NeuroImage 277 (2023) 120226

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Neural entrainment underpins sensorimotor synchronization to dynamic rhythmic stimuli

Mattia Rosso[a,b,*], Bart Moens[a], Marc Leman[a], Lousin Moumdjian[a,c,d]

[a] IPEM Institute for Systematic Musicology, Ghent University, Ghent, Belgium
[b] Université de Lille, ULR 4072 – PSITEC – Psychologie: Interactions, Temps, Emotions, Cognition, Lille, France
[c] REVAL Rehabilitation Research Center, Faculty of Rehabilitation Sciences, Hasselt University, Hasselt, Belgium
[d] UMSC Hasselt, Pelt, Belgium

ARTICLE INFO

Keywords:
Neural entrainment
Source separation
Instantaneous frequency
EEG
Finger-tapping
Perturbation

ABSTRACT

Neural entrainment, defined as unidirectional synchronization of neural oscillations to an external rhythmic stimulus, is a topic of major interest in the field of neuroscience. Despite broad scientific consensus on its existence, on its pivotal role in sensory and motor processes, and on its fundamental definition, empirical research struggles in quantifying it with non-invasive electrophysiology. To this date, broadly adopted state-of-the-art methods still fail to capture the dynamic underlying the phenomenon.

Here, we present event-related frequency adjustment (ERFA) as a methodological framework to induce and to measure neural entrainment in human participants, optimized for multivariate EEG datasets. By applying dynamic phase and tempo perturbations to isochronous auditory metronomes during a finger-tapping task, we analyzed adaptive changes in instantaneous frequency of entrained oscillatory components during error correction. Spatial filter design allowed us to untangle, from the multivariate EEG signal, perceptual and sensorimotor oscillatory components attuned to the stimulation frequency.

Both components dynamically adjusted their frequency in response to perturbations, tracking the stimulus dynamics by slowing down and speeding up the oscillation over time. Source separation revealed that sensorimotor processing enhanced the entrained response, supporting the notion that the active engagement of the motor system plays a critical role in processing rhythmic stimuli. In the case of phase shift, motor engagement was a necessary condition to observe any response, whereas sustained tempo changes induced frequency adjustment even in the perceptual oscillatory component. Although the magnitude of the perturbations was controlled across positive and negative direction, we observed a general bias in the frequency adjustments towards positive changes, which points at the effect of intrinsic dynamics constraining neural entrainment.

We conclude that our findings provide compelling evidence for neural entrainment as mechanism underlying overt sensorimotor synchronization, and highlight that our methodology offers a paradigm and a measure for quantifying its oscillatory dynamics by means of non-invasive electrophysiology, rigorously informed by the fundamental definition of entrainment.

1. Introduction

Humans exhibit a natural inclination to synchronize their movement with rhythmic signals in the environment. The phenomenon can be observed across a range of contexts such as music playing and dance (Leman, 2016; Clayton and Eerola, 2020), sports (Cohen et al., 2010), joint rhythmic tasks (Rosso et al., 2021; Rosso et al., 2022) and verbal communication (Richardson and Shockley, 2008; Shockley et al., 2003). How does the human brain represent an external rhythm, how does it track it over time, and how does it temporally match motor behavior to it? These questions are of exceptional interest in neuroscience nowadays. Evidence converges towards sensorimotor entrainment, namely the alignment of motor and sensory rhythms at the neural level, as a plausible answer.

Neural entrainment is defined as unidirectional synchronization of neural oscillations to an external rhythmic stimulus (Haegens and Zion Golumbic, 2018; Lakatos et al., 2019), and comes with the assumption of endogenous oscillatory activity in the brain which can be driven towards a state of phase- and frequency-locking. From the viewpoint of perception, oscillations reflect changes in the weight of sensory inputs via rhythmic fluctuations in neuronal excitability (Lakatos et al., 2005; Fries, 2005; Buzsáki and Draguhn, 2004). Their

* Corresponding author at: IPEM Institute for Systematic Musicology, Ghent University, Ghent, Belgium.
E-mail address: mattia.rosso@ugent.be (M. Rosso).

https://doi.org/10.1016/j.neuroimage.2023.120226.
Received 25 January 2023; Received in revised form 2 May 2023; Accepted 12 June 2023
Available online 14 June 2023.

entrainment to rhythmic events is thought to subserve the selection of relevant information and to reduce the interference of competing input streams (Lakatos et al., 2019). Perceptual representations are enhanced via low-frequency rhythmic fluctuations of sensory gain (Obleser and Kayser, 2019), resulting in what is considered to be a 'rhythmic mode' of attention (Rosso et al., 2022; Schroeder and Lakatos, 2009; Schroeder et al., 2010; Obleser et al., 2017; Zoefel and VanRullen, 2015). Notably, low-frequency oscillations originating from cortical motor areas (Morillon et al., 2014; Morillon et al., 2015; Morillon and Baillet, 2017) are thought to guide selection of environmental information and active sampling implemented in motoric routines (Schroeder and Lakatos, 2009), which leads to perceptual enhancement of attended stimuli (Schroeder et al., 2010; Morillon et al., 2014; Chemin et al., 2014; Park et al., 2015; Rimmele et al., 2018). Thus, the engagement of the motor system enhances and stabilizes the internal representation of environmental rhythms (Kliger Amrani and Zion Golumbic, 2022) by scaffolding the prediction of incoming sensory events (Rosso et al., 2022; Morillon et al., 2014; Morillon et al., 2015; Morillon and Baillet, 2017; Rimmele et al., 2018; Morillon et al., 2019; Arnal and Giraud, 2012). A network of cortical and subcortical motor areas was recently put forward as responsible for beat-based time-keeping, by dynamically tracking the phase of the stimulus cycle and enabling overt motor alignment (Cannon and Patel, 2021).

The rationale of current state-of-the-art approaches to quantify neural entrainment is that the frequencies of a rhythmic stimulation can be tagged in the power spectrum of electrophysiological timeseries (Nozaradan et al., 2011). In practice, 'frequency tagging' transforms the brain signal into the frequency domain, quantifying the power of stimulus-related frequencies over the whole spectrum. The observation that target frequencies dominate the spectrum has been commonly taken as evidence for the underlying entrainment of neural oscillations (e.g., Nozaradan et al., 2011; Nozaradan et al., 2012; Lenc et al., 2018), which is not exempt from critiques. The most evident of these highlights that periodic stimulation elicits a train of time-locked transient responses in the brain, resulting in prominent peaks in the power spectrum at the related frequencies (Novembre and Iannetti, 2018). Furthermore, allocating attentional resources to a predictable stimulus results in top-down modulation of the evoked response amplitude (Legrain et al., 2011; Breska and Deouell, 2017; Nobre and van Ede, 2018) and, in turn, of power at the related frequency. Periodic evoked responses hinder the measurement of real neural entrainment, represent a critical confound for the empirical investigation with non-invasive electrophysiology, and pose a major methodological challenge for the researchers in the field (Haegens and Zion Golumbic, 2018).

Fundamentally, the entrainment of two oscillatory signals cannot be described by their spectral profile alone, because this is not a test for the underlying oscillatory process (Obleser and Kayser, 2019). Given that the temporal structure of a signal strictly depends on the phase of its oscillatory components (Rajendran and Schnupp, 2019), it is not possible to make inferences on local oscillatory dynamics of the signal if phase information is neglected. Neural entrainment is just one possible process leading to the match of two spectral profiles, and therefore frequency tagging is not a sensitive method to its realization. Entrainment in the strict sense is based on the phase of the signal, and is in principle dissociated from its amplitude. Therefore, the process should be defined in terms of changes in frequency over time (Rosenblum et al., 2001) rather than being inferred by the power spectrum. It is the unidirectional process that leads to a state of phase-locking with a driving force (Lakatos et al., 2019), not the phase-locked state itself (Rosenblum et al., 2001).

Motivated by these limitations, we previously proposed the *Stability Index (SI)* as a measure to quantify neural entrainment from electroencephalography (EEG) recordings of healthy participants engaged in finger-tapping to a steady auditory metronome (Rosso et al., 2021). In our work, we extracted from the EEG signal a component attuned to the stimulation frequency, and computed the *SI* based on the fluctuations of the component's frequency over time. Critically, we reported significant correlations between the *SI* and behavioral measures of synchronization: the more stable the neural component, the more stable and more accurate the synchronization performance. We argued that, in contrast to previous amplitude-based approaches, the measure explicitly captures the dynamic phase adjustment of entrained neural oscillations. The putative entrained component would adaptively speed up and slow down, fluctuating around the target center frequency to reach stable synchronization over time (Rosso et al., 2021). However, one main limitation of our previous work is that we could only provide a global measure of these fluctuations for a given time window, and correlate it to the global behavioral performance. Further inferences on the dynamics of the frequency adjustment over time were not possible, due to the fact that we did not dispose of an experimental paradigm to induce them in a controlled fashion.

With these limitations in mind, we hereby present *event-related frequency adjustment* (*ERFA*) as a novel experimental paradigm for investigating neural and behavioral entrainment with auditory rhythmic stimuli. A way to induce controlled frequency adjustment is to manipulate the frequency of an auditory metronome while experimental subjects are attempting to synchronize their finger-taps to it. The event discloses a post-perturbation window wherein the entrained component is expected to adjust its frequency according to the stimulus dynamics, and correct the error to return to a stable state. In finger-tapping studies, error correction is traditionally investigated with tempo changes and phase shifts, in both positive and negative directions (for a review, see Repp and Su, 2013). We implemented the former as a step change of +/- 10% from the baseline frequency (1.67 Hz), and the latter as a +/- 90° shift of the metronome beat along its cycle. Given that control mechanisms are thought to underpin error correction depending on the nature and the direction of perturbations (Praamstra et al., 2003; Repp, 2001; Repp, 2001; Bavassi et al., 2017; Jantzen et al., 2018), these variables were expected to reveal different underlying oscillatory dynamics.

The ERFA curves, which constitute the neural measure within our experimental paradigm, were computed in three main blocks: 1) *attunement*: from the continuous multivariate EEG signal, a single component maximally attuned to the metronome's frequency was extracted via spatial filtering (Rosso et al., 2021; Cohen and Gulbinaite, 2017); 2) *instantaneous frequency*: changes in frequency over time were computed based on the rate of change of the component's phase (Rosenblum et al., 2001; Cohen, 2014); 3) *event-based segmentation*: instantaneous frequency responses were time-locked to the perturbations, and aggregated by perturbation type and direction. The *attunement* block was carried out separately on two perturbation-free periods, while actively listening to the stimuli or tapping along, which allowed to untangle a perceptual and a sensorimotor entrained component within the same rhythmic task, and to assess their relative contributions in adjusting to perturbations.

Our aim was to use ERFA to track neural entrainment dynamics within the post-perturbation windows, and to model them as a function of time. We hypothesized that the instantaneous frequency response would track the stimulus dynamics across perturbation types and directions, and that the active engagement in the behavioral task would boost the entrainment of the sensorimotor component in the brain signal as compared to the perceptual component. Fig. 1 provides a graphical representation of paradigm, trial structure and expected results. Procedures are explained in detail in the *Materials and methods* section.

2. Materials and methods

2.1. Participants

Twenty ($N = 20$) right-handed healthy participants took part in the experiment (11 females, 9 males; mean age = 32.8 years, std = 6.2 years). All participants had normal hearing and normal or corrected-to-normal vision; none reported any history of major medical, psychiatric or neurological conditions; none reported to be a professional mu-

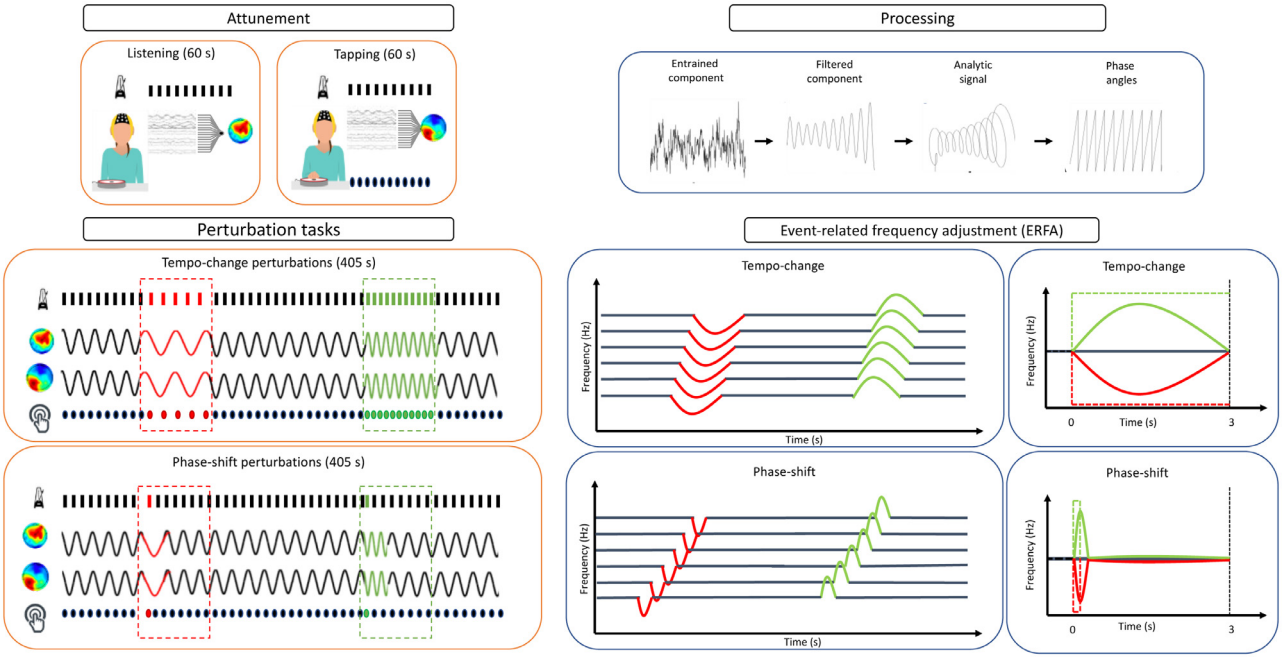

Fig. 1. ERFA paradigm workflow. Orange boxes represent the context of the experimental tasks, blue boxes represent data processing. The approach started with an attunement phase, where participants were exposed to a perturbation-free auditory metronome for 60 seconds. The purpose was to induce a stable oscillation attuned to the stimulation frequency, facilitating its separation from the broadband multivariate EEG signal via generalized eigendecomposition (GED) (Cohen and Gulbinaite, 2017; Cohen, 2022). Participants underwent this period of steady stimulation under distinct conditions of active listening and finger-tapping: using these two datasets, we designed for every participant two different spatial filters to extract a purely perceptual (without movement) and a sensorimotor (with movement) entrained component, respectively. After 60 seconds of finger-tapping, metronomes started to be perturbed by introducing unpredictable tempo changes or phase shifts during 405 seconds, depending on the experimental condition. During these *perturbation tasks*, participants were instructed to synchronize their finger-taps to the metronome, therefore following the stimulus dynamics. The spatial filters designed based on the *attunement* phase were applied to the EEG signal recorded during the tasks, allowing us to track the dynamic changes in the frequency of both perceptual and sensorimotor oscillatory components. Examples of negative (red) and positive (green) perturbation windows are represented for tempo changes and phase shifts, illustrating how the metronomes' onset were manipulated and how neural components and finger-taps were expected to entrain. These figures are presented only as conceptual examples for illustrative purposes, and are not meant to be realistic. In order to provide a measure for the neural and behavioral adjustments, the following *processing* pipeline based on Rosso et al. (2021) was applied. Entrained components were narrow-band filtered around the metronome's center frequency and Hilbert-transformed to produce analytic signals and extract phase timeseries. These were then unwrapped to prevent phase resets, differenced and scaled to Hz to produce instantaneous frequency timeseries, namely an estimate of the oscillation's frequency at every timepoint. The processing up to this stage was performed on the continuous EEG recording. Instantaneous frequency timeseries were eventually segmented based on perturbation windows, and aggregated by perturbation type and direction. The average of the event-based segments provided the *event-related frequency adjustment* (*ERFA*) curves as our outcome of interest, which were expected to follow the stimulus dynamics as expressed in terms of instantaneous frequency. In the rightmost boxes of the figure, we show how we hypothesized the ERFA curves would look like. We expected a gradual and sustained response to tempo changes (top box) and a transient response to phase shifts (bottom box), tracking the stimulus dynamic and direction (represented as dashed lines).

sician. Upon arrival in the laboratory, they received extensive briefing on the experimental procedure and signed the informed consent form. Following the administration of the Edinburgh Handedness Inventory (Oldfield, 1971) and a questionnaire to collect information on demographics and musical training, the experimenter proceeded to prepare the EEG equipment.

The study was approved by the Ethics Committee of Ghent University (Faculty of Arts and Philosophy) and informed written consent was obtained from each participant. A 20€ coupon was given to all participants as economic compensation for their time.

2.2. Experimental tasks

The experiment consisted of one listening task without perturbations and two finger-tapping tasks with perturbations, performed in a fully-randomized order. Two different perturbation types were used as separate experimental conditions, these were tempo changes and phase shifts (see Fig. 1). The listening task lasted 60 seconds, while both finger-tapping tasks lasted 465 seconds, of which the first 60 seconds were free of perturbations and the following 405 seconds contained perturbations.

The listening task was used to collect EEG data in the absence of movement and design a spatial filter, to extract the perceptual component maximally attuned to the auditory stimulus. Stimulation consisted of a metronome set at 100 Bpm (1.67 Hz), which was chosen as optimal rate for sensorimotor synchronization (Kliger Amrani and Zion Golumbic, 2022; London, 2012; Zalta et al., 2020). To verify that participants were actively listening, they were instructed that the regular rhythm could be disrupted by a phase shift at any moment, and they would have to tap their finger on the pad as fast as possible upon detection. One single perturbation was presented after 60 seconds of steady rhythm, at a random moment within a 5 second time window, and was successfully detected by all participants except one. The importance of not moving any body part during the listening was stressed, as well as the importance of staring at a black dot painted on the center of the pad to minimize eye-movement artifacts in the EEG recordings.

In finger-tapping tasks, participants were instructed to synchronize their finger-taps to the metronome all the time, and to keep trying to synchronize even in case they would find themselves out of sync. The importance of minimizing body and head movements during the task was stressed, as well as the importance of staring at a black spot (i.e. a fixation point) painted on the tapping pad. Both measures were taken to

minimize the contamination of EEG recordings. Immediately before the start of the experiment, participants were given a simple demonstration by the experimenter and familiarized themselves with the finger-tapping to the non-perturbed metronome until they felt confident with the procedure.

Each finger-tapping task lasted a total of 465 seconds, the first 60 of which were free from perturbations. This steady period was used to design a spatial filter to extract the sensorimotor component maximally attuned to the metronome, based on EEG data collected in condition of overt movement. From that moment onwards, perturbations were presented at random times every 5 to 15 seconds: perturbation onsets were therefore unpredictable while participants had enough time to re-stabilize behavioral and neural responses. A total of 20 negative perturbations, 20 positive perturbations, and 40 perturbation-free windows were available to analyze per participant within a time frame of 405 seconds. The order of perturbations' direction was randomized within finger-tapping condition, while tempo changes and phase shifts were presented in separate experimental conditions. Tempo changes consisted of a +/- 10% step change with respect to the baseline tempo (100 bpm, 1.667 Hz), sustained for 3 seconds before a second step change back to the baseline. Phase shifts consisted of a discrete +/- 90° shift on the beat cycle, perceived as a shorter or a longer inter-beat interval, respectively (see Fig. 1). The order of the individual conditions (60 seconds of perturbation-free listening, 465 seconds of finger-tapping with tempo changes, 465 seconds of finger-tapping with phase shifts) was fully randomized across participants.

Besides the experimental conditions hereby presented, participants also performed the same tasks with different kinds of ecological musical stimuli. However, these extra conditions were meant to answer research questions not relevant to the present paper, hence will be addressed in future work.

2.3. Experimental apparatus

Participants were sitting on a comfortable armrest chair, and a circular pad was placed on the table in front of them for recording finger-tapping responses. Stimuli were presented via DefenderShield® airtube earbuds. Volume was adjusted before the beginning of the experiment, to make sure that stimuli were clearly audible without creating any discomfort. A circular tapping pad, containing a strain gauge pressure sensor, was used to detect finger-tapping onsets with a 1 ms resolution. A folded towel was placed underneath the pad, so that no auditory feedback from the finger-taps was perceivable. The pressure sensor was connected to a Teensy 3.2 microcontroller, which worked as serial/MIDI hub for data logging and communication between the stimulation computer and the EEG recording computer. Upon occurrence of events of interest (i.e., perturbations), a TTL trigger was sent from the Teensy microcontroller to the EEG amplifier via BNC connection, granting the alignment between behavioral and neural timeseries. The EEG signals were recorded with an ANT-Neuro eego™mylab system at a sampling rate of 1 kHz. Each participant was equipped with an EEG headset (64-channel waveguard™original with Ag/AgCl electrodes). Recordings were performed with a referential montage, with 'CPz' being the reference for all electrodes.

The stimuli sequence was played back by specifically designed software developed in Max MSP 8 (Cycling '74, USA), running on the stimulation computer (Windows 10, Intel core i7 8th gen, Focusrite Rednet PCIExpress ASIO low-latency soundcard). Prior to each trial, a randomized balanced list of perturbations was generated and inspected by the experimenter. A pre-generated .wav file, containing a non-perturbed sequence of metronome ticks at 1.667 Hz (600 ms inter-beat intervals), was played back and manipulated in real-time based on the perturbations list and timing configuration (i.e., initial perturbation-free period and minimum and maximum time between perturbations). MIDI events were generated when a beat was perceived by the listener and when a perturbation occurred. These MIDI events were logged using the Teensy 3.2 microcontroller, along with the finger-tapping timestamps.

2.4. Data analysis

The data processing pipelines were implemented in Matlab 2019a (MathWorks Inc., USA). Statistical modeling was carried out in R Studio version 4.0.3 (R Core Team), using the lme4 package (Bates et al., 2014) for model fitting. Participant #6 was excluded from the analysis, due to a technical issue during data acquisition which resulted in the loss of EEG data.

2.4.1. Behavioral data processing

Finger-taps were processed in order to return timeseries aligned to the neural ERFA, expressed in the same unit of measure (i.e., Hz over ms). Whenever a timestamp was followed by a second timestamp with < 350 ms interval, it was considered as a false positive and therefore removed (false positives could occasionally be recorded when a participant pushed the tapping pad for too long or accidentally laid the hand on it). The intervals between the remaining timestamps were then linearly interpolated from 0 to 1 at 1 kHz sampling rate. The resulting ramp wave was scaled to 2π, providing an estimate of the finger-taps phase with a temporal resolution of 1 ms. Instantaneous frequency timeseries were computed as the first derivative of the unwrapped phase angles time series (Boashash, 1992), and converted to Hz as indicated in (Cohen, 2014):

$$Hz_t = \frac{s\left(\phi_t - \phi_{t-1}\right)}{2\pi}$$

where s indicates the data sampling rate and ϕ_t indicates the (unwrapped) phase angle at time t. Unwrapping was necessary in order to remove discontinuities in the timeseries caused by phase resets.

2.4.2. Neural data processing

Pre-processing. The pre-processing pipeline was realized integrating functions from the *Fieldtrip* toolbox (Oostenveld et al., 2011) for Matlab. Bad channels were identified by visually inspecting the raw timeseries in combination with the distribution of variance across channels. The recording was re-referenced to the average of all electrodes after rejection, to avoid noise leakage into the common average. 3.23 bad channels per participant were removed on average (std = 1.95). A sixth-order Butterworth high-pass filter with 1 Hz cut-off was applied to the raw recordings to remove slow drifts. We show in *Supplementary material 1* that, for these parameters, the high-pass filter does not influence the oscillatory dynamics of interest in the present work. A low-pass sixth-order Butterworth filter with 40 Hz cut-off was applied to remove high-frequency muscular activity. A fourth-order notch filter centered at 50 Hz was applied to remove power-line noise up to the 3rd harmonic.

Subsequently, independent component analysis (ICA) was conducted and used to remove stereotyped artifacts by means of visual inspection of topographical maps and timeseries of component activation, as implemented in the 'runica' *Fieldtrip* algorithm. The reference 'CPz' and the bad channels' timeseries were excluded from the input matrix. Under optimal conditions, removal was limited to those components which exhibited the stereotypical frontal distribution generated by blinks and lateral eye movements, or bilateral temporo-mastoidal distribution with periodic peaks in the activation timeseries plausibly generated by heart beats. Extra components were removed in instances where recurrent artifacts with clearly abnormal amplitude were detected. 5.03 artifactual components per participant were identified and removed on average (std = 4.36). The dataset was inspected prior to ICA decomposition and following ICA back-projection to assess the quality of the artifact removal. Special attention was given to the frontal clusters of electrodes maximally contaminated by eye-related artifacts. Rejected bad channels were reconstructed after artifact removal, by computing the average activity from neighboring electrodes indicated by the template provided by ANT-Neuro for 64-channel waveguard™original caps.

No segmentation in epochs was performed up to this point, since continuous recording was needed for performing source separation.

Source separation. Generalized eigendecomposition (GED) (Cohen, 2022) was used to extract a perceptual and the sensorimotor components attuned to the stimulation frequency. The procedure described below is the same for both components, with the only difference that the respective inputs were 60 seconds of data recorded during the listening task, and 60 seconds of data recorded during finger-tapping in the absence of perturbations. This allowed to design two separate spatial filters, which were subsequently applied to the data recorded during the 405 seconds of finger-tapping in the presence of perturbations (see Fig. 1).

As first described in the context of source separation for rhythmic entrainment (Cohen and Gulbinaite, 2017), GED allows to avoid channel selection bias while optimizing the signal-to-noise ratio between the entrained component and the broadband neural activity. The technique consists of a spatial filter to separate sources and reduce data dimensionality, guided by some criteria. In this case, the criterion was the attunement to the stimulation frequency. Dimensionality was reduced by computing the weighted average of the timeseries from all 64 channels, where the set of vectors W (weights) was calculated by solving the following equation:

$$RW\Lambda = SW$$

where S is the covariance matrix of the narrow-band signal; R is the reference covariance matrix of the broad-band signal; Λ is the set of eigenvalues. GED identifies eigenvectors W that best separate the signal ('S') covariance from the reference ('R') covariance matrix. The eigenvector associated to the largest eigenvalue was taken as a spatial filter, transposed and multiplied by the broadband data matrix to reconstruct the time series of our target entrained component.

Given we were explicitly looking for frequency fluctuations, our narrow-band filter needed to be wide enough to leave room for fluctuations around the target frequency. We designed our filter as a Gaussian function in the frequency domain, with center at 1.667 Hz and a width of 0.3 Hz at half of the maximum gain (Rosso et al., 2021). We then filtered the broadband data at all channels via spectral multiplication between broadband signal and wavelet kernel in the frequency domain, and transformed the resulting narrow-band signal back in the time domain with inverse fast Fourier transform.

The S covariance matrix was computed from the narrowband signal, the R covariance matrix was here computed from the broadband multivariate signal. In this regard, the choice aligns with the approach presented in (Rosso et al., 2021) rather than (Cohen and Gulbinaite, 2017), because the former is optimized for low target frequencies < 2 Hz. Covariance matrices were computed within 600 ms time windows locked to the finger-taps onsets, and grand-average S and R covariance matrices were computed. Matrices whose z-normalized Euclidean distance from the grand-average exceeded the 2.23 z-scores (corresponding to a probability of 0.01) were removed, and the grand-average S and R wererecalculated free of transient artifactual activity. 1% regularization was applied to the R matrix.

In previous work (Rosso et al., 2022; Rosso et al., 2021), we implemented an optimization of GED based on a macro-selection of regions of interest, justified by experimental design and some prior assumptions. In this study, however, we made no prior assumptions on scalp distribution and therefore inputted timeseries from all channels, allowing the data collected to drive the source separation in the different conditions ('listening' and 'finger-tapping').

Instantaneous frequency. The same Gaussian filter used for GED was applied to the perceptual and to the sensorimotor components (center at 1.667 Hz and 0.3 Hz width at half maximum) to extract reliable phase timeseries from the analytic signal (Rosenblum et al., 2001). These were extracted by means of Hilbert transform. In order to remove discontinuities caused by phase resets, the timeseries were unwrapped, differenced and finally converted to Hz (Cohen, 2014):

$$Hz_t = \frac{s(\phi_t - \phi_{t-1})}{2\pi}$$

The resulting instantaneous frequency timeseries were smoothed with a sliding moving median (window width of 400 samples), to remove transient artifactual activity that may distort the phase timeseries (Cohen, 2014).

It should be to pointed out that, in electrophysiological data, the estimation of instantaneous frequency of oscillatory activity is sensitive to the aperiodic 1/f component of the spectrum (Donoghue et al., 2020). This feature can result in bias towards lower frequencies under certain conditions. Samaha and Cohen (Samaha and Cohen, 2022) demonstrated that a low periodic/aperiodic spectral power ratio, in combination with broad filter width, is problematic for reliable estimations. We highlight that our spatial filter application was specifically aimed at maximizing the target narrowband energy relative to the broadband energy (Rosso et al., 2021; Cohen and Gulbinaite, 2017; Haufe et al., 2014), resulting in outstanding spectral peaks indicating prominent oscillatory activity (see Fig. 2). This, along with our conservative choice of a narrow filter, complies with the good practices for reducing the slope of the 1/f, and therefore increases the reliability of the estimated instantaneous frequency (Samaha and Cohen, 2022).

Event-related frequency adjustment (ERFA). In order to analyze neural and behavioral responses to perturbations within the same framework, the following procedure was applied to all instantaneous frequency timeseries to compute the related instances of ERFAs: one for the finger-tapping (behavioral), one for the EEG perceptual component (neural), and one for the EEG sensorimotor component (neural). The approach was inspired by event-related potentials (ERPs) (Luck, 2014) and shares with it most of the features presented in this paragraph. The crucial difference is that ERFAs express frequency (y-axis) over time (x-axis), while ERPs express amplitude (y-axis) over time (x-axis). Furthermore, the pre-stimulus baseline of ERFAs requires steady rhythmic stimulation to provide a stable baseline frequency level.

Perturbation onsets were identified in the instantaneous frequency timeseries based on the timestamp logs. Time windows were defined as the time span from −500 to +3000 ms with respect to the perturbation onset, aggregated per perturbation type and direction, baseline-normalized and averaged. Baseline normalization was performed by subtracting the average of the 500 to 0 ms interval from the rest of the ERFA, dividing by the target stimulation frequency (i.e., 1.667 Hz) and multiplied by 100. The resulting ERFA is expressed in percent change with respect to the baseline stimulation frequency. For every participant, 19 ERFA curves were aggregated per perturbation type and direction, and the average response was computed. The responses to negative perturbations were sign-flipped in order to avoid trivial results at statistical comparisons with positive perturbations (Jantzen et al., 2018). Finally, the ERFA computation was repeated shifting all time-windows by −400 ms along the instantaneous frequency timeseries, namely into perturbation-free periods. This provided for every participant a baseline for further statistical comparisons.

2.5. Statistical modeling

ERFAs to tempo change perturbations were modeled via polynomial fitting to the curves, downsampled by a factor of 10 for computational feasibility. For phase shifts, the discrete integral of the ERFA from 0 to 1500 ms interval was computed via trapezoidal method with unit spacing, using the *trapz()* Matlab function. This returned signed areas under the curves, which we interpreted as a measure of entity of the error correction sensitive to the direction of the corrective response. Baseline normalization is crucial in order to obtain signed areas, as data need to be zero-centered.

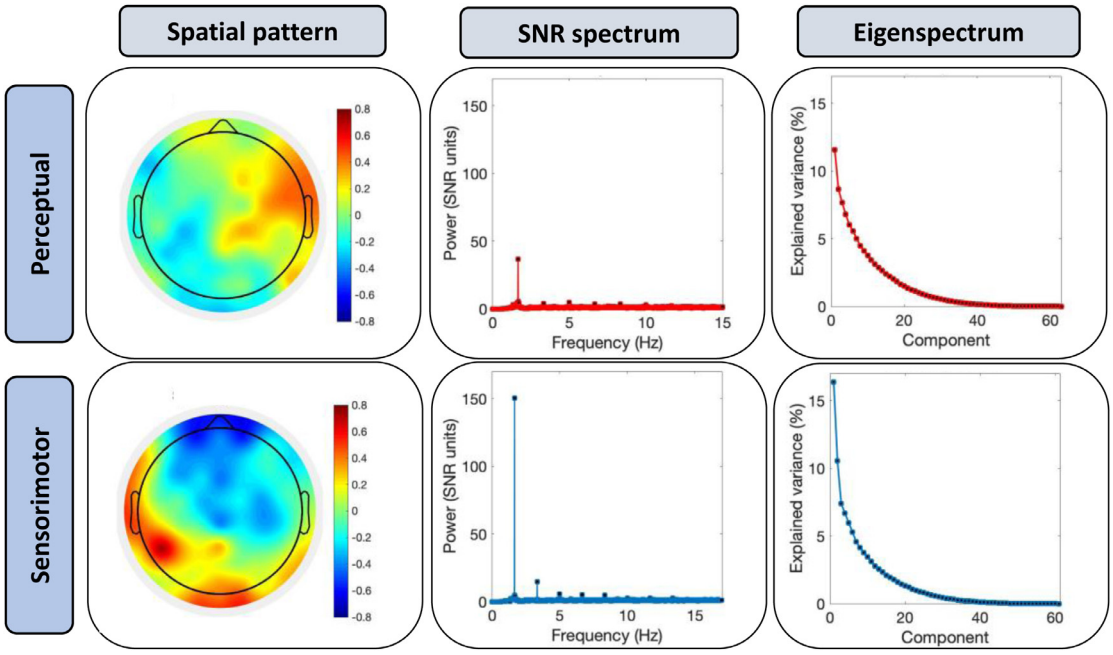

Fig. 2. Source separation. The criteria for the assessment of source separation via GED are here shown at the group level (N = 19). The *spatial pattern* of the perceptual component did not exhibit the centro-frontal activation expected from EEG auditory evoked responses (Nozaradan et al., 2011; Nozaradan et al., 2012; Nozaradan et al., 2015), whereas the activation of the sensorimotor component exhibited both frontal negativity and a peak of activation over left centroparietal regions as previously reported during finger-tapping task performed with the contralateral hand (Rosso et al., 2021). The *spectrum* of both components was dominated by a prominent peak at the stimulation frequency (i.e., 1.67 Hz) and harmonics, indicating that our filter successfully separated from the broad-band signal a component oscillating around the center frequency of the metronome. Spectra were computed over the whole finger-tapping session, and normalized to signal-to-noise ratio (SNR) units with respect to the neighboring bins to remove the 1/f component. The *eigenspectrum* showed that the weights associated to the largest eigenvalue explained considerably more variance than the others, hence was chosen as spatial filter for our sensor data. We observed a general additive effect of the sensorimotor component on all 3 criteria: overt movement resulted in stronger spatial activation, finer attunement to the stimulation frequency and more explained variance in the reconstructed sources. Taken together, these criteria support that the source separation successfully extracted two components attuned to the metronome's frequency during listening (perceptual component) and finger-tapping (sensorimotor component), with the latter being more effectively separated from the broadband signal. The dynamic frequency adjustment of the two components in response to perturbations was further operationalized as event-related frequency adjustment (ERFA).

For behavioral ERFAs only, reaction times were computed as the x-coordinate of the inflexion point of sigmoid function which best fitted the ERFA from 0 to 800 ms. The fitting was performed with the *sigm_fit()* Matlab function, defining no fixed parameters.

For the neural ERFAs, the effects on the response variables of interest were tested in 3 × 2 factorial designs with Direction (Negative, Null, Positive) and Component (Perceptual, Sensorimotor) as factors. For the behavioral ERFAs, the effects on the response variables of interest were tested in 3 × 1 factorial design with Direction (Negative, Null, Positive) as the only factor. When pairwise comparisons between Negative and Positive directions were performed after omnibus tests, Bonferroni correction was applied for a significance level of $\alpha = 0.05$.

2.5.1. Orthogonal polynomials (for tempo change)

For tempo changes, growth curve analysis (Rosso et al., 2021; Mirman, 2017) was used to model the change in instantaneous frequency of the neural entrained components (i.e., the ERFAs) within 3-second perturbed windows (i.e., metronomes at +/- 10% of the baseline frequency). Parameters of categorical predictors were estimated relative to Null Direction and Perceptual Component, respectively. The rationale of such factor leveling is that no systematic change in instantaneous frequency was expected in Null trials, resulting in the flattening of random fluctuations when averaging over trials. As for the Component, we intended to test the significance of additive motor processing in the Sensorimotor component during entrainment, with respect to purely Perceptual processing.

The order of the polynomial is ideally chosen based on hypothesis and on the nature of the data, should be confirmed by the empirical data, and should allow a straightforward interpretation of the effects (Mirman, 2017). A full model based on orthogonal polynomials includes all terms up to the chosen order, which in this case was 2: a linear and a quadratic term were sufficient to capture the effect of our manipulation. In a framework of polynomial fitting, a flat line can be considered a pure intercept and a 'zero-order' polynomial, in the sense that it exhibits zero change in any direction. When instantaneous frequency is time-invariant, it indicates a steady oscillation within a given time window (e.g., the behavior of a non-perturbed metronome). The intercept of our model provides hence valuable information on the average frequency within the perturbed window, and how it changes across the factors' levels: significant main effects of categorical predictors indicate global differences across experimental conditions, independently from the temporal profile of the response variable. On the other hand, significant interaction effects on the polynomial terms indicate that instantaneous frequency in systematically modulated by the temporal dynamics of the perturbation. The parameter estimate of the linear component (1st order) corresponds to the slope of the line and the consequent shift in the vertex for the parabola, while the quadratic component (2nd order) corresponds to the parabolic curvature. In summary, intercepts indicate average frequency while polynomial terms of Time are used to model instantaneous changes in frequency. Our model also included random effects of Subject on all polynomial terms, and their interactions with the factors: the random effects structure was maximized in order to minimize false alarm rates without substantial loss of power (Barr et al., 2013).

Here, the formulas for the models fitted to neural and behavioral ERFAs, respectively (in R syntax):

$$nERFA \sim (Time + Time^2) * Component * Direction + (Time + Time^2 \mid Subject) + (Time + Time^2 \mid Subject : Component : Direction)$$

$$bERFA \sim (Time + Time^2) * Direction + (Time + Time^2 \mid Subject) + (Time + Time^2 \mid Subject : Direction)$$

2.5.2. Integrals (for phase shifts)

For phase shifts, we deemed a model based on integrals to be more suited and parsimonious as compared to polynomials. This was motivated by the observation that a parabolic curve did not provide a good fit for transient responses elicited by phase shifts. We therefore opted for a method insensitive to the particular curve-shape, quantifying instead the area under the curve within a shorter time window (1500 ms) following the perturbation onset. The adoption of integrals in the context of event-related neural responses was validated in the ERP literature (Luck, 2014), and in principle equally valid in the context of these ERFAs given the comparable post-stimulus window sizes and a response more localized in time, in contrast to the more sustained responses to tempo changes.

The integrals provided us with a measure of ERFA magnitude, which is sensitive to the direction of the response since the areas are signed with respect to the 0% change baseline. Notably, we verified on the data logged from the metronomes that the expected absolute area underneath positive and negative phase shifts were equal irrespective of the difference in shape.

Here are the formulas for the models fitted to neural and behavioral ERFAs, respectively, plus the behavioral reaction times (in R syntax):

$$nERFA\_integral \sim Component * Direction + (1 \mid Subject)$$

$$bERFA\_integral \sim Direction + (1 \mid Subject)$$

$$bERFA\_rt \sim Direction + (1 \mid Subject)$$

Data and scripts are available upon request to the authors with a formal data sharing agreement, in line with the conditions of the local ethics committee which approved the present study.

3. Results

Event-related frequency adjustments (ERFAs), computed as the instantaneous frequency response within the post-perturbation window, are the object of our analyses. Neural and behavioral ERFAs were computed from EEG and finger-tapping timeseries, respectively. Two distinct approaches were used to model the ERFAs in response to the tempo change and phase shift perturbations. For both neural and behavioral responses, orthogonal polynomials (Mirman, 2017) were used to model tempo change, and integrals (Luck, 2014) were used to quantify the entity of phase shifts. ERFAs in response to all negative perturbations were sign-flipped, in order to make the responses directly comparable with the positive counterparts (Jantzen et al., 2018). Whereas in the literature stimulus rates are often expressed in milliseconds to describe the duration of inter-stimulus intervals, ERFAs will be consistently expressed in Hz, or % change with respect to a baseline frequency. Therefore, increasing values indicate that the oscillatory component is speeding up, and decreasing values indicate that the oscillatory component is slowing down.

Below, an assessment of the EEG source separation is provided, followed by the results of our models presented per perturbation type.

3.1. EEG source separation

A perceptual and a sensorimotor component were extracted from the multivariate signal from separate experimental conditions (see Fig. 1). This represented the *attunement* phase of our approach, where spatial filters were designed to target the perceptual and sensorimotor components maximally entrained to the metronome. We hereby provide the results of the qualitative assessment of the GED according to three criteria: (a) spatial activation pattern, (b) spectral profile, and (c) eigenspectrum. These results are shown and described in detail in Fig. 2.

3.2. Tempo change perturbations

The ERFA response curves of the tempo change conditions were modeled with 2nd order orthogonal polynomials, as the timeseries followed parabolic growth-and-decay, tracking the direction of the perturbation (see Fig. 3A and C). This parabolic term was capable of capturing the most prominent effect in the model. Fixed effects of Direction (Negative, Null, Positive) and Component (Perceptual, Sensorimotor) on the polynomial terms were tested. The Null level (i.e., perturbation-free windows) was treated as a baseline for contrasting Negative and Positive levels of Direction, while the Perceptual level was treated as a baseline for contrasting the Sensorimotor level of Component. Model parameters were estimated with respect to the levels defined as baseline. The same polynomial model was fitted to the behavioral ERFA curves, with the exception that the Component factor was removed. The specifics of our statistical modeling are explained in detail in the *Materials and methods* section.

3.2.1. Neural ERFA

The orthogonal polynomial model revealed a significant main effect of Positive Direction (Estimate = 1.689, SE = 0.450, $p < 0.001$) and a significant two-way interaction between Positive Direction and the quadratic term of Time (Estimate = −8.049, SE = 4.039, $p = 0.046$). The former indicates that both Perceptual and Sensorimotor components oscillated on average significantly faster within Positive perturbed windows, the latter indicates that the higher average was accompanied by more parabolic modulation of the ERFA (i.e., inverse U-shape). While the main effect of Negative Direction reached significance (Estimate = 1.160, SE = 0.450, $p = 0.010$), no significant two-way interaction with the polynomial terms was found for Negative Direction. Taken together, these effects suggest that oscillations are generally biased to follow positive frequency changes, as quantified by the ERFAs of both perceptual and sensorimotor components.

The nature of the Component becomes relevant for Negative Direction, as indicated by the significant two-way interaction (Estimate = 1.461, SE = 0.636, $p = 0.022$): the oscillation was on average significantly slower within Negative perturbed windows, but only for the Sensorimotor type. In contrast, in response to Positive perturbations, both Component types exhibited significantly faster oscillations.

A significant three-way interaction was found between quadratic term of Time, Component and Direction (Positive: Estimate = −19.129, SE = 5.711, $p < 0.001$; Negative: Estimate = −20.266, SE = 5.711, $p < 0.001$). These effects indicate that the Sensorimotor Component significantly boosted the parabolic modulation of the ERFA as compared to the Perceptual Component, regardless of whether the metronome was speeding up or slowing down. This additive effect can be seen in Fig. 3A and B. Results from the statistical model are reported in Table 1.

3.2.2. Behavioral ERFA

The orthogonal polynomial model revealed significant main effects of both Positive (Estimate = 8.816, SE = 0.567, $p < 0.001$) and Negative (Estimate = 9.276, SE = 0.567, $p < 0.001$) Directions, indicating that both signs resulted in an average frequency offset in the expected direction. The dynamic is captured by the significant interaction of the

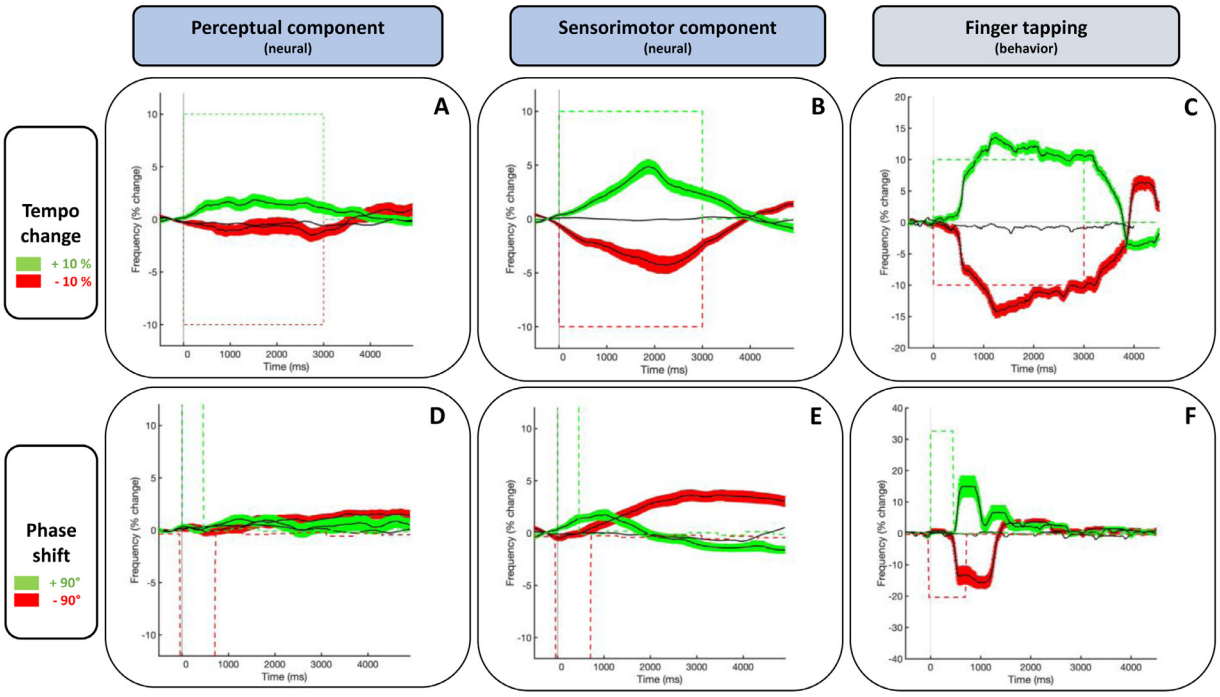

Fig. 3. ERFA curves. Results are here represented as grand-averages across ERFA types and perturbation types ($N = 19$; 19 trials per participant for every perturbation type), and expressed in percentage change with respect to the pre-perturbation stimulation frequency. Positive and negative perturbations are color-coded in green and red, respectively. For more visual clarity, the average curves are always represented as a black continuous line, whereas the contour representing the standard error of the mean (SEM) is color-coded according to the direction of the perturbation. The black continuous line without contour, approximately flat, represents the average frequency change in a sample of 19 non-perturbed time windows, where the frequency of the metronomes was stable. Dashed lines represent the instantaneous frequency timeseries of the metronome. Due to different magnitude of neural and behavioral responses, these are shown on different scales to better visualize the dynamic. Labels correspond to the following: A- neural ERFA (perceptual component, tempo change); B- neural ERFA (sensorimotor component, tempo change); C- behavioral ERFA (tempo change); D- neural ERFA (perceptual component, phase shift); E- neural ERFA (sensorimotor component, phase shift); F- behavioral ERFA (phase shift).

Table 1
Neural ERFA: tempo change ($N = 19$).

Predictors	Estimate	SE	t value	p
(Intercept)	−0.441	0.318	−1.388	0.165
Time	−0.814	2.665	−0.305	0.760
Time^2	0.011	2.978	0.003	0.996
Component (SM)	0.475	0.450	1.057	0.291
Direction (-)	1.160*	0.450	2.580	0.010
Direction (+)	1.689***	0.450	3.754	< 0.001
Time:Component	1.407	3.623	0.388	0.698
Time^2:Component	1.540	4.038	0.381	0.703
Time:Direction (-)	−1.255	3.623	−0.346	0.729
Time:Direction (+)	−1.936	3.623	−0.534	0.593
Time^2:Direction (-)	−6.141	4.039	−1.520	0.128
Time^2:Direction (+)	−8.049*	4.039	−1.993	0.046
Component:Direction (-)	1.461*	0.636	2.296	0.022
Component:Direction (+)	0.743	0.636	1.167	0.243
Time:Comp:Dir (-)	−3.250	5.124	−0.634	0.526
Time:Comp:Dir (+)	1.413	5.124	0.275	0.783
Time^2:Comp:Dir (-)	−20.266***	5.711	−3.548	< 0.001
Time^2:Comp:Dir (+)	−19.129***	5.711	−3.349	< 0.001

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

quadratic term and Direction (Positive: Estimate = −92.034, SE = 4.562, p < 0.001; Negative: Estimate = −85.739, SE = 4.562, p < 0.001), in line with the patterns of the neural ERFAs. When testing for the interaction between Direction and the linear term of the model, which indicates the position of the vertex of the parabolic curve (see Rosso et al., 2021), for application and interpretation of orthogonal polynomials fitted to asymmetric curves), we did not find an asymmetry across levels of Direction at our significance level alpha = 0.05, only a trend (Estimate = 9.293, SE = 4.971, p = 0.061). The behavioral ERFA to tempo changes is represented in Fig. 3C.

Table 2
Behavioral ERFA: tempo change ($N = 19$).

Predictors	Estimate	SE	t value	p
(Intercept)	−0.709	0.560	−1.267	0.205
Time	−1.018	4.834	−0.211	0.833
Time^2	3.957	3.871	1.022	0.307
Direction (-)	9.276***	0.567	16.345	< 0.001
Direction (+)	8.816***	0.567	15.534	< 0.001
Time:Direction (-)	9.293	4.971	1.869	0.061
Time:Direction (+)	0.830	4.971	0.167	0.867
Time^2:Direction (-)	−85.739***	4.562	−18.794	< 0.001
Time^2:Direction (+)	−92.034***	4.562	−20.174	< 0.001

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

No significant difference was found at post-hoc comparisons when testing for Positive against Negative directions, neither for the main effect nor for the interaction on the quadratic terms. Results from the statistical model are reported in Table 2. Note that, since the Null level was set as a reference for both Positive and Negative in the model, effects of Direction are reported pairwise.

3.3. Phase shift perturbations

The signed area (i.e., discrete integrals) under the ERFA curves was computed as response variable, and a 3 × 2 factorial model was fitted: Direction (Negative, Null, Positive) x Component (Perceptual, Sensorimotor). In this context, integrals provided a measure of entity of the shift in instantaneous frequency regardless of the curve shape, while

Table 3
Neural ERFA: phase shift ($N = 19$).

Predictors	Estimate	SE	t value	p
(Intercept)	557.929	413.656	1.349	0.177
Component (SM)	−617.651	552.777	−1.117	0.264
Direction (-)	−775.485	552.777	−1.403	0.161
Direction (+)	297.630	552.777	0.538	0.590
Component:Direction (-)	−24.904	781.745	−0.032	0.974
Component:Direction (+)	1763.756*	781.745	2.256	0.024

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4
Behavioral ERFA: phase shift ($N = 19$).

Predictors	Estimate	SE	t value	p
(Intercept)	−372.017	1772.633	−0.210	0.834
Direction (-)	10,920.459***	2506.882	4.356	< 0.001
Direction (+)	9197.843***	2506.882	3.670	< 0.001

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

being sensitive to the direction of the shift. This sensitivity is due to the fact that the area is signed, returning positive or negative values for portions above or below the x-axis, respectively. In addition, for behavioral ERFAs, reaction times were computed as the intersection of the curve's inflection point with the time x-axis. The specifics of our statistical modeling are explained in detail in the *Materials and methods* section.

3.3.1. Neural ERFA

We found a significant two-way interaction between Component and Positive Direction (Estimate = 1763.756, SE = 781.745, p < 0.024), indicating that phase shifts in such direction elicited significant ERFAs only when the component was Sensorimotor. The difference between components can be seen in Fig. 3D and E, where a flat grand-average ERFA is observed for the Perceptual component due to the absence of systematic changes in instantaneous frequency across participants. It is noteworthy that the ERFA of the Sensorimotor component in Negative Direction did not mirror its Positive counterpart: rather than dropping below the baseline frequency before re-stabilizing, the curve underwent a later transition and stabilization at a higher frequency (Fig. 3E). This resulted in a very small area under the ERFA in the defined post-stimulus window, and a non-significant effect in our model. Results from the statistical model are reported in Table 3.

3.3.2. Behavioral ERFA

A significant main effect of Direction with respect to the Null level (Positive: Estimate = 9197.843, SE = 2506.882, p < 0.001; Negative: Estimate = 10,920.459, SE = 2506.882, p < 0.001) was found. A post-hoc comparison between Positive and Negative levels did not yield a significant effect of Direction. This indicates that there were no differences in magnitude of error adjustments across perturbation directions. However, a *t*-test revealed a significant difference between reaction times to Positive and Negative perturbations (Estimate = 49.316, SE = 11.143, p < 0.001), indicating that participants corrected the errors significantly faster in response to Negative phase shifts. The behavioral ERFA to phase shifts is represented in Fig. 3C. Results from the statistical model are reported in Tables 4 and 5.

4. Discussion

The aim of the present work was to present a paradigm and a measure capable of quantifying neural entrainment from an electrophysiological brain signal, rigorously informed by the fundamental definition of the process (Lakatos et al., 2019; Novembre and Iannetti, 2018; Rajendran and Schnupp, 2019). To tackle this methodological challenge (Haegens and Zion Golumbic, 2018), we moved away from a frequency-domain representation of oscillatory components in the EEG signal. Instead, we modeled their frequency adjustment as a function of time, provided a controlled manipulation of the stimulus dynamics. In the context of a finger-tapping synchronization task with perturbed auditory metronomes, event-related frequency adjustments (ERFAs) revealed how oscillatory components entrain to the stimulus by speeding up and slowing down, tracking dynamic rhythmic changes within critical time windows.

Table 5
Reaction times: phase shift ($N = 19$).

Predictors	Estimate	SE	t value	p
(Intercept)	490.789	7.879	62.290	< 0.001
Direction (+)	49.316***	11.143	4.426	< 0.001

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Crucially, our experimental design allowed us to disentangle in the brain signal a perceptual and a sensorimotor component, separately attuned to the auditory metronome via GED (Cohen, 2022), and to statistically compare their ERFAs (see Fig. 1). The results showed that sensorimotor processing is critical for neural entrainment, if not a necessary condition for it to take place. When compared to perceptual ERFAs, sensorimotor ERFAs exhibited a significantly stronger modulation, congruent with the direction of the perturbation (Fig. 3). It should be noted that, while with tempo changes the effect was observed when either speeding up or slowing down the stimulus by 10% of its frequency, only phase shifts of 90° in the positive direction (perceived as a shorter inter-beat interval) were tracked following the stimulus dynamics (Fig. 3B and E). These patterns resemble a more smoothed version of the behavioral ERFAs computed from finger-tapping data (Fig. 3E and F). On the other hand, negative phase shifts (perceived as a longer inter-beat interval) elicited a sensorimotor ERFA qualitatively different than expected, deviating from both stimulus and behavioral dynamics. Specifically, we observed in this case an initial destabilization followed by a gradual increase in frequency, suggestive of a dissociation between the observed behavioral adjustments and their underlying neural mechanisms (Fig. 3E and F).

Although in EEG data it is not possible to fully disentangle the entrainment of endogenous oscillations from regularly evoked responses which are time-locked to the stimulus, we highlight how our approach contributes to tackle this hard problem in the investigation of neural entrainment (Haegens and Zion Golumbic, 2018). We propose that identifying and statistically contrasting the perceptual and sensorimotor components serve as an initial progression towards better understanding the interplay of the mechanisms involved. Notably, while the perceptual ERFAs can in principle be driven by changes in the stimulation rate, the additional modulation observed in the sensorimotor ERFAs cannot be explained by a bottom-up processing alone. This is because the only difference across the two components is the engagement of the motor system, while the stimulation remains constant. Furthermore, our results reveal some features in the ERFAs which are not fully explainable by changes in the physical stimuli, calling into question the functioning of endogenous sensory rhythms in the brain and their own intrinsic dynamics. To elaborate, we reported asymmetries across positive and negative directions for both perceptual and sensorimotor ERFAs. For tempo changes, the perceptual component exhibited a moderate frequency adjustment expected for the positive tempo-change, but not for the negative tempo-change (Fig. 3A). If the perceptual ERFA were entirely driven bottom-up, we would expect positive and negative responses to mirror each other, because the absolute magnitude of the tempo change was the same in both directions. However, the involvement of sensorimotor processing resulted in a significantly more prominent slowing down of the oscillation in response to negative tempo change (compare the

depth of the red curves in Fig. 3A and B). Taken together, we argue that this evidence points to a general bias towards speeding up, which suggests the presence of intrinsic oscillatory dynamics at play. Following this argument, we propose that adjusting the frequency to the slowing metronome requires an extra deployment of neural resources, which are recruited by engaging the motor system to a greater extent (Kliger Amrani and Zion Golumbic, 2022) and therefore reflected in the enhanced sensorimotor ERFA.

We put forward two putative mechanisms to explain the interaction between perceptual and sensorimotor components. The first possibility (A) is that auditory rhythms do not entrain at all, and instead pure evoked responses are generated in sensory areas. The function of these periodic responses may be to form an amplitude-based representation of the external rhythm at the cortical level, so that endogenous motor rhythms (Morillon et al., 2014; Morillon et al., 2015; Morillon and Baillet, 2017) can entrain by dynamically adapting their frequency via phase-based alignment. Alternatively (B), endogenous auditory rhythms might be effectively entraining to the metronome (Lakatos et al., 2007; Lakatos et al., 2008; Lakatos et al., 2009; Lakatos et al., 2013; Lakatos et al., 2016), and in turn drive the alignment of endogenous motor rhythms. We argue that the evidence hereby presented leans in favor of the latter mechanism, which is also more coherent with recent neurophysiological models proposing that neural firing rates encode an abstract representation of stimulus and movement cycles', enabling both beat perception and overt synchronization to the beat (Cannon and Patel, 2021). Backed by evidence from monkeys (Gámez et al., 2019; Cadena-Valencia et al., 2018) and humans (Teki et al., 2011; Bengtsson et al., 2009; Chen et al., 2008; Grahn and Brett, 2007), the authors proposed that the supplementary motor area (SMA) is the key structure responsible for this cyclic sensorimotor process, working as interface between auditory pathways and motor areas.

For the sake of completeness, we provide as *Supplementary material 2* the grand-average activation timeseries for perceptual and sensorimotor components during tempo changes, to give a visual impression of the evoked responses in the signal from which ERFAs were computed. Although our task was not designed for ERPs, it is still noteworthy to see that there is no visible pattern of evoked responses expected by the stimulation rates. Additionally, we compared in a series of simulations ERFAs produced by an oscillatory model and ERFAs produced by an alternative model of evoked responses (see *Supplementary material 3*). Overall, the ERFAs computed from oscillations appear to be more robust to varying levels of noise and are still reliable in conditions of poor signal-to-noise ratio, which better approximates the reality of signals recorded with EEG. Given all the above, we argue that the parabolic ERFAs observed in Fig. 3A and B are better explained by an oscillatory model, as compared to evoked responses passively tracking changes in stimulation rate.

For phase shifts, the asymmetry across directions was more radical: in response to negative perturbations, the sensorimotor ERFA showed a gradual transition towards higher frequency (Fig. 3E). Given this class of perturbation was more localized in time and more prominent than a sustained −10% tempo change, it may be less demanding for the brain to gradually speed up an entrained oscillation and catch up with the beat over some cycles, rather than suddenly slowing down to track the dynamic of a phase shift. Evidence from neuropathology supports the idea that a bias for faster tempi is a feature in healthy adult population, which can be impaired by lesions in the cerebellum. When such structure, critical for event-based timing (Schwartze et al., 2016), is compromised, high stimulation rate becomes detrimental for the neural tracking of the beat as quantified by frequency tagging (Nozaradan et al., 2017), while a case study suggests that behavioral synchronization improves when stimulation rate is below the spontaneous rate of movement (Moumdjian et al., 2022). Another observation in favor of the entrainment of sensory rhythms is the fact that, while perceptual ERFAs to tempo changes exhibit a certain degree of curvature (Fig. 3A), they are flat in response to phase shifts (Fig. 3D). Endogenous oscillations would in fact need several cycles to entrain, a condition that is met with sustained tempo changes.

Finally, the topography of the perceptual component (Fig. 2) does not resemble the fronto-central cluster expected from auditory evoked responses (Nozaradan et al., 2011; Nozaradan et al., 2012; Nozaradan et al., 2015). We observed instead a more distributed pattern, whose activation was considerably weak compared to the sensorimotor component. Active motor engagement resulted once again in an additive effect, with maximal left centro-parietal activation as previously reported during finger-tapping performed with the right hand (Rosso et al., 2021). Although, in line with recent research (Kliger Amrani and Zion Golumbic, 2022; Cheng et al., 2022), the evidence hereby provided points at a special influence of the motor system on neural entrainment, we cannot rule out that the differences observed across components may be partly explained by improved signal-to-noise ratio in the sensorimotor component. Furthermore, we do not know whether the same neural adjustments would have been observed in the absence of the motor requirement of the finger-tapping task, and to what degree. We emphasize the need for future research to expand on the present experimental design and better address the role of motor involvement in neural entrainment.

Alongside the neural ERFAs discussed so far, we also analyzed behavioral ERFAs as the change in the instantaneous frequency of finger-tapping during overt error correction responses to perturbations. We observed highly consistent dynamics at the group level, but also some degree of interindividual variability in response to phase shifts, showing that some participants adopted different strategies to correct the synchronization error by deviating from the stimulus dynamic. The systematic classification of such strategies is out of the scope of this paper. Adaptation to tempo changes accurately tracked the stimulus dynamics, and results were consistent with the reported neural responses: a significant difference in average frequency and parabolic curvature was found in both positive and negative perturbations as compared to the baseline. When testing the two directions at post-hoc comparisons, no significant differences were found in the dynamics of the behavioral responses. As for the adaptation to phase shifts, participants adapted as expected in both directions. Although no significant difference in the entity of the correction across directions was found, a test on the reaction times revealed that participants were significantly faster in adapting to negative phase shifts (perceived as a longer inter-beat interval).

A strength of ERFA to be highlighted is that it provides an overarching analysis framework for signals of different nature. The temporal dynamics of rhythmic stimulation, behavioral responses and electrophysiological signals were all processed in terms of instantaneous frequency, allowing direct comparisons across different measurements. In the particular case of finger-tapping, data were processed modeling rhythmic behavior as an oscillator (Rosso et al., 2021; Heggli et al., 2019), congruently with a neural oscillatory framework. ERFAs were also comparable across perturbation types, namely tempo changes and phase shifts. Despite the discussion on different cognitive mechanisms underlying phase and tempo corrections (Repp, 2001; Repp and Keller, 2004), from a signal processing perspective they can be operationalized as the same phenomenon on different timescales, expressed in the same unit of measure. Instantaneous frequency was in fact computed entirely based on the rate of change of an oscillation's phase, and its expression in Hz units is just a matter of re-scaling (Cohen, 2014). Phase information is needed for estimating changes in frequency, and is a necessary condition for operationalizing entrainment according to its fundamental definition (Rajendran and Schnupp, 2019).

Our work was mainly focused on oscillatory dynamics recorded with EEG. Despite the constraints of the poor spatial resolution and low signal-to-noise ratio characterizing the technique, we address the importance of applying our experimental paradigm and metric in populations affected by neurological deficits, to generate testable predictions on the functional role of neuroanatomical structures compromised by

the pathology (Cannon and Patel, 2021). Furthermore, future research may deploy the paradigm with magnetoencephalography or intracranial recordings, to investigate entrained activity in cortical and subcortical anatomical structure during sensorimotor synchronization.

5. Conclusions

The major methodological contribution of our work consisted of a paradigm and a measure for investigating neural entrainment in human participants, optimized for non-invasive electrophysiological recordings. By perturbing isochronous auditory metronomes in tempo and phase during a finger-tapping task, we induced behavioral synchronization errors and showed that oscillatory neural components dynamically adjusted their frequency to stimulus changes during error correction responses. By means of spatial filters design, we were able to disentangle perceptual and sensorimotor oscillatory components from the multivariate EEG signal, revealing that active engagement of the motor system enhanced neural entrainment. This evidence, along with clues of intrinsic brain dynamics not explicable by bottom-up processing of the stimuli, strongly suggests that actual neural entrainment underlies tracking and sensorimotor synchronization to dynamic auditory rhythms. In addition to these fundamental findings, ERFA proved to be a sensitive measure of neural entrainment, reflecting an oscillatory model of brain functioning while mitigating the influence of bottom-up evoked responses.

Declaration of Competing Interest

The authors declare no competing interests.

Credit authorship contribution statement

Mattia Rosso: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Project administration, Funding acquisition. **Bart Moens:** Software, Resources. **Marc Leman:** Writing – review & editing, Supervision. **Lousin Moumdjian:** Conceptualization, Writing – review & editing, Project administration.

Data availability

Data and scripts are available upon request to the authors with a formal data sharing agreement, in line with the conditions of the local ethics committee which approved the present study.

Acknowledgments

The present study was funded by Bijzonder Onderzoeksfonds (BOF) from Ghent University (Belgium), in the context of a joint-PhD project with the University of Lille (France) (I-SITE ULNE program; grant number: 01D21819). The authors are grateful to Ivan Schepers for building the hardware of the finger-tapping device, to Canan Nuran Gener for her precious help in collecting the data, and to the three anonymous reviewers whose thorough comments considerably improved the quality of the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.120226.

References

Arnal, L.H., Giraud, A.L., 2012. Cortical oscillations and sensory predictions. Trends Cogn. Sci. 16, 390–398.
Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68, 255–278.
Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. arXiv [stat.CO] (2014).
Bavassi, L., Kamienkowski, J.E., Sigman, M., Laje, R., 2017. Sensorimotor synchronization: neurophysiological markers of the asynchrony in a finger-tapping task. Psychol. Res. 81, 143–156.
Bengtsson, S.L., et al., 2009. Listening to rhythms activates motor and premotor cortices. Cortex 45, 62–71.
Boashash, B., 1992. Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals. In: Proceedings of the IEEE.
Breska, A., Deouell, L.Y., 2017. Neural mechanisms of rhythm-based temporal prediction: delta phase-locking reflects temporal predictability but not rhythmic entrainment. PLoS Biol. 15, e2001665.
Buzsáki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. Science 304, 1926–1929.
Cadena-Valencia, J., García-Garibay, O., Merchant, H., Jazayeri, M., de Lafuente, V., 2018. Entrainment and maintenance of an internal metronome in supplementary motor area. Elife 7.
Cannon, J.J., Patel, A.D., 2021. How beat perception Co-opts motor neurophysiology. Trends Cogn. Sci. 25, 137–150.
Chemin, B., Mouraux, A., Nozaradan, S., 2014. Body movement selectively shapes the neural representation of musical rhythms. Psychol. Sci. 25, 2147–2159.
Chen, J.L., Penhune, V.B., Zatorre, R.J., 2008. Listening to musical rhythms recruits motor regions of the brain. Cereb. Cortex 18, 2844–2854.
Cheng, T.H.Z., Creel, S.C., Iversen, J.R., 2022. How do you feel the rhythm: dynamic motor-auditory interactions are involved in the imagination of hierarchical timing. J. Neurosci. 42, 500–512.
Clayton, J., Eerola, 2020. Interpersonal entrainment in music performance: theory, method, and model. Music Percept..
Cohen, M.X., Gulbinaite, R., 2017. Rhythmic entrainment source separation: optimizing analyses of neural responses to rhythmic sensory stimulation. Neuroimage 147, 43–56.
Cohen, E.E.A., Ejsmond-Frey, R., Knight, N., Dunbar, R.I.M, 2010. Rowers' high: behavioural synchrony is correlated with elevated pain thresholds. Biol. Lett. 6, 106–108.
Cohen, M.X., 2014. Fluctuations in oscillation frequency control spike timing and coordinate neural networks. J. Neurosci. 34, 8988–8998.
Cohen, M.X., 2022. A tutorial on generalized eigendecomposition for denoising, contrast enhancement, and dimension reduction in multichannel electrophysiology. Neuroimage 247, 118809.
Donoghue, T., et al., 2020. Parameterizing neural power spectra into periodic and aperiodic components. Nat. Neurosci. 23, 1655–1665.
Fries, P., 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn. Sci. 9, 474–480.
Gámez, J., Mendoza, G., Prado, L., Betancourt, A., Merchant, H., 2019. The amplitude in periodic neural state trajectories underlies the tempo of rhythmic tapping. PLoS Biol. 17, e3000054.
Grahn, J.A., Brett, M., 2007. Rhythm and beat perception in motor areas of the brain. J. Cogn. Neurosci. 19, 893–906.
Haegens, S., Zion Golumbic, E., 2018. Rhythmic facilitation of sensory processing: a critical review. Neurosci. Biobehav. Rev. 86, 150–165.
Haufe, S., Dähne, S., Nikulin, V.V., 2014. Dimensionality reduction for the analysis of brain oscillations. Neuroimage 101, 583–597.
Heggli, O.A., Cabral, J., Konvalinka, I., Vuust, P., Kringelbach, M.L, 2019. A Kuramoto model of self-other integration across interpersonal synchronization strategies. PLoS Comput. Biol. 15, e1007422.
Jantzen, K.J., Ratcliff, B.R., Jantzen, M.G, 2018. Cortical networks for correcting errors in sensorimotor synchronization depend on the direction of asynchrony. J. Mot. Behav. 50, 235–248.
Kliger Amrani, A., Zion Golumbic, E., 2022. Memory-paced tapping to auditory rhythms: effects of rate, speech, and motor engagement. J. Speech Lang. Hear. Res. 65, 923–939.
Lakatos, P., et al., 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J. Neurophysiol. 94, 1904–1911.
Lakatos, P., Chen, C.M., O'Connell, M.N., Mills, A., Schroeder, C.E, 2007. Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron 53, 279–292.
Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 320, 110–113.
Lakatos, P., et al., 2009. The leading sense: supramodal control of neurophysiological context by attention. Neuron 64, 419–430.
Lakatos, P., et al., 2013. The spectrotemporal filter mechanism of auditory selective attention. Neuron 77, 750–761.
Lakatos, P., et al., 2016. Global dynamics of selective attention and its lapses in primary auditory cortex. Nat. Neurosci. 19, 1707–1717.
Lakatos, P., Gross, J., Thut, G., 2019. A new unifying account of the roles of neuronal entrainment. Curr. Biol. 29, R890–R905.
Legrain, V., Iannetti, G.D., Plaghki, L., Mouraux, A., 2011. The pain matrix reloaded: a salience detection system for the body. Prog. Neurobiol. 93, 111–124.
Leman, M., 2016. The Expressive Moment: How Interaction (With Music) Shapes Human Empowerment. MIT Press.
Lenc, T., Keller, P.E., Varlet, M., Nozaradan, S., 2018. Neural tracking of the musical beat is enhanced by low-frequency sounds. Proc. Natl. Acad. Sci. U. S. A. 115, 8221–8226.
London, J. Hearing in time: psychological aspects of musical meter. (2012).
Luck, S.J., 2014. An Introduction to the Event-Related Potential Technique, 2nd ed. MIT Press.
Mirman, D., 2017. Growth Curve Analysis and Visualization Using R. CRC Press.
Morillon, B., Baillet, S., 2017. Motor origin of temporal predictions in auditory attention. Proc. Natl. Acad. Sci. U. S. A. 114, E8913–E8921.
Morillon, B., Schroeder, C.E., Wyart, V., 2014. Motor contributions to the temporal precision of auditory attention. Nat. Commun. 5, 5255.

Morillon, B., Hackett, T.A., Kajikawa, Y., Schroeder, C.E, 2015. Predictive motor control of sensory dynamics in auditory active sensing. Curr. Opin. Neurobiol. 31, 230–238.
Morillon, B., Arnal, L.H., Schroeder, C.E., Keitel, A., 2019. Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. Neurosci. Biobehav. Rev. 107, 136–142.
Moumdjian, L., et al., 2022. A case-study of a person with multiple sclerosis and cerebellar ataxia synchronizing finger-taps and foot-steps to music and metronomes. Neuroimmunol. Rep. 2, 100101.
Nobre, A.C., van Ede, F., 2018. Anticipated moments: temporal structure in attention. Nat. Rev. Neurosci. 19, 34–48.
Novembre, G., Iannetti, G.D., 2018. Tagging the musical beat: neural entrainment or event-related potentials? Proc. Natl. Acad. Sci. U.S.A. 115, E11002–E11003.
Nozaradan, S., Peretz, I., Missal, M., Mouraux, A., 2011. Tagging the neuronal entrainment to beat and meter. J. Neurosci. 31, 10234–10240.
Nozaradan, S., Peretz, I., Mouraux, A., 2012. Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. J. Neurosci. 32, 17572–17581.
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.
Nozaradan, S., Schwartze, M., Obermeier, C., Kotz, S.A., 2017. Specific contributions of basal ganglia and cerebellum to the neural tracking of rhythm. Cortex 95, 156–168.
Obleser, J., Kayser, C., 2019. Neural entrainment and attentional selection in the listening brain. Trends Cogn. Sci. 23, 913–926.
Obleser, J., Henry, M.J., Lakatos, P., 2017. What do we talk about when we talk about rhythm? PLoS Biol. 15, e2002794.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 156869.
Park, H., Ince, R.A.A., Schyns, P.G., Thut, G., Gross, J., 2015. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Curr. Biol. 25, 1649–1653.
Praamstra, P., Turgeon, M., Hesse, C.W., Wing, A.M., Perryer, L., 2003. Neurophysiological correlates of error correction in sensorimotor-synchronization. Neuroimage 20, 1283–1297.
Rajendran, V.G., Schnupp, J.W.H, 2019. Frequency tagging cannot measure neural tracking of beat or meter. Proc. Natl. Acad. Sci. U.S.A. 116, 2779–2780.
Repp, B.H., Keller, P.E., 2004. Adaptation to tempo changes in sensorimotor synchronization: effects of intention, attention, and awareness. Q. J. Exp. Psychol. A 57, 499–521.
Repp, B.H., Su, Y.H., 2013. Sensorimotor synchronization: a review of recent research (2006–2012). Psychon. Bull. Rev 20, 403–452.
Repp, B.H., 2001a. Processes underlying adaptation to tempo changes in sensorimotor synchronization. Hum. Mov. Sci. 20, 277–312.
Repp, B.H., 2001b. Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. J. Exp. Psychol. Hum. Percept. Perform. 27, 600–621.
Richardson, D., Shockley, 2008. Synchrony and swing in conversation: coordination, temporal dynamics, and communication. Embodied Commun..
Rimmele, J.M., Morillon, B., Poeppel, D., Arnal, L.H, 2018. Proactive sensing of periodic and aperiodic auditory patterns. Trends Cogn. Sci. 22, 870–882.
Rosenblum, M., Pikovsky, A., Kurths, J., Schäfer, C. & Tass, P.A. Chapter 9 phase synchronization: from theory to data analysis. in Handbook of Biological Physics (eds. Moss, F. & Gielen, S.) vol. 4 279–321 (North-Holland, 2001).
Rosso, M., Maes, P.J., Leman, M., 2021a. Modality-specific attractor dynamics in dyadic entrainment. Sci. Rep. 11, 18355.
Rosso, M., Leman, M., Moumdjian, L., 2021b. Neural entrainment meets behavior: the stability index as a neural outcome measure of auditory-motor coupling. Front. Hum. Neurosci. 15, 668918.
Rosso, M., Heggli, O.A., Maes, P.J., Vuust, P., Leman, M., 2022. Mutual beta power modulation in dyadic entrainment. Neuroimage 257, 119326.
Samaha, J., Cohen, M.X., 2022. Power spectrum slope confounds estimation of instantaneous oscillatory frequency. Neuroimage 250, 118929.
Schroeder, C.E., Lakatos, P., 2009. Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. 32, 9–18.
Schroeder, C.E., Wilson, D.A., Radman, T., Scharfman, H., Lakatos, P., 2010. Dynamics of active sensing and perceptual selection. Curr. Opin. Neurobiol. 20, 172–176.
Schwartze, M., Keller, P.E., Kotz, S.A., 2016. Spontaneous, synchronized, and corrective timing behavior in cerebellar lesion patients. Behav. Brain Res. 312, 285–293.
Shockley, K., Santana, M.V., Fowler, C.A, 2003. Mutual interpersonal postural constraints are involved in cooperative conversation. J. Exp. Psychol. Hum. Percept. Perform. 29, 326–332.
Teki, S., Grube, M., Kumar, S., Griffiths, T.D, 2011. Distinct neural substrates of duration-based and beat-based auditory timing. J. Neurosci. 31, 3805–3812.
Zalta, A., Petkoski, S., Morillon, B., 2020. Natural rhythms of periodic temporal attention. Nat. Commun. 11, 1051.
Zoefel, B., VanRullen, R., 2015. Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. J. Neurosci. 35, 1954–1964.