, J., Milton, F., Salas, D., Smith, A., Simler, A., Winlove, C., & Zeman, A. (2018). The neural correlates of visual imagery vividness—an fmri study and literature review. *Cortex*, 105, 26–40.
- Ganis, G., & Schendan, H. E. (2008). Visual mental imagery and perception produce opposite adaptation effects on early brain potentials. *NeuroImage*, 42(4), 1714–1727.
- Halpern, A. R. (1988a). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 434.
- Halpern, A. R. (1988b). Perceived and imagined tempos of familiar songs. *Music Perception*, 6(2), 193–202.
- Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A pet investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, 9(7), 697–704.
- Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, 42(9), 1281–1292.
- Hinvest, N. S., Ashwin, C., Carter, F., Hook, J., Smith, L. G. E., & Stothart, G. (2022). An empirical evaluation of methodologies used for emotion recognition via EEG signals. *Social Neuroscience*.
- Kaufmann, T., Schulz, S. M., Köblitz, A., Renner, G., Wessig, C., & Kübler, A. (2013). Face stimuli effectively prevent brain-computer interface inefficiency in patients with neurodegenerative disease. *Clinical Neurophysiology*, 124(5), 893–900.
- Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, 2(9), 635–642.

- Kayashima, Y., Yamamuro, K., Makinodan, M., Nakanishi, Y., Wanaka, A., & Kishimoto, T. (2017). Effects of Canon chord progression on brain activity and motivation are dependent on subjective feelings, not the chord progression per se. *Neuropsychiatric Disease and Treatment*, *13*, 1499–1508.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge: Harvard University Press.
- Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.
- Janata, P. (2001). Brain electrical activity evoked by mental formation of auditory expectations and images. *Brain Topography*, *13*(3), 169–193.
- Jin, J., Allison, B. Z., Kaufmann, T., Kübler, A., Zhang, Y., Wang, X., & Cichocki, A. (2012). The changing face of P300 BCIs: A comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement. *PLoS One*, *7*(11), e49688.
- Klein, I., Paradis, A.-L., Poline, J.-B., Kosslyn, S. M., & Le Bihan, D. (2000). Transient activity in the human calcarine cortex during visual-mental imagery: An event-related fMRI study. *Journal of Cognitive Neuroscience*, *12*(Supplement 2), 15–23.
- Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., & Ganis, G. (1999). The role of area 17 in visual imagery: Convergent evidence from PET and rTMS. *Science*, *284*(5411), 167–170.
- Lanfranco, R. C., Rivera-Rei, Á., Huepe, D., Ibáñez, A., & Canales-Johnson, A. (2021). Beyond imagination: Hypnotic visual hallucination induces greater lateralized brain activity than visual mental imagery. *NeuroImage*, *1*(239), Article 118282.
- Le Bihan, D., Turner, R., Zeffiro, T. A., Cuenod, C. A., Jezzard, P., & Bonnerot, V. (1993). Activation of human primary visual cortex during visual recall: A magnetic resonance imaging study. *Proceedings of the National Academy of Sciences*, *90*(24), 11802–11805.
- Lee, S. H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, *59*(4), 4064–4073.
- Lee, M.-H., Williamson, J., Kee, Y.-J., Fazli, S., & Lee, S.-W. (2019). Robust detection of event-related potentials in a user-voluntary short-term imagery task. *PLoS One*, *14*(12), e0226236.
- Leoni, J., Strada, S., Tanelli, M., Jiang, K., Brusa, A., & Proverbio, A. (2021). Automatic stimuli classification from ERP data for augmented communication via Brain-Computer Interfaces. *Expert Systems with Applications* 184(1 December 2021) [10.1016/j.eswa.2021.115572].
- Leoni, J., Tanelli, M., Strada, S. C., Jiang, K., Brusa, A., & Proverbio, A. M. (2020). Automatic stimuli classification from ERP data for augmented communication via brain-computer interfaces. In *2020 IEEE international conference on human-machine systems* (pp. 1–6).
- Leoni, J., Tanelli, M., Strada, S., Brusa, A., & Proverbio, A. M. (2022). Single-Trial Stimuli Classification from Detected P300 for Augmented Brain-Computer Interface: A Deep Learning Approach. *Machine Learning with Applications*. <https://doi.org/10.1016/j.mlwa.2022.100393>
- Lewis, K. J. S., Borst, G., & Kosslyn, S. M. (2011). Integrating visual mental images and visual percepts: New evidence for depictive representations. *Psychological Research Psychologische Forschung*, *75*, 259–271.
- Llorella, F. R., Patow, G., & Azorín, J. M. (2020). Convolutional neural networks and genetic algorithm for visual imagery classification. *Physical and Engineering Sciences in Medicine*, *43*(3), 973–983.
- Lu, N., & Yin, T. (2015). Motor imagery classification via combinatory decomposition of ERP and ERSP using sparse nonnegative matrix factorization. *Journal of Neuroscience Methods*, *15*(249), 41–49.
- Marion, G., Di Liberto, G. M., & Shamma, S. A. (2021). The Music of Silence: Part I: Responses to Musical Imagery Encode Melodic Expectations and Acoustics. *The Journal of Neuroscience*, *1*(35), 7435–7448.
- Maier, M., Frömer, R., Rost, J., Sommer, W., & Rahman, R.A. (2020). Time course and shared neurocognitive mechanisms of mental imagery and visual perception. *bioRxiv* 2020.01.14.905885.
- Meyer, M., Elmer, S., Baumann, S., & Jancke, L. (2007). Short-term plasticity in the auditory system: Differential neural responses to perception and imagery of speech and music. *Restorative Neurology and Neuroscience*, *25*(3–4), 411–431.
- Moro, V., Berlucchi, G., Lerch, J., Tomaiuolo, F., & Aglioti, S. M. (2008). Selective deficit of mental visual imagery with intact primary visual cortex and visual perception. *Cortex*, *44*(2), 109–118.
- Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage*, *105*, 215–228.
- Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*(4), 713–719.
- Page, J. W., Duhamel, P., & Crognale, M. A. (2011). ERP evidence of visualization at early stages of visual processing. *Brain and Cognition*, *75*(2), 141–146.
- Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *Neuroreport*, *19*(2), 209–213.
- Paulmann, S., Seifert, S., & Kotz, S. A. (2010). Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. *Social Neuroscience*, *5*(1), 59–75.
- Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, *20*(10), 624–634.
- Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends in Cognitive Sciences*, *19*(10), 590–602.
- Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences*, *112*(33), 10089–10092.
- Pires, G., Nunes, U., & Castelo-Branco, M. (2011). Statistical spatial filtering for a P300-based BCI: Tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis. *Journal of neuroscience methods*, *195*(2), 270–281.
- Power, S. D., Falk, T. H., & Chau, T. (2010). Classification of prefrontal activity due to mental arithmetic and music imagery using hidden Markov models and frequency domain near-infrared spectroscopy. *Journal of Neural Engineering*, *7*(2), 26002.
- Proverbio, A. M., De Benedetto, F., & Guazzone, M. (2020). Shared neural mechanisms for processing emotions in music and vocalizations. *The European Journal of Neuroscience*, *51*(9), 1987–2007.
- Proverbio, A. M., Riva, F., Zani, A., & Martin, E. (2011). Is it a baby? Perceived age affects brain processing of faces differently in women and men. *Journal of Cognitive Neuroscience*, *23*(11), 3197–3208.
- Proverbio, A. M., Del Zotto, M., & Zani, A. (2007). The emergence of semantic categorization in early visual processing: ERP indices of animal vs. artifact recognition. *BMC Neuroscience*, *8*(1), 1–16.
- Proverbio, A. M., Tacchini, M., & Jiang, K. (2022). Event-related brain potential markers of visual and auditory perception: A useful tool for brain computer interface systems. *Frontiers in Behavioral Neuroscience*, *16*, 1025870. <https://doi.org/10.3389/fnbeh.2022.1025870>
- Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55.
- Regev, M., Halpern, A. R., Owen, A. M., Patel, A. D., & Zatorre, R. J. (2021). Mapping Specific Mental Content during Musical Imagery. *Cereb Cortex*, Mar 22:bh036.
- Richlan, F. (2020). The Functional Neuroanatomy of Developmental Dyslexia Across Languages and Writing Systems. *Frontiers in Psychology* *11*, 155.
- Rosson, B., & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Research* *51*(12), 1297–1311.
- Salmaso, D., & Longoni, A. M. (1985). Problems in the assessment of hand preference. *Cortex*, *21*, 533–549.
- Shan, H., Liu, H., & Stefanov, T. P. (2018). A simple convolutional neural network for accurate P300 detection and character spelling in brain computer interface. In *IJCAI*, 1604–1610.
- Shergill, S. S., Bullmore, E., Brammer, M., Williams, S., Murray, R., & McGuire, P. (2001). A functional study of auditory verbal imagery. *Psychological Medicine*, *31*(2), 241.
- Spagna, A., Hajhate, D., Liu, J., & Bartolomeo, P. (2021). Visual mental imagery engages the left fusiform gyrus, but not the early visual cortex: A meta-analysis of neuroimaging evidence. *Neuroscience & Biobehavioral Reviews*, *122*, 201–217.
- Su, J., Yang, Z., Yan, W., & Sun, W. (2020). Electroencephalogram classification in motor-imagery brain-computer interface applications based on double-constraint nonnegative matrix factorization. *Physiological Measurement*, *41*(7), 075007, 11.
- Taghizadeh-Sarabi, M., Daliri, M. R., & Niksirat, K. S. (2015). Decoding objects of basic categories from electroencephalographic signals using wavelet transform and support vector machines. *Brain Topography*, *28*(1), 33–46.
- Thorudottir, S., Sigurdardottir, H. M., Rice, G. E., Kerry, S. J., Robotham, R. J., Leff, A. P., & Starfrel, R. (2020). The architect who lost the ability to imagine: The cerebral basis of visual imagery. *Brain Sciences*, *10*(2), 59.
- Wang, Y., Luo, J., Guo, Y., Du, Q., Cheng, Q., & Wang, H. (2021). Changes in EEG Brain Connectivity Caused by Short-Term BCI Neurofeedback-Rehabilitation Training: A Case Study. *Frontiers in Human Neuroscience*, *24*(15), Article 627100.
- Wu, J., Mai, X., Chan, C. C., Zheng, Y., & Luo, Y. (2006). Event-related potentials during mental imagery of animal sounds. *Psychophysiology*, *43*(6), 592–597.
- Wu, J., Yu, Z., Mai, X., Wei, J., & Luo, Y. (2011). Pitch and loudness information encoded in auditory imagery as revealed by event-related potentials. *Psychophysiology*, *48*(3), 415–419.
- Xie, S., Kaiser, D., & Cichy, R. M. (2020). Visual Imagery and Perception Share Neural Representations in the Alpha Frequency Band. *Current Biology*, *30*(15), 3062, 3.
- Xu, F. Z., Zheng, W. F., Shan, D. R., Yuan, Q., & Zhou, W. D. (2020). Decoding spectrotemporal representation for motor imagery recognition using ECoG-based brain-computer interfaces. *Journal of Integrative Neuroscience*, *19*(2), 259–272, 30.
- Yomogida, Y., Sugiura, M., Watanabe, J., Akitsuki, Y., Sassa, Y., Sato, T., ... Kawashima, R. (2004). Mental visual synthesis is originated in the fronto-temporal network of the left hemisphere. *Cerebral Cortex*, *14*(12), 1376–1383.
- Yan, W., Liu, X., Shan, B., Zhang, X., & Pu, Y. (2021). Research on the Emotions Based on Brain-Computer Technology: A Bibliometric Analysis and Research Agenda. *Frontiers Psychology* *12*, 771591.
- Zago, S., Corti, S., Bersano, A., Baron, P., Conti, G., Ballabio, E., ... Bresolin, N. (2010). A cortically blind patient with preserved visual imagery. *Cognitive and Behavioral Neurology*, *23*(1), 44–48.
- Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2021). More than words: Word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proceedings of the Biological Sciences*, *288*(1955), 20210500.
- Zhang, Y., Zhao, Q., Jin, J., Wang, X., & Cichocki, A. (2012). A novel BCI based on ERP components sensitive to configural processing of human faces. *Journal of Neural Engineering*, *9*(2), Article 026018.
- Zanto, T. P., Rubens, M. T., Thangavel, A., & Gazzaley, A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience*, *14*(5), 656.

- Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: A role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, 22(4), 775–789.
- Zhou, P., Garnsey, S., & Christianson, K. (2019). Is imagining a voice like listening to it? Evidence from ERPs. *Cognition* 182, 227–241.

- Zhou, P., Garnsey, S. M., & Christianson, K. (2020). ERP data on auditory imagery of native and non-native English speech during silent reading. *Data Brief.*, Feb 12;29: 105242.