

## Research report

## Contingent magnetic variation and beta-band oscillations in sensorimotor temporal decision-making

Lu Guo<sup>a,b</sup>, Ming Bao<sup>c,\*</sup>, Zhifei Chen<sup>c</sup>, Lihan Chen<sup>d,e,f,g,\*\*,1</sup><sup>a</sup> The Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China<sup>b</sup> State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China<sup>c</sup> School of Materials Science and Intelligent Engineering, Nanjing University, Suzhou 215163, China<sup>d</sup> School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China<sup>e</sup> Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China<sup>f</sup> National Engineering Laboratory for Big Data Analysis and Applications, Peking University, Beijing 100871, China<sup>g</sup> State Key Laboratory of General Artificial Intelligence, Peking University, Beijing, 100871, China

## ARTICLE INFO

## Keywords:

Perceptual averaging  
 Temporal prediction  
 Auditory sequence  
 Magnetoencephalograph  
 Beta oscillations

## ABSTRACT

The ability to accurately encode the temporal information of sensory events and hence to make prompt action is fundamental to humans' prompt behavioral decision-making. Here we examined the ability of ensemble coding (averaging multiple inter-intervals in a sound sequence) and subsequent immediate reproduction of target duration at half, equal, or double that of the perceived mean interval in a sensorimotor loop. With magnetoencephalography (MEG), we found that the contingent magnetic variation (CMV) in the central scalp varied as a function of the averaging tasks, with a faster rate for buildup amplitudes and shorter peak latencies in the "half" condition as compared to the "double" condition. ERD (event-related desynchronization) -to-ERS (event-related synchronization) latency was shorter in the "half" condition. A robust beta band (15–23 Hz) power suppression and recovery between the final tone and the action of key pressing was found for time reproduction. The beta modulation depth (i.e., the ERD-to-ERS power difference) was larger in motor areas than in primary auditory areas. Moreover, results of phase slope index (PSI) indicated that beta oscillations in the left supplementary motor area (SMA) led those in the right superior temporal gyrus (STG), showing SMA to STG directionality for the processing of sequential (temporal) auditory interval information. Our findings provide the first evidence to show that CMV and beta oscillations predict the coupling between perception and action in time averaging.

## 1. Introduction

Predictive timing is the phenomenon by which people can effectively anticipate future events based on the perception of temporal regularities or associative contingencies between ongoing sensory events (Arnal and Giraud, 2012; Nobre et al., 2007). This remarkable ability is indispensable for human adaptive behavior, in that it enables us to proactively plan our actions and promptly make behavioral decisions (Acerbi et al., 2012; Nobre and Van, 2017; Rohenkohl et al., 2012; Zhao et al., 2019). Evidence of predictive timing has been shown across a wide range of circumstances, such as timing of the intervals between syllables and words and controlling movements in time (e.g., Arnal et al., 2015;

Breska and Deouell, 2017; Cope et al., 2012; Morillon et al., 2016). This typical ability of temporal prediction has been recently implemented in ecologically valid scenarios, including crossmodal interactions such as audiovisual integration with prompt perceptual decision making (Zeng and Chen, 2019; Chen et al., 2018). The ensemble coding of sensory properties has been recently summarized in the framework of (temporally) attentional selection (Obleser and Kayser 2019; Lakatos et al., 2008). However, how the extracted mean interval could be reproduced in a sensorimotor loop remains largely known, and the underlying neurocognitive mechanism is far from clear (Tark et al., 2021).

In temporal prediction, the contingent negative variation (CNV), a slow negative deflection in the frontal-central scalp

\* Corresponding author.

\*\* Corresponding author at: School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China.

E-mail addresses: [baomingnju@nju.edu.cn](mailto:baomingnju@nju.edu.cn) (M. Bao), [clh@pku.edu.cn](mailto:clh@pku.edu.cn) (L. Chen).<sup>1</sup> <https://orcid.org/0000-0002-0337-4177><https://doi.org/10.1016/j.brainresbull.2024.111021>

Received 5 February 2024; Received in revised form 7 June 2024; Accepted 21 June 2024

Available online 26 June 2024

0361-9230/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

electroencephalogram (EEG), has been revealed to be strongly associated with the temporal predictability of incoming events and the concomitant preparation of responses, hence providing us an index to measure the neuronal response to (predictive) timing behavior (Macar and Vidal, 2004; Nobre et al., 2007; Ruchkin et al., 1977; Walter et al., 1964). For example, faster CNV buildup and longer CNV latencies are correlated with shorter expected interval durations (Breska and Deouell, 2017; Macar et al., 1999; Praamstra et al., 2006; Ruchkin et al., 1977) or post-adaptation of short intervals (Li et al., 2021). In terms of cortical oscillations, beta-band power dynamics (13–30 Hz) have been proposed as predictive neural correlates in time estimation (Bartolo and Merchant, 2015; Fujioka et al., 2012; Iversen et al., 2009), during auditory rhythm perception (Fujioka et al., 2015), and in motor timing of motor-related planning and execution (Baker, 2007; Donner et al., 2009; Jasper and Penfield, 1949; Pfurtscheller et al., 2005). Using transcranial alternating current stimulation (tACS), Wiener et al. (2018) showed that beta stimulation contributed to prolonged perceived visual duration, suggesting that beta oscillations are intrinsically involved in the perception of time and are particularly associated with the retention and comparison of a memory standard for the duration (Wiener et al., 2018). A previous study has adopted a delayed-target detection task in which subjects were required to detect whether the last tone of an isochronous sequence was delayed with regard to the beat with MEG recordings, and found that delta-beta coupled oscillations underpin prediction accuracy (Arnal et al., 2015). Therefore, beta oscillations have been reported to feature the sensory interval timing as well as action timing, which suggests that in a given sensorimotor event for prompt timing decision, we'll expect to observe the commonality of beta oscillations interfacing the sensation and action, based on the predictive timing of the preceding sensory events (typically for sound sequence).

The present study exploited the sensorimotor paradigm and went further to examine whether/how humans could encode accurately the multiples of the averaging interval in both isochronous and anisochronous sequences when they were about to make decisions for target time intervals. Here, we used a sensorimotor paradigm in which human observers produced a probe interval as multiples (half, equal, or double) of the mean interval in a preceding sound sequence. We combined MEG recordings with behavioral psychophysics to investigate the characteristics of the CMV (the magnetic counterpart of the CNV component) and the beta oscillations underlying ensemble (en)coded interval timing. We further analyzed the functional connectivity between critical regions of the sensorimotor and auditory cortices to show the driving direction for predictive (auditory) timing and behavioral selection. The results showed that the CMV varied as a function of the averaging tasks in terms of the rate for buildup amplitudes and peak latencies. A robust beta band (15–23 Hz) power suppression and recovery were found across the time window between the final marker tone and the key pressing for duration reproduction. Moreover, there is a functionally link between sensory encoding and motor activation (Abbasi and Gross, 2020; Morillon and Baillet, 2017), the beta modulation depth (i.e., event-related desynchronization to-event-related synchronization power difference) was larger in motor areas than that in primary auditory areas, and phase slope index (PSI) analysis indicated that left Supplementary motor area (SMA) activities lead the brain (beta) oscillations in right Superior temporal gyrus (STG) activities, it provides evidence of the SMA to STG directionality during the processing of sequential (temporal) auditory interval information.

## 2. Materials and methods

### 2.1. Participants

A total of 20 healthy undergraduate and graduate students (9 female; mean age =  $22 \pm 2.77$  SD) participated in the experiment after providing written informed consent. They were all right-handed, with normal or corrected-to-normal vision and normal hearing. None of them

had a history of psychological or neurological disorders. Data from two participants were excluded from further analysis because of excessive artifact movements during the experiments. Thus, the final analyzed dataset included 18 participants (9 female; mean age =  $22.42 \pm 2.71$  SD). The experiment was implemented in compliance with all institutional guidelines set by the Academic Affairs Committee, School of Psychological and Cognitive Sciences at Peking University.

### 2.2. Stimuli and apparatus

#### 2.2.1. Stimuli

Visual stimuli were used as instruction cues for the corresponding tasks. The visual cue was composed of two parallel horizontal bars, presented so that the length ratio of the lower bar to the upper bar was 1:2, 1:1, or 2:1 (with labels of “1/2 T”, “1 T”, and “2 T” above the bars accordingly), to indicate the task of reproducing half, equal, or double of the mean of the preceding multiple auditory inter-intervals.

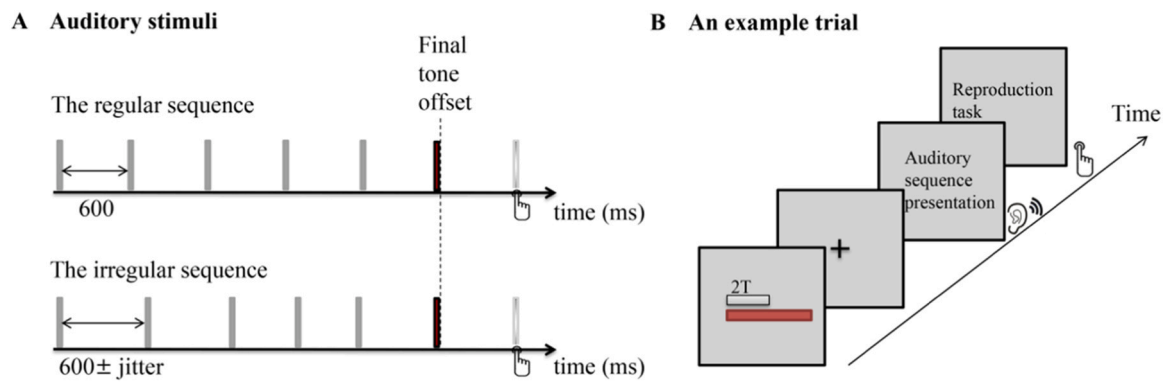
We used two types of auditory stimuli: paired tones and a sound sequence. First, a pair of pure tones (durations of 30 ms, 1000 Hz for the first tone and 500 Hz for the second) was used in a pretest. Motor functional localizer was determined in this pretest, during which participants reproduced an interval equal to the given interval (600 ms) between the two tones. Second, sound sequences, each containing four, five, or six intervals separated by 30 ms beeps, were used for the auditory functional localizer task and the main experiment. The sequences for the auditory functional localizer task always contained five intervals, while a few sequences for the main task consisted of either four or six intervals (as filler stimuli). The intervals were either fixed (600 ms, regular sequence) or variable (mean = 600 ms, between 480 ms and 720 ms, irregular sequence). The frequencies of the first five beeps were maintained at 1000 Hz, while the last beep (signaling tone) changed to 500 Hz to prompt the participants to reproduce the interval as requested (Fig. 1A).

#### 2.2.2. Apparatus and data acquisition

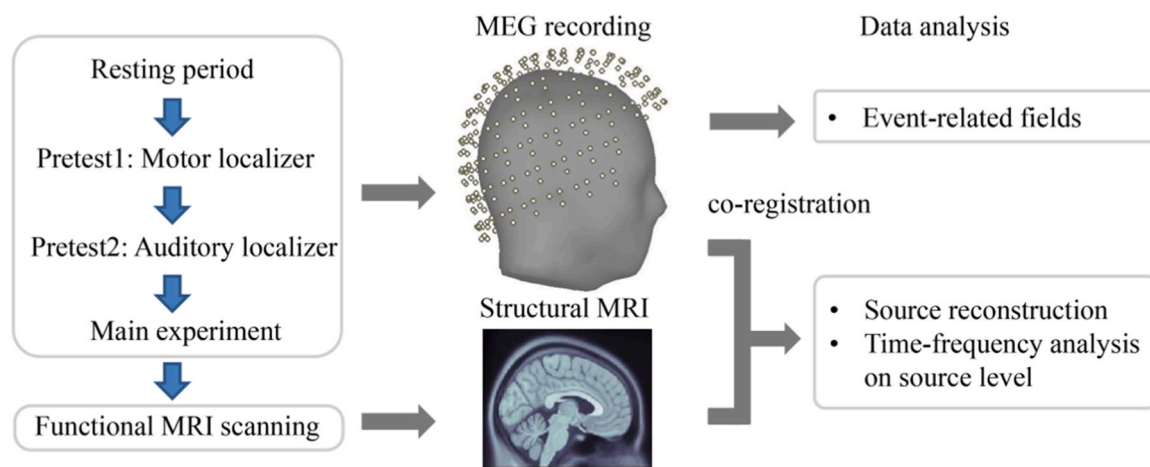
We used a sound card with high temporal precision (RME Fireface UFX) to generate the auditory stimuli (around 50 dB, adjusted individually for each participant) and a tube phone to deliver the sounds binaurally to the participants. We presented visual stimuli on a plexiglass board, relayed by an LCD screen (100 Hz refresh rate and  $1024 \times 768$  pixel resolution). The viewing distance was 60 cm. Neuro-magnetic signals were recorded with a 306-channel, whole-head MEG system (204 planar gradiometers and 102 magnetometers; Elekta-Neuromag, Helsinki, Finland) in a magnetically shielded room at Peking University. The sampling rate of the MEG signal was 1000 Hz. Maxfilter software (Elekta-Neuromag, Helsinki, Finland) with temporal signal space separation (tsss) was first used to remove external noises from the raw MEG data. After MEG recordings, participants underwent magnetic resonance imaging (MRI) scans in a Magnetom Prisma 3 T MRI scanner (Siemens Healthcare), to obtain whole-head T1-weighted structural anatomical images. The parameters for MRI were: 192 sagittal slices; Field of view (FOV) =  $256 \text{ mm} \times 256 \text{ mm}$ ; and slice thickness = 1 mm. The computer programs for controlling the experiment were developed with MATLAB (MathWorks Inc.) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

### 2.3. Design and procedures

Before the formal MEG recording, resting brain activity was recorded for 4 minutes and these baseline data were used to compute the noise covariance. Participants then completed two localizer pre-tests (Fig. 2). In the motor functional localizer task, participants completed a total of 40 trials. In a typical trial, after a fixation cross (lasting for 300–500 ms), a pair of pure tones were presented. Participants were instructed to press a button immediately and reproduce the duration equal to the given interval (600 ms) enclosed by the two tones, as soon as the second tone



**Fig. 1.** Stimuli and procedure for the main experiment. A, Schematic configuration of the auditory sequences. Regular sequences contained four to six intervals (mean duration of 600 ms), separated by auditory beeps (30 ms, 1000 Hz, but the last signaling beep was at 500 Hz). Irregular sequences also contained four to six intervals with intervals randomly distributed between 480 and 720 ms (the mean interval was maintained at 600 ms). B, An example trial. At the beginning, a visual cue was given to instruct the participants in the task condition (“half”, “equal”, and “double”) for the coming trial. After a fixation, an auditory sequence appeared. Upon the offset of the final tone, the participants reproduced a time interval according to the instructions.



**Fig. 2.** Diagram for the experimental paradigm and the MEG data analysis. We recorded 306-channel, whole-head MEG signals from participants while they were both resting and performing the task. The structural MRIs were subsequently acquired with a 3.0 T MRI scanner. After the MEG-MRI co-registration, we converted MEG sensor data to source activities, based on which time-frequency analysis on the source level and on functional whole-brain activity were implemented.

was presented. The auditory functional localizer pretest consisted of 80 trials, in which participants were required to listen to a sound sequence. Each trial started with a fixation cross appearing at the center of the screen for 300–500 ms to indicate the coming of the following sound sequence. Participants simply listened to a sequence of 6 tones with fixed or variable inter-intervals. The inter-trial interval (ITI) was 1200–1400 ms. The 80 trials were divided into 2 blocks. Participants took a short break of 1 minute between blocks.

For the main experiment, 2 (auditory sequences: regular, irregular)  $\times$  3 (tasks: “half”, “equal”, or “double”) factorial design was implemented. As shown in Fig. 1B, a typical trial started with a visual cue (2800–3000 ms) to inform participants of the upcoming timing task. After a visual fixation cross (300–500 ms) as well as a blank interval of 100 ms, a sound sequence was delivered, which contained four to six 30-ms tones (1000 Hz) but ended with a 30-ms tone (marker) of 500 Hz. Upon hearing this marker tone, participants reproduced half, equal, or double of the mean intervals in the preceding sequence, by pressing a button (“8” on the keyboard) with their right index finger to demarcate this interval.

Each participant completed 5 acquisition runs during the MEG recordings. Each experimental condition had 12 trials in succession in each run: 10 trials with an auditory sequence consisting of 6 sound beeps, and the remaining 2 catch trials with jitter stimuli (5 or 7 beeps). The presentation order of the experimental conditions within each run

was randomized. In total, there were 360 trials, consisting of 60 repetitions for each of the 6 conditions. After each acquisition run, participants took 2–3 minutes to rest but remained stationary in their seat. They were reminded by the experimenter to maintain their body posture during this break, as well as with the feedback of the online video monitoring in the operator room.

Individual MRIs were acquired after the MEG recording session, which were used to co-register with the functional data for the purposes of following source reconstruction and localization.

#### 2.4. Data processing

##### 2.4.1. Analysis methods

We applied a repeated measure of analysis of variances (ANOVA) for the data that followed the normal distributions and applied Non parametric analysis of Friedman’s Test/Two Way Analysis of Variance by ranks for those violated the normal distributions (see 3.3.3 Timing of beta desynchronization/synchronization and beta modulation depth). The analysis was implemented with MATLAB version: 9.13.0 (R2022b) (Natick, Massachusetts: The MathWorks Inc) and SPSS26.0 (Armonk, NY: IBM Corp).

##### 2.4.2. Behavioral data analysis

We screened the data by removing outliers according to the

following criteria. First, trials containing MEG artifacts were discarded. Then, in the dataset for reproducing durations in the “half” task condition (with the physical reference value of 300 ms), any data points shorter than 150 ms or longer than 800 ms were removed. For the “equal” (600 ms) condition, the dataset of reproduced durations shorter than 150 ms or longer than 1000 ms were removed. For the “double” (1200 ms) condition, the dataset of reproduced durations shorter than 150 ms or longer than 1600 ms were removed. Finally, for each participant, the dataset in which the reproduced intervals deviated beyond 3 standard deviations from the mean of the corresponding conditions was also discarded.

We applied a repeated measure of analysis of variances (ANOVA) for the remaining datasets, with regularities (regular, irregular) of the sound sequence and tasks (“half”, “equal”, and “double”) as within-participants factors. We then conducted Bonferroni-corrected post hoc analysis to evaluate any significant differences between the task conditions.

#### 2.4.3. MEG preprocessing and analysis

The analysis of MEG data was performed using the *Brainstorm* software (Tadel et al., 2011) and in-house Matlab codes according to the guidelines (Gross et al., 2013), as well as Fieldtrip (Oostenveld et al., 2011).

**2.4.3.1. MEG-MRI co-registration.** With *Brainstorm*, we co-registered the MRI image of each participant to the MEG coordinate system in two steps. First, the initial registration was based on the three fiducial references that defined the subject coordinate system (SCS): nasion and bilateral pre-auricular points. These three anatomical landmarks were manually identified in the individual’s MRI and then pair-matched with the same reference points as measured during the MEG acquisition. Second, to improve the registration, the locations of additional scalp points were acquired during the MEG session, using a 3D digitizer device (Polhemus, Fastrak system). This additional alignment was run for each participant, based on an iterated closest point algorithm using *Brainstorm*.

After making sure that the sensors were properly aligned with the MRI of each participant, we used the preprocessed MEG signals for further analysis at two levels: the magnetic field distribution measured at the sensor surface (*sensor level*) and the estimated current sources that underlay the recorded magnetic fields (*source level*).

**2.4.3.2. Analysis of event-related fields (*sensor level*).** Segments containing artifacts (e.g., subjective movements, muscle noise) were first discarded. Calculations of the signal space projection (SSP, a common approach based on the spatial decomposition of MEG/EEG recordings for eliminating reproducible artifacts) for eye blinks were employed over the data from all 306 channels (Uusitalo and Ilmoniemi, 1997). The first SSP component, which correlated relatively well with electro-oculographic (EOG) trace, was chosen to be removed from the data. The event-related fields (ERF) analysis was time-locked to the onset of the final tone of the sequence (3180-ms duration of the auditory sequence starting from the onset of the first tone to the offset of the final tone). The epochs for analysis were of 6400 ms duration, including a 4000 ms pre-onset period. The data were baseline-corrected to the  $-4000$  ms to  $-3150$  ms pre-onset time window (immediately before the onset of the whole sound sequence; no sounds were present in this time range) and low-pass filtered at 60 Hz.

The ERF analysis was conducted in a predefined central cluster of sensors (gradiometers selected for less “brain noise”) from *Brainstorm*: MEG0723, MEG0732, MEG0742, MEG1112, MEG1123, MEG1133, MEG1142, MEG1833, MEG2212, MEG2223, MEG2232, MEG2243, and MEG2443, which covered the main sensors that reflected the modulations of the CMV potentials. We tested two a priori hypotheses regarding the CMV in the central cluster. First, we examined whether the CMV

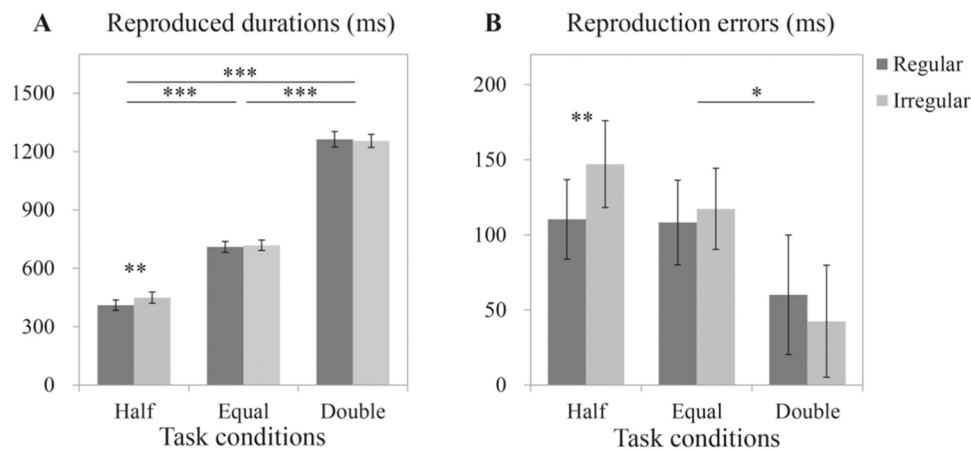
buildup was modulated by the expected interval duration, i.e., faster buildup when the expected duration was shorter (Breska and Deouell, 2017; Miniussi et al., 1999; Praamstra et al., 2006). Second, we compared the peak latency of the CMV between different conditions. We also analyzed the early ERF components, such as N1m and P2m, evoked by the final beep in the sequence.

To visualize neuromagnetic responses in this cluster, epochs were averaged across repeated trials within each experimental condition for each participant. The final tone evoked ERFs with clearly identifiable N1m, P2m, and contingent magnetic variation components (CMV) from the central cluster (Fig. 3A). Based on an unbiased visual inspection of the averaged waveforms, the N1m and P2m peak latencies were estimated when the amplitude reached the minima in the 80–130 ms time segment and the maxima in the 150–240 ms time segment after the onset of the final tone, respectively. For the CMV peak latency, we calculated mean amplitudes over 100 ms sliding windows with a midpoint ranging from 200 to 600 ms. The latency was determined by the midpoint of the sliding window with the maximum amplitude (Li et al., 2017; Pfeuty et al., 2003). In addition, the CMV buildup amplitude was defined as the mean value in the 200–320 ms time segment. Within each experimental condition, we conducted a two-way repeated-measures ANOVA on the amplitudes and latencies of the N1m, P2m, and CMV components, with regularities (regular and irregular) and tasks (“half”, “equal”, and “double”) as within-participants factors.

**2.4.3.3. Source reconstruction.** The spatiotemporal dynamics of the cortical sources underlying the measured magnetic field distributions were determined from the motor localizer pretest, the auditory localizer pretest, and the main experiment. With *Brainstorm*, we imaged the foci of activations that were time-locked to the onset of the final tone using the depth-weighted minimum L2 norm estimator of cortical current density (Hämäläinen and Ilmoniemi, 1994) and the unconstrained source model (three values for each grid point), which could subsequently be normalized with the Z-transformation to obtain the within-participants averages in different conditions. In order to get one statistic and one p-value per grid point in the output, a flattening step was used in which the operator  $|A|$  was interpreted as the norm of the three orientations, i.e.  $|A| = \sqrt{A_x^2 + A_y^2 + A_z^2}$ . Then individual source maps were projected onto the ICBM-152 brain template to compute the grand averages of all participants, with the source locations transformed according to the Montreal Neurological Institute (MNI) coordinate system.

With the auditory motor timing paradigm, we explored the temporal dynamics in the auditory and motor areas. We identified regions of interest (ROIs) in the superior temporal gyrus that responded (bilaterally) to the auditory localizer task and in the left supplementary motor area (see Table 1) that responded to the motor localizer task. Those anatomical areas were used for the analysis of oscillatory neural activities.

**2.4.3.4. Time-frequency analysis on source level.** Time-frequency analysis was conducted for each epoch ( $-4400$ – $2400$  ms) of source activities using the complex Morlet wavelet transform, resulting in an estimate of oscillatory power (squared absolute value) at each time sample and between 1 and 40 Hz with 1 Hz resolution. We calculated the averaged response across repeated trials within each experimental condition for each participant. Mean power was normalized according to the mean over the baseline ( $-4000$  to  $-3150$  ms immediately before the sequence onset). Based on the power change ratio, we constructed time-frequency maps of event-related desynchronization and synchronization (ERD/ERS) for each condition for each participant, and we compared the ERD/ERS maps between conditions using a non-parametric paired t-test approach by applying Monte Carlo permutation statistics (5000 randomizations) (Maris and Oostenveld, 2007) with false discovery rate (FDR) correction for multiple comparisons (Genovese et al., 2002).



**Fig. 3.** Group means of the reproduced duration and the reproduced error with task conditions (“half”, “equal”, and “double”) and auditory sequence regularity (regular and irregular) as within-participant factors. All the error bars show the associated standard errors of the means (\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ).

**Table 1**

Summary of ROIs. The table shows the coordinates of the seed, the number of vertices, and the specific areas in each region.

Brain area	Hemisphere	Center MNI coordinates	Size	Area
Temporal lobe, superior temporal gyrus	Left	(-70.6, -20.9, 5.7)	80	16.99 cm <sup>2</sup>
Temporal lobe, superior temporal gyrus	Right	(-1.1, -21.8, 22.9)	80	17.32 cm <sup>2</sup>
Parietal lobe, supplementary motor area	Left	(-7.3, -8.8, 59.1)	80	13.71 cm <sup>2</sup>

ROI: Region of Interest; MNI: Montreal Neurological Institute

For the ROI-based time course abstraction of the beta oscillation (15–23 Hz), we performed statistical analysis of the following four measures using a two-way ANOVA with within-participants factors of tasks and ROIs. The former two were the latencies of ERD and ERS maxima indicating the onset and offset of beta modulation. Specifically, we calculated mean power change ratios over 50 ms sliding windows with a midpoint ranging from 100 to 650 ms after the onset of the final tone. The latency of ERD maxima was determined by the time point of the sliding window’s midpoint with minimal amplitude. Latency of ERS maxima was defined as the time point with the highest beta power change ratio in the interval starting from the latency of ERD maxima to 2000 ms after the onset of the final tone. The latter two measures were the ERD-to-ERS latency, which combined the latencies of the ERD and ERS maxima into a single peak to peak expression to reveal the dynamic timing of beta modulation, and power differences, which reflected the modulation depth during beta desynchronization and synchronization.

**2.4.3.5. Functional connectivity analysis.** We used the fieldtrip toolbox for the functional connectivity analysis (Oostenveld et al., 2011). We focused our analyses on the 200 ms pre- to 2200 ms post-the onset of the final tone as the target time window. To further examine the directionality of the auditory-sensorimotor circuit, we used Phase slope index as an effective index to show the granger causality measurement between ROI. The PSI (Nolte et al., 2008) is a measure of directionality of coupling between brain areas based on the spectral properties of electrophysiological data. Specifically, it estimates which is the leading source in a pair by relying on the sign of the discrete frequency derivative of the phase difference between two signals. As an index of dominant unidirectional interaction, PSI indicates the direction of coupling between two systems. PSI quantifies difference as a function of frequency, with a positive phase slope indicating that the signal from the first structure is leading the signal from the second structure.

### 3. Results

#### 3.1. Behavioral results

For the reproduced durations, a  $3 \times 2$  ANOVAs with the task (“half”, “equal”, and “double”) and auditory sequence (regular, irregular stimuli) as within-subject factors indicated a significant interaction effect between auditory sequences and tasks [ $F(2,34) = 4.074, p < 0.05, \eta_p^2 = 0.193$ ], and a significant main effect of tasks [ $F(2,34) = 515.621, p < 0.001, \eta_p^2 = 0.968$ ] (Fig. 3A). Bonferroni corrected comparisons revealed significant differences in the mean reproduced durations between the “half” ( $430.1 \pm 27.1$  ms), “equal” ( $713.9 \pm 26.2$  ms), and “double” ( $1258.8 \pm 35.9$  ms) conditions,  $ps < 0.001$ . The main effect of sequence regularities was not significant,  $F(1,17) = 1.678, p = 0.212, \eta_p^2 = 0.090$ . Furthermore, simple main effects analysis suggested that the regularity of the sound sequences affected the mean interval reproduction solely in the “half” condition,  $F(1,17) = 14.66, p < 0.001$ , in which the reproduced duration for irregular sequences ( $449.1 \pm 28.7$  ms) was longer than the one for regular sequences ( $411.0 \pm 26.4$  ms). Overall, participants performed well in the temporal averaging tasks with fairly well-estimated durations for “half”, “equal”, and “double” trials. The pattern of results indicated a general overestimation of the mean interval of the sequence in different experimental tasks. This over-estimation could be caused by a general delay in motor responses.

An analogous ANOVA performed on the reproduction errors (defined as the difference between the produced interval and the target interval) also revealed a significant main effect of task conditions,  $F(2,34) = 4.599, p < 0.05, \eta_p^2 = 0.213$  (Fig. 3B). Bonferroni corrected comparisons indicated shorter reproduced errors in the “double” task condition ( $51.3 \pm 37.9$  ms) than in the “equal” condition ( $112.7 \pm 26.5$  ms), with  $ps < 0.05$ . The main effect of auditory regularity was not significant,  $F(1,17) = 0.894, p = 0.358, \eta_p^2 = 0.050$ . The interaction effect between type of sequences and task conditions was significant,  $F(2,34) = 5.536, p < 0.01, \eta_p^2 = 0.246$ . Simple main effects suggested that the regularity of sound sequences affected the reproduction error in the “half” condition,  $F(1,17) = 12.17, p < 0.01$ , with larger errors for irregular sequences ( $147.1 \pm 28.8$  ms) but a smaller reproduced error for regular sequences ( $110.3 \pm 26.5$  ms). Taken together, these results indicate that the over-estimation of to-be-reproduced intervals occurred for different tasks but in varying degrees. The interval reproduction was more accurate in the “double” condition and less accurate in the “half” condition.

Furthermore, we analyzed the behavioral data for the motor localizer

pretest, in which participants reproduced the interval (a fixed 600-ms duration) between two tones. Using paired-samples t-tests, we compared the mean reproduced durations and the mean reproduction errors between this pretest and the “equal” condition in the main experiment. The difference in the mean reproduced durations ( $732.7 \pm 20.3$  ms in the motor pretest and  $713.9 \pm 26.2$  ms in “equal” trials) was not significant [ $t(1,17) = 0.932, p = 0.364$ ], nor was the difference in the mean reproduction errors ( $128.2 \pm 19.6$  ms in the motor pretest and  $112.7 \pm 26.5$  ms in “equal” trials) [ $t(1,17) = 0.729, p = 0.476$ ]. Therefore, despite the unequal demands for these two tasks (a single interval reproduction task vs. multiple intervals averaging task), participants displayed similar overestimation performance.

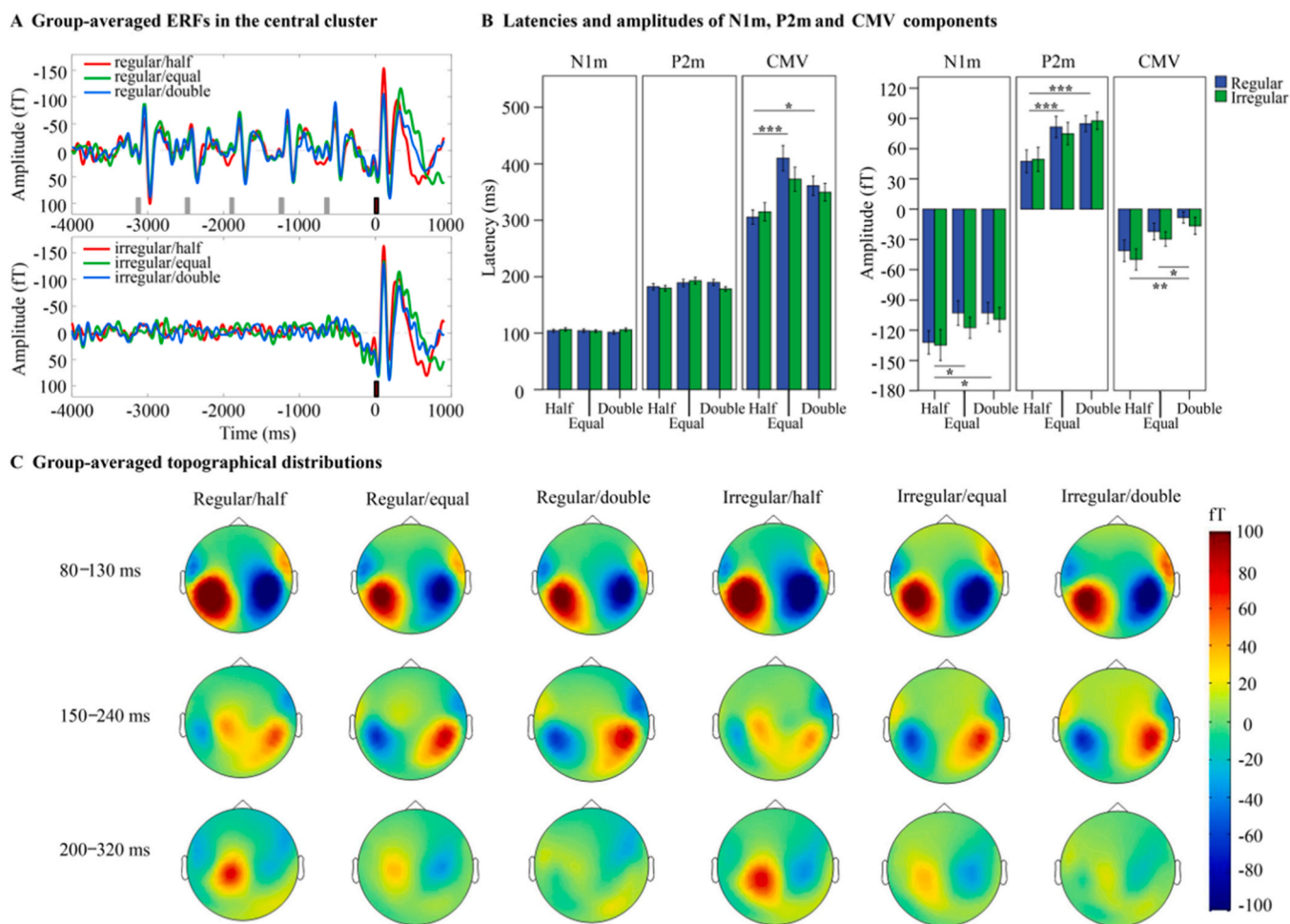
### 3.2. Analysis of event-related fields

Grand averaged ERFs during tone sequence presentation under different experimental conditions are presented in Fig. 4A. We observed periodic auditory responses with the unfolding of the regular sequence (Fig. 4A, top row), in sharp contrast to the absence of typical auditory responses prior to the final tone in the irregular sequence (Fig. 4A, bottom row) due to averaging across the temporally jittered stimuli. Furthermore, we observed a typical CMV component after the final tone for both the regular and irregular sequences, implying that the CMV was not driven automatically by the rhythm but was rather evoked by preparing for the timing of the expected action, when participants were

exposed to the auditory streams and used the temporal structure for time prediction and estimation.

#### 3.2.1. CMV component

We compared two major parameters of the CMV across all experimental conditions: CMV buildup amplitude and peak latency. First, repeated-measures ANOVA on the CMV buildup amplitude revealed a significant main task effect,  $F(2,34) = 7.774, p < 0.01, \eta_p^2 = 0.314$  (Fig. 4B). Bonferroni corrected comparisons indicated significant differences in the buildup amplitude of the CMV between “half” trials ( $-45.5 \pm 10.0$  fT) and “double” trials ( $-12.3 \pm 6.2$  fT) ( $ps < 0.01$ ) and between “equal” trials ( $-25.8 \pm 7.0$  fT) and “double” trials ( $ps < 0.05$ ). However, no further main effect or interaction was obtained, with all  $F_s < 2.4, ps > 0.15$ . The results revealed shorter peak latencies of the CMV under the “half” condition ( $310.3 \pm 13.1$  ms) than under the “equal” condition ( $391.3 \pm 15.4$  ms,  $ps < 0.001$ ) and the “double” condition ( $355.2 \pm 12.3$  ms,  $ps < 0.05$ ) [ $F(2,34) = 10.993, p < 0.001, \eta_p^2 = 0.393$ ] (Fig. 4B), but no further main effect or interaction was detected, with all  $F_s < 1.15, ps > 0.3$ . In this context, the CMV buildup amplitudes and peak latencies were both modulated by the task demands. This modulation resulted in a shorter peak latency when participants performed a “half” averaging task, as well as a slower rate of CMV buildup that was characterized by less negative waveforms during the “double” trials.



**Fig. 4.** A, Grand averaged ERFs during tone sequence presentation under different experimental conditions in the central cluster (see the colored lines), with time 0 locked to the final tone. Grey bars: tone stimuli (prior to the final tone), red bars: the final tone in the sequence. B, Latencies and amplitudes of the N1m, P2m, and CMV components under different conditions, collapsed over all participants (see the colored bars). Error bars represent standard errors under each condition (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ). C, Grand averaged scalp topographies for all measurement windows under different conditions.

### 3.2.2. N1m and P2m components

As shown in Fig. 4B, ANOVA of the N1m and P2m amplitudes both revealed a significant main effect of tasks, [ $F(2,34) = 6.705, p < 0.01, \eta_p^2 = 0.283$ ] and [ $F(2,34) = 21.397, p < 0.001, \eta_p^2 = 0.557$ ], respectively. Bonferroni corrected comparison confirmed that the N1m amplitude under the “half” condition ( $-133.4 \pm 12.6$  fT) was more negative than that under the “equal” ( $-110.2 \pm 10.8$  fT) and “double” conditions ( $-106.1 \pm 10.7$  fT), with both  $ps < 0.05$ . The P2m amplitude under the “half” condition ( $48.5 \pm 10.8$  fT) was smaller than that under the “equal” ( $78.1 \pm 10.4$  fT) and “double” conditions ( $86.1 \pm 7.4$  fT), with both  $ps < 0.001$  (Table 2). No other statistically positive results or other significant differences were found in the analysis.

Considering that both N1m and P2m partially represent an exogenous response to the sound such that the P2m co-varies with N1m along many stimulus dimensions (Crowley and Colrain, 2004), to eliminate the differential effects of the stimulus and

participant-related variables on each individual component, we calculated the N1m/P2m peak-to-peak amplitudes for each experimental condition and each participant, and conducted an analogous ANOVA on these data. However, there was no significant main effect of sequence regularities [ $F(1,17) = 2.345, p = 0.144, \eta_p^2 = 0.121$ ] as well as task conditions [ $F(2,34) = 0.772, p = 0.470, \eta_p^2 = 0.043$ ], and no significant interaction between these two factors [ $F(2,34) = 0.038, p = 0.963, \eta_p^2 = 0.002$ ]. Therefore, the peak amplitudes differed for N1m and P2m when they were analyzed individually. However, with a combined “peak to peak” measurement to examine the unitary response to a sound stimulus, the essential differences between these two peaks were statistically equal under the different experimental conditions.

For the N1m and P2m peak latencies, no significant main effect was observed, with all  $F_s < 2.5, ps > 0.10$  (Table 2). Moreover, the interaction between sequence regularities and task conditions was not significant,  $F_s < 1.4, ps > 0.25$ . Taken together, the manipulation of the sequence regularity and timing tasks did not modulate the N1m and P2m peak latencies.

### 3.3. Results of time-frequency analysis

#### 3.3.1. Oscillatory activity in defined ROIs

We first assessed the time-frequency profiles of power responses within each defined ROI during the presentation of the tone sequences. As shown in Fig. 5, the most pronounced power modulation was the fluctuation of alpha power during the sequence presentation period as well as beta power during the “waiting for response” period. For the alpha band (10–13 Hz), a sharp increase of power was observed

following each tone in the sequence, which then fell back slightly just prior to, or at the time of, the occurrence of the next tone. For the beta band (15–30 Hz), we found a noteworthy power decrease following the final tone and a subsequent power increase (“rebound”) above the baseline level after the critical action, suggesting the potential role of beta oscillations in temporal encoding in this time reproduction paradigm. Therefore, we focused our analyses on the 200 ms pre- to 2200 ms post-onset of the final tone time window and further assessed the statistical significance of the observed beta oscillations.

#### 3.3.2. Beta power responses differ across task conditions rather than across stimulus rhythms

We examined power responses for “regular” and “irregular” trials and tested whether this modulation of beta band power uniquely responds to the rhythms of the preceding tone sequences. As a result of a permutation paired t-test, there was no compelling difference for regular/irregular pairs either in the bilateral STG (paired t-test,  $ps > 0.2$ , FDR corrected) or the left SMA (paired t test,  $ps > 0.28$ , FDR corrected), implying that the beta power modulation was manifested to the same extent in the temporal encoding and prediction period before the moment of participants’ final key pressing, regardless of whether the prediction was based on the rhythm or not.

Next, we assessed power responses from different averaging tasks within each defined ROI. As shown in Fig. 6A, in general, there was some apparent similarity between conditions in term of the modification of betaoscillations (i.e., initial decrease followed by a rebound), but with a key difference in the time courses of beta ERD/ERS. In order to reveal this difference, we also performed a permutation paired t-test analysis to compare the induced power responses in pairs between task conditions irrespective of whether the sound sequence was regular or not. From the result within each ROI (see Fig. 6B), the “half” condition produced significantly stronger power responses in the 15–23 Hz range compared to the “equal” (left column) and “double” conditions (right column) around 1 s after the onset of the final tone (paired t test,  $ps < 0.05$ , FDR corrected). Comparison of “equal” and “double” conditions (Fig. 6B, middle column) also showed significant differences (paired t test,  $ps < 0.05$ , FDR corrected) in this frequency range between 1–1.5 s, further suggesting that the beta power responses (15–23 Hz) differed across task conditions rather than across the tone sequence regularities.

#### 3.3.3. Timing of beta desynchronization/synchronization and beta modulation depth

For a closer look at the temporal dynamics of beta power modulation, we extracted the band-limited beta power (15–23 Hz) from each ROI under each of the three task conditions for each participant, summarized the group averaged beta responses (see Fig. 6C), and further

**Table 2**

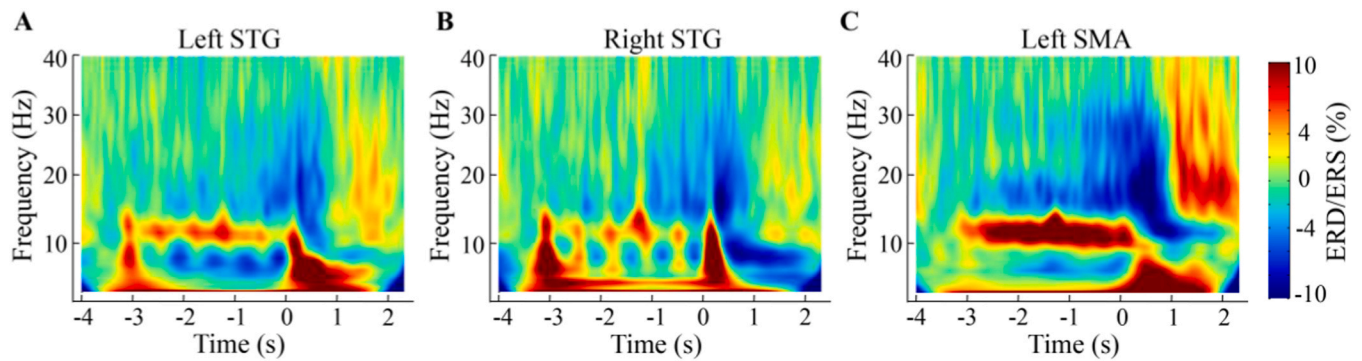
Effect of the task conditions on the peak amplitude (amp.) and latency (lat.) of the N1m, P2m, and CMV components in the predefined central cluster. The values of these measures are the means of all eighteen participants and their standard errors (SE). The highlighted bold figures indicated featured significant one (s) for comparison.

Tasks	N1m		P2m		CMV	
	Peak amp. (fT $\pm$ SE)	Peak lat. (ms $\pm$ SE)	Peak amp. (fT $\pm$ SE)	Peak lat. (ms $\pm$ SE)	Buildup amp. (fT $\pm$ SE)	Peak lat. (ms $\pm$ SE)
Half	<b>-133.4 <math>\pm</math> 12.6</b>	105.5 $\pm$ 2.0	<b>48.5 <math>\pm</math> 10.8</b>	181.0 $\pm$ 4.0	-45.5 $\pm$ 10.0	<b>310.3 <math>\pm</math> 13.1</b>
Equal	-110.2 $\pm$ 10.8	104.0 $\pm$ 2.0	78.1 $\pm$ 10.3	191.1 $\pm$ 5.1	-25.8 $\pm$ 7.0	391.3 $\pm$ 15.4
Double	-106.1 $\pm$ 10.7	103.8 $\pm$ 2.0	86.1 $\pm$ 7.4	184.2 $\pm$ 3.5	<b>-12.3 <math>\pm</math> 6.2</b>	355.2 $\pm$ 12.3
ANOVA	$F(2,34) = 6.7$	$F(2,34) = 0.2$	$F(2,34) = 21.4$	$F(2,34) = 2.4$	$F(2,34) = 6.7$	$F(2,34) = 11.0$
	$p < 0.01$	$p = 0.796$	$p < 0.001$	$p = 0.107$	$p < 0.01$	$p < 0.001$
	Half vs. Equal ( $p = 0.045$ )		Half vs Equal ( $p = 1.8E-4$ )		Half vs. Double ( $p = 0.008$ )	Half vs. Equal
	Half vs. Double ( $p = 0.026$ )		Half vs Double ( $p = 10E-5$ )		Equal vs. Double ( $p = 0.027$ )	( $p = 3.1E-4$ )
						Half vs. Double ( $p = 0.044$ )

N1m: Event-related Magnetic Field Counterpart of N1;

P2m: Event-related Magnetic Field Counterpart of P2;

CMV: Contingent Magnetic Variation; SE: Standard errors of mean.



**Fig. 5.** Grand averaged time-frequency responses during the tone sequence presentation in each defined ROI expressed in ERD/ERS percentage (A: left STG; B: right STG; C: left SMA), with time 0 locked to the onset of the final tone. Each tone in the sequence elicited a typical power increase in the alpha band (10–13 Hz), followed by a notable broadband suppression in the beta band (15–30 Hz) after the whole tone sequence presentation, which then rebounded above the baseline level following the critical action. ERD: event related desynchronization; ERS: event related synchronization. SMA:Supplementary motor area; STG: Superior temporal gyrus.

tested for significant differences between the three task conditions on the following four measures: the latencies of beta ERD and ERS maxima, ERD-to-ERS latency difference (i.e., beta modulation timing), and power difference (i.e., beta modulation depth).

Given the data did not follow normal distributions, we applied non parametric analysis (Friedman's Test/Two Way Analysis of Variance by ranks). As shown in Fig. 6D, upon within-participant factors of task conditions (“half”, “equal”, and “double”) and ROIs (left STG, right STG, and left SMA), the analysis revealed that there were no significant main or interaction effects ( $p > 0.10$ ) for the latency of the ERD maxima, with a grand mean of  $349.8 \pm 21.1$  ms. By contrast, the latency of ERS maxima differed significantly across the task conditions,  $p < 0.001$ . The group mean ERS maxima and its SD were  $1168.6 \pm 49.1$  ms in “half” trials,  $1381.2 \pm 46.6$  ms in “equal” trials, and  $1618.7 \pm 71.9$  ms in “double” trials, which suggested that the beta desynchronization reached its maximum around 350 ms after the onset of the final tone in the sequence, while the timing of beta synchronization varied according to the expected interval duration in different tasks (Table 3).

A two-way Friedman's Analysis of Variance by ranks was also performed on the ERD-to-ERS latency differences and power differences. Concerning latency differences (see Fig. 6D, bottom), we found a significant main effect of task conditions,  $p < 0.001$ , showing that latency differences under the “half” ( $840.3 \pm 56.1$  ms) and “equal” ( $992.5 \pm 456.1$  ms) conditions were significantly smaller than those under the “double” ( $1286.3 \pm 81.1$  ms) condition, with  $p < 0.01$  (Table 3). In the right STG and left SMA, ERD-to-ERS latency differences were modulated by the task conditions,  $p < 0.01$ .

For ERD-to-ERS power differences (see Fig. 6D), the main effects of the task condition and the ROIs were significant,  $p < 0.05$ . The power difference under the “equal” condition ( $30.4 \pm 3.2$  %) was significantly larger than that under the “double” condition ( $25.0 \pm 2.8$  %),  $p < 0.05$ . Meanwhile, the power difference in the left SMA ( $40.6 \pm 6.0$  %) was significantly larger than that in the left STG ( $22.4 \pm 2.2$  %) and right STG ( $22.3 \pm 2.3$  %), with both  $p < 0.01$ .

The beta modulation depth differed across task conditions in defined auditory areas,  $p < 0.001$  for the left STG and  $p < 0.05$  for the right STG. These results demonstrated that beta modulation timing changes with different ensemble coding tasks, in that, when expecting an interval with a relatively short duration under the “half” condition, the beta recovery seemed to be prompt to arrive at the ERS maximum from the ERD maximum. Meanwhile, when expecting an interval with a long duration under the “double” condition, the recovery process was blunted. By contrast, the beta modulation depth varied with task conditions and brain regions. In primary auditory areas, despite the obviously smaller beta modulation depth compared with the motor areas, significant differences were observed between the different task conditions, further

implying a greater sensitivity of beta modulation depth in the auditory cortices when they respond to anticipated durations.

### 3.4. Functional connectivity analysis

We further compared the amplitudes across the left STG, the right STG, and the left SMA, and found higher amplitudes in the left SMA than in the left STG and right STG (with permuted multiple comparison correction,  $p < 0.05$ , from the time range 1697–2158 ms locked to the onset of the last tone) (Fig. 7).

Phase slope index (PSI) analysis indicated that left SMA activities lead the brain (beta) oscillations in right STG activities (Fig. 8). It indicates evidence of the SMA to STG directionality during processing of sequential (temporal) auditory interval information.

## 4. General discussion

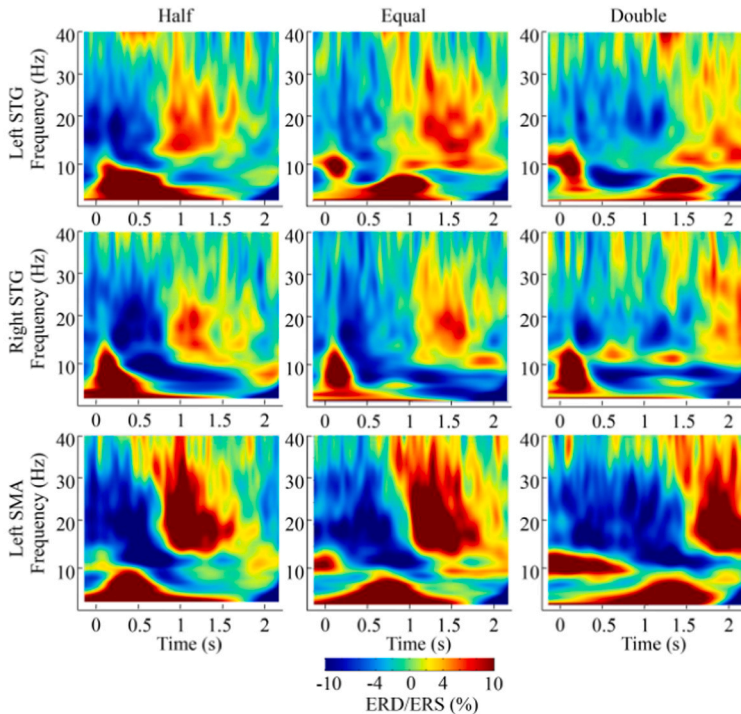
The present study used psychophysics and MEG recordings in human subjects to investigate the underlying mechanisms of ensemble coding (time averaging) and prediction in auditory scenes. Specifically, we manipulated the temporal structure of a tone sequence and asked participants to reproduce intervals whose length was half, equal, or double that of the mean temporal interval in a preceding sequence. ERFs extraction on the sensor level and power analysis of neural oscillations on the source level were employed to examine characteristics of the CMV component and of the beta oscillations in the temporal averaging process. The main finding was that humans are able to summarize the temporally statistical feature (i.e., the mean inter-interval) of the regular (isochronous) and irregular (random) tone sequences with multiple time fractions (scales). Moreover, this summary ability (ensemble coding) on the time domain is represented and featured by the characteristic underpinning of brain activity across sensory and motor areas.

### 4.1. Perceptual averaging in the time dimension

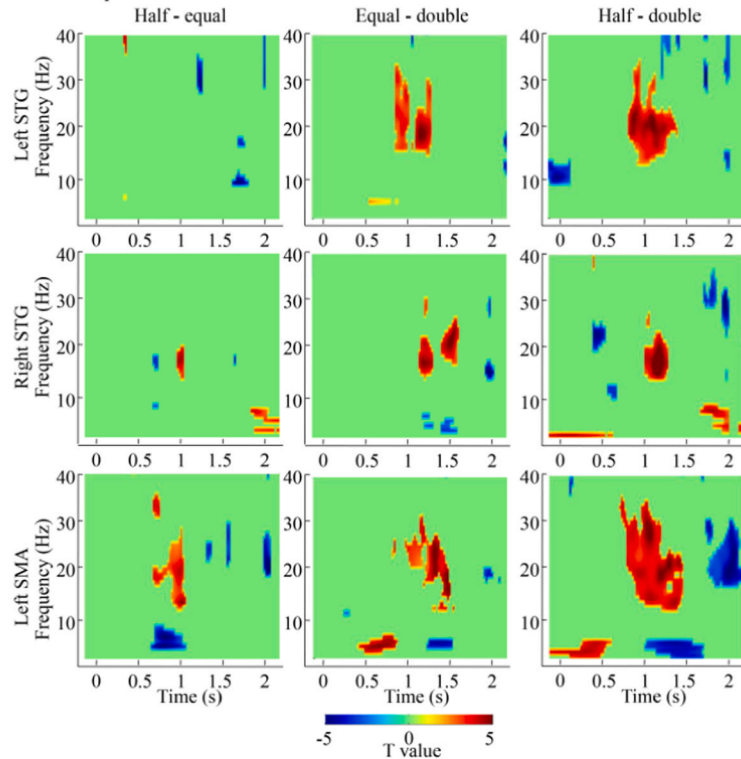
Extracting summary statistics of collections of sensory events could provide us with a perceptual strategy to cope with the limitations of attention and the complexity of our environment (Chen et al., 2018; Chetverikov et al., 2016). Previous studies have mainly focused on statistical summary in the visual domain, and demonstrated that observers are able to efficiently perceive and report average stimulus properties, such as motion direction, orientation, and emotion (Dakin and Watt, 1997; Haberman and Whitney, 2007; Parkes et al., 2001). However, ensemble coding is not a process or an ability that is confined to vision. Auditory averaging has been realized in listeners' estimates of the mean tone frequency (Albrecht et al., 2012; Piazza et al., 2013) or



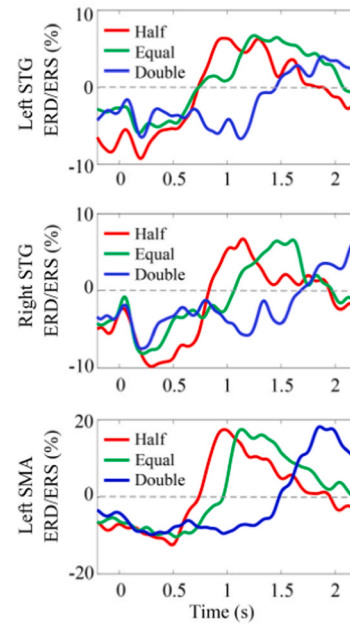
**A Grand averaged power responses for different task conditions within each defined ROI**



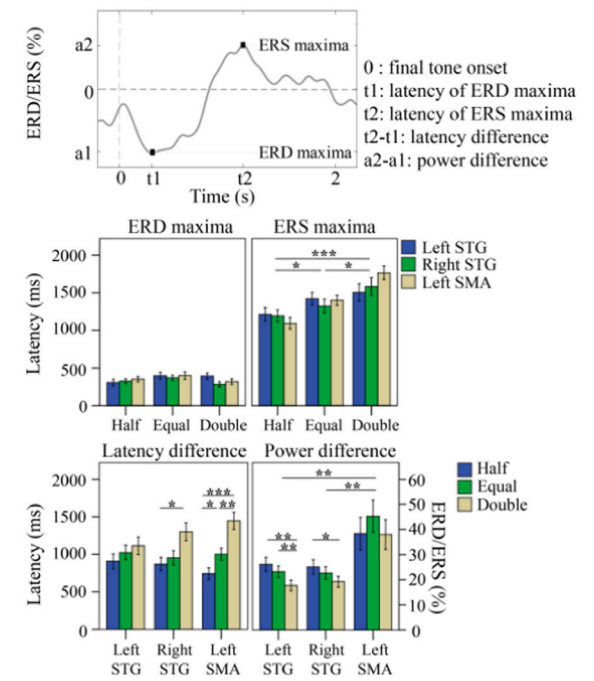
**B Power response differences between task conditions within each defined ROI**



**C Time course of band-limited beta power (15–23 Hz)**



**D Latency of beta desynchronization and synchronization**



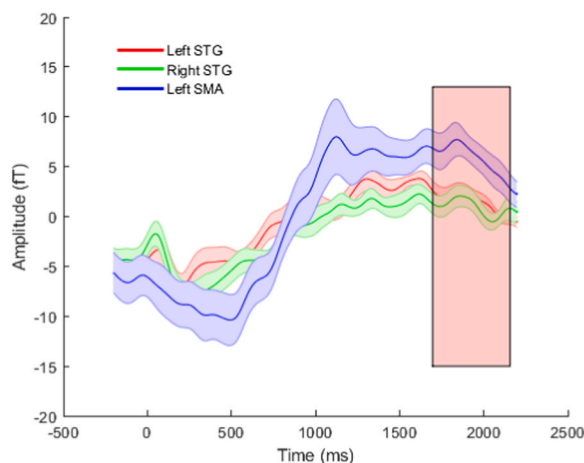
**Fig. 6.** A, Grand-averaged power responses in the bilateral STG and left SMA for the “half” (left column), “equal” (middle column), and “double” (right column) conditions. Time 0 was locked to the onset of the final tone in the sequence. B, Power response differences between task conditions in each defined ROI threshold by non-parametric paired t-test (FDR corrected;  $p < 0.05$ ; left: half-equal; middle: equal-double; right: half-double). C, Grand averaged beta power (15–23 Hz) time-courses for the three task conditions in the left STG (top), right STG (middle), and left SMA (bottom). The red trace depicts the “half” condition, green the “equal” condition, and blue the “double” condition. D, Schema for measuring the peak latency of beta ERD and ERS, latency and power difference (top) and group means of these four measures (middle, bottom). All the error bars show the associated standard errors of the means ( $*p < 0.05$ ;  $**p < 0.01$ ;  $***p < 0.001$ ). ERD: event related desynchronization; ERS: event related synchronization. SMA:Supplementary motor area; STG: Superior temporal gyrus.

**Table 3**

Summary of results concerning beta modulation. This table shows the mean values of four measures (i.e., beta ERD and ERS latency, beta modulation timing, and beta modulation depth) and the standard errors of the mean (SE) observed in different tasks and ROIs. The highlighted bold figures indicated featured significant one (s) for comparison.

Measures	ROIs	Tasks			Interaction	Tasks
		Half	Equal	Double		
ERD latency (ms ± SE)	ISTG	307.6 ± 44.5	397.4 ± 45.4	394.0 ± 41.7	<i>p</i> =0.307	
	rSTG	324.3 ± 32.8	369.7 ± 35.0	283.9 ± 34.4		
	ISMA	352.8 ± 35.2	399.0 ± 48.1	319.5 ± 38.5		
ERS latency (ms ± SE)	ISTG	<b>1215.4 ± 87.4</b>	<b>1420.3 ± 85.0</b>	<b>1506.4 ± 115.3</b>	<i>p</i> <0.001	
	rSTG	<b>1195.0 ± 79.6</b>	<b>1323.2 ± 90.1</b>	<b>1583.4 ± 118.5</b>		
	ISMA	<b>1095.4 ± 78.3</b>	<b>1400.2 ± 67.5</b>	<b>1766.3 ± 89.4</b>		
Beta modulation timing (ms ± SE)	ISTG	907.8 ± 98.3	1022.9 ± 96.5	1112.4 ± 117.1	<i>p</i> <0.001	Half vs.Double, <i>p</i> <0.01 Half vs.Double, <i>p</i> <0.001 Equal vs.Double, <i>p</i> <0.01
	rSTG	<b>870.6 ± 88.9</b>	<b>953.6 ± 96.3</b>	<b>1299.6 ± 119.8</b>		
	ISMA	<b>742.6 ± 79.7</b>	<b>1001.2 ± 81.1</b>	<b>1446.8 ± 115.5</b>		
Beta modulation depth (% ± SE)	ISTG	26.2 ± 2.8	23.2 ± 2.3	17.7 ± 2.1	<i>p</i> <0.001	Half (Equal) vs.Double, <i>p</i> <0.01 Half vs.Double, <i>p</i> =0.033 Equal vs.Double, <i>p</i> =0.045
	rSTG	25.1 ± 3.0	22.7 ± 2.5	19.2 ± 2.2		
	ISMA	38.4 ± 6.5	<b>45.3 ± 6.5</b>	<b>38.0 ± 5.9</b>		

ERD: event related desynchronization; ERS: event related synchronization;  
ISMA: Left Supplementary motor area;  
ISTG-Left Supplementary motor area; rSTG: right Superior temporal gyrus.  
SE: Standard errors of mean.



**Fig. 7.** Amplitude (fT) as a function of time points across the three ROIs. The shaded areas indicate the standard error of the mean. The rectangular area shows the significant differences between amplitudes across the left SMA vs. the left STG and across the left SMA vs. the right STG. ROI: Region of Interest; SMA: Supplementary motor area; STG: Superior temporal gyrus.

the mean tone duration in a sound sequence (Schweickert et al., 2014) and has been adopted to resolve the perceptual uncertainty of visual motion categorization (Chen et al., 2018). However, little is yet known regarding ensemble coding in other sensory aspects of auditory ensembles. To the best of our knowledge, our results provide the first direct evidence of the ensemble coding for task-relevant auditory empty intervals in the time domain.

In the current study, listeners estimated half, equal, or double of the mean duration of multiple intervals in a sequence of six pure tones that were presented with either a regular or irregular rhythm. Notably, the data showed a general overestimation of the duration under the “half”, “equal”, and “double” task conditions, and this overestimation occurred in the different tasks with varying degrees, with the smallest reproduction error in the “double” condition. This consistent overestimation bias could be the result of two factors: sensory noise, which often leads to unfaithful temporal duration measurement, and motor noise, which could hamper the precise timing of movement. It is important to note that, in the present reproduction paradigm, we used a “reproduction by waiting” protocol (i.e., an empty interval), rather than reproduction

with a continuous button press (i.e., synchronizing with given sensory intervals by motor tracking) (Collier and Wright, 1995; Wright and Collier 1994), or experience (context) induced timing calibration (Jazayeri and Shadlen, 2010).

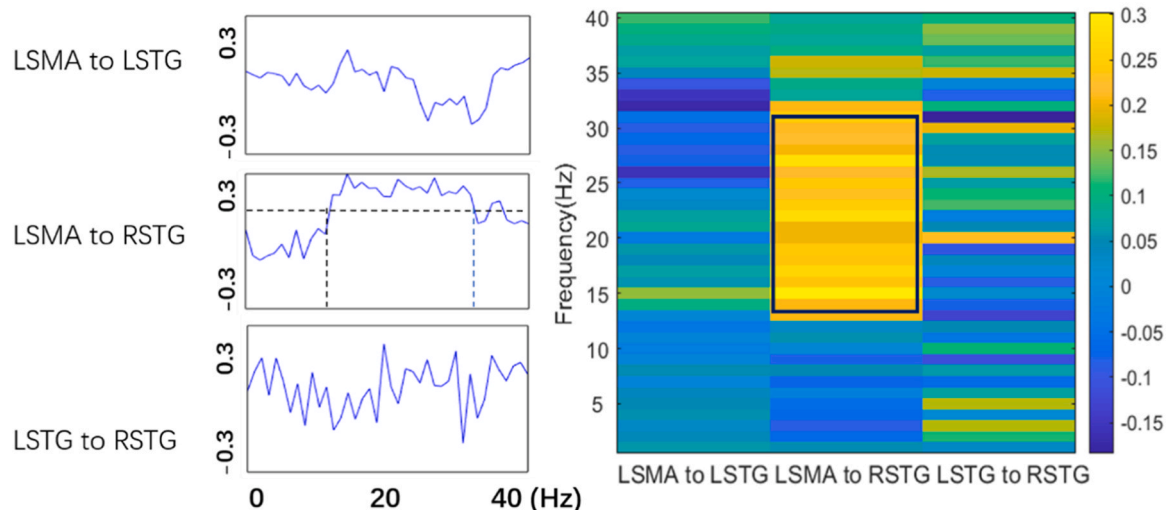
In our paradigm, during the task execution period, participants waited after the tone sequence presentation until the elapsed time was close to the expected duration, then initiated and executed a response (i.e., a key press). It has been suggested that this paradigm might involve an additional component of motor control and planning in time encoding (Wearden, 2003), which could contribute to the delay in the motor response and account for the overestimation found here. We nonetheless adopted this paradigm in the present study because we were considering the tight link between sensation and action, with the immediate motor response post an ensemble encoding of the preceding auditory intervals.

#### 4.2. Temporal encoding featured in the CMV component

In the present study, we investigated slow magnetic fields (i.e., the CMV component) related to sub- and supra-second interval timing by recording magnetic brain signals. Notably, we focused on the “temporal waiting” stage in the task execution period to reveal the characteristic modulation of the CMV component across different averaging tasks. Our results showed that the CMV buildup amplitudes recorded in the central cluster were significantly less negative under the “double” condition than under the “half” and “equal” conditions. Although the difference in the CMV buildup amplitude between the “half” and “equal” conditions did not reach statistical significance, the CMV component tended to build up faster in the “half” condition than in the “equal” condition. The faster buildup process of the CMV corresponded to the shorter reproduced duration, suggesting that a duration-related modulation of the CMV neural generator might account for the temporal accumulation process (Kononowicz and van Rijn, 2015; Kononowicz et al., 2015; Macar and Vidal, 2004; Ruchkin et al., 1986; van Rijn et al., 2011; Walter et al., 1964; Wiener et al., 2015). When the to-be-reproduced interval was shorter in the “half” condition, a relatively precipitous increase of the neuronal excitability in the CMV generator enabled timely estimation and action, quite the contrary to the situation under the “double” condition.

#### 4.3. Robust beta oscillations in interval timing tasks

Our study demonstrated four main findings regarding the beta



**Fig. 8.** Phase slope index (PSI) analysis indicated that left SMA activities significantly led the beta oscillation in the right STG (as shown by the steady increase of slope index at the left middle figure, and the enclosed area (beta range) for the right figure), showing the directionality from SMA to STG during processing of sequential (temporal) auditory interval information. SMA: Supplementary motor area; STG: Superior temporal gyrus.

oscillations induced in an auditory interval timing task. First, robust beta power modulation consisted of a noteworthy power decrease and a subsequent power rebound (Abbasi and Gross, 2020; Arnal et al., 2015; Spitzer and Haegens, 2017). In particular, the former touched the trough at a fixed latency around 350 ms following the tone sequence, but the latter was influenced by the to-be-reproduced intervals. Second, beta-band networks during auditory timing included the auditory cortex and motor-related areas despite their differential functions (Cao et al., 2017; Engel and Fries, 2010; Sedley et al., 2016). Third, the time course of beta-band modulations, in response to and with adjustment to the task-specific interval duration, was not distinct for temporally regular and irregular rhythms. Fourth, the modulation depth of beta-band oscillation differed across conditions only in the auditory cortices, suggesting that those brain areas are the main neural substrates to exhibit more sensitive beta power signatures in sensory areas regarding auditory temporal performance (Wiener et al., 2018).

According to Fujioka et al. (2012), the time course of beta modulation synchronization provides an internal mechanism for predictive timing. The authors presented participants with isochronous tone sequences of different stimulus rates while they were watching a subtitled movie. The auditory evoked responses showed that the initial beta-ERD did not change with the stimulus conditions, while the following beta-ERS reached the maximum around the time of the next stimulus, in which the brain activity provided a temporal representation of the stimulus rate. In line with these earlier results, by using a reproduction paradigm, we also observed the similar behavior of beta oscillatory activity, which first decreased to the ERD maxima at a fixed latency and then rebounded with a temporal adjustment to the task conditions in both the auditory and motor-related cortices. Moreover, by analyzing the beta modulation depth, our data also revealed a greater sensitivity to the expected durations in the bilateral STG.

However, when we compared the temporal dynamics of beta oscillatory activity in the present study vs. the one by Fujioka et al. (2012), we found divergent results. Contrary to the previous evidence, here in the regular stimulus condition, it was alpha oscillations rather than beta activities that synchronized with the beat interval. Indeed, alpha oscillations were the first to be associated with timing functions (Anliker, 1963). There have been a number of studies suggesting that alpha rhythm contributes to sensory processing and involves temporal attention (for a review, see Klimesch, 2012). In our opinion, the rhythmic alpha activities observed with the unfolding of the regular sequence might reflect temporal orienting of attention, in which attention is

deployed with anticipation (focus) on the beat.

Moreover, considering the depth of stimulus processing (actively attending vs. passively listening), response requirements (motor reproduction vs. none), and the analyzed band-limited data (15–23 Hz vs. 20–22 Hz), which differed for the present and previous studies, we supposed that there are different timing functions of beta oscillation. In a passive listening paradigm, as in Fujioka et al. (2012), the periodic beta power modulations were driven by cortical entrainment. In other words, the brain picked up on the periodic characteristic in auditory inputs and thus established automatically endogenous processing of internalized timing, indexed by the same periodic beta synchronization. We could view this function of the periodic beta activity as a reflection of automatic processing of the rhythmic auditory background. In contrast, in the present study with an explicit motor timing task, in which participants attended to the timing of a tone sequence and reproduced an unfilled (empty) interval closely following the sequence according to given rules, the nature of the observed beta oscillation could be different. Here, for an identically rhythmic or non-rhythmic series of tones, the beta rebound varied as a function of the different task-specific time intervals, revealing the explicit and active role of beta-rebound in timing computations.

#### 4.4. The coupling between lSMA and rSTG

While the auditory cortices are critical to encoding the (average) auditory time intervals, this active auditory sensing was typically orchestrated by the functional connectivity pattern between the sensorimotor and auditory areas. We found the directional flow of brain oscillations from the left sensorimotor area to the right superior temporal gyrus, with noticeable leading time for activations in the sensorimotor areas. This observation indicates that in dynamic and active sensing, including motor decision-making based on perceptual averaging for task-relevant temporal information, the motor system is actively involved in predictive timing and is coupled with the featured neural signatures from auditory regions (Morillon and Baillet, 2017).

#### 4.5. Limitations of the present study

In the present study, the music experience of the participants, as well as their differential educational backgrounds, might affect the perceived the auditory duration and lead to individual differences. Evidence has shown that musicians could maintain a more accurate and efficient

mental model for metrical structures, they in general exhibited significantly reduced mismatch negativities (MMNs), compared to their non-musician counterparts (Zhao et al., 2017). The error in reproducing the duration of a stimulus might be overall lower for musically trained than nontrained participants, and the accuracy in estimating the duration of the music clips might be correlated positively with years of musical training (Plastira and Avraamides, 2021). However, we did not record the individual experience of music training in the present study. Moreover, we had a relatively small sample size of only 18 cases. Those limitations could be addressed in future studies.

## 5. Conclusion

The present study substantiated the ability of human beings to summarize the temporal characteristics (i.e., the mean interval) of auditory ensembles, irrespective of the regular or irregular temporal structure. More importantly, this perceptual expertise was delineated by the featured temporal profiles of the CMV component in central sites, which indexed the temporal accumulation process with varying rates of buildup of the waveforms and with beta band activities in auditory and motor areas, which changed the temporal dynamics and modulation depth of the ERD-to-ERS oscillatory pattern with adjustment to the expected durations. Moreover, the sensorimotor area actively engages in temporal predictions of auditory temporal attention, with leading brain activities and resonance in beta-band oscillations. Together, our findings deciphered the contributions to predictive timing of timing-sensitive neural activities in both sensory and motor regions.

## Significance statement

In daily life we encounter often the quick decision making in a sensorimotor loop. How precisely to extract the mean interval from multiple events and respond immediately, remains a challenge yet is critical for human performance. Here we examined how human listeners can precisely extract (ensemble coding) the mean duration of sub-second empty auditory intervals in a sequence and reproduce the target interval with various time fractions (half, equal, and double) of the mean. We demonstrated that the contingent magnetic variation (CMV) and its buildup in the central scalp and beta-band cortical oscillations in the auditory and motor-related regions encode sensorimotor temporal decision making processes. Upon target interval reproduction, a robust beta desynchronization and synchronization oscillatory pattern and featured beta neural oscillations originating from the left sensorimotor cortex and directed toward auditory regions was observed. This coupling of neural routes of auditory encoding and active time reproduction has been the critical neural processing mechanism underlying predictive averaging timing in a sensorimotor loop.

## Author statement

We declare no conflict of interest in this research, and the manuscript has not been submitted

## CRediT authorship contribution statement

**Ming Bao:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Lu Guo:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Lihan Chen:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Zhifei Chen:** Writing – review & editing, Methodology.

## Declaration of Competing Interest

We declare no conflict of interest in this research.

## Data availability

Data will be made available on request.

## Acknowledgements

This study was supported by grants from Natural Science Foundation of China (T2192932, 62061136001, 31861133012), STI2030-Major Project 2021ZD0202600 to L.C. and the Natural Science Foundation of China (11774379) to M.B. We thank the National Center for Protein Sciences at Peking University for assistance with MEG experiments.

## References

- Abbasi, O., Gross, J., 2020. Beta-band oscillations play an essential role in motor-auditory interactions. *Hum. brain Mapp.* *41* (3), 656–665.
- Acerbi, L., Wolpert, D.M., Vijayakumar, S., 2012. Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *Plos Comput. Biol.* *8* (11), e1002771.
- Albrecht, A.R., Scholl, B.J., Chun, M.M., 2012. Perceptual averaging by eye and ear: computing summary statistics from multimodal stimuli. *Atten. Percept. Psychophys.* *74* (5), 810.
- Anliker, J., 1963. Variations in alpha voltage of the electroencephalogram and time perception. *Science* *140* (3573), 1307–1309.
- Arnal, L.H., Doelling, K.B., Poeppel, D., 2015. Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cereb. Cortex* *25* (9), 3077–3085.
- Arnal, L.H., Giraud, A.L., 2012. Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* *16* (7), 390–398.
- Baker, S.N., 2007. Oscillatory interactions between sensorimotor cortex and the periphery. *Curr. Opin. Neurobiol.* *17* (6), 649–655.
- Bartolo, R., Merchant, H., 2015. Oscillations are linked to the initiation of sensory-cued movement sequences and the internal guidance of regular tapping in the monkey. *J. Neurosci.* *35* (11), 1–8.
- Brainard, D.H., 1997. The psychophysics toolbox. *Spat. Vis.* *10* (4), 433–436.
- 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* (2), e2001665.
- Cao, L., Thut, G., Gross, J., 2017. The role of brain oscillations in predicting self-generated sounds. *NeuroImage* *147*, 895–903.
- Chen, L., Zhou, X., Müller, H.J., Shi, Z., 2018. What you see depends on what you hear: temporal averaging and crossmodal integration. *J. Exp. Psychol.: Gen.* *147* (12), 1851–1864.
- Chetverikov, A., Campana, G., Kristjánsson, Á., 2016. Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition* *153*, 196–210.
- Collier, G.L., Wright, C.E., 1995. Temporal rescaling of sample and complex rations in rhythmic tapping. *J. Exp. Psychol. Hum. Percept. Perform.* *21* (3), 602–627.
- Cope, T.E., Grube, M., Griffiths, T.D., 2012. Temporal predictions based on a gradual change in tempo. *J. Acoust. Soc. Am.* *131* (5), 4013–4022.
- Crowley, K.E., Colrain, I.M., 2004. A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin. Neurophysiol.* *115* (4), 732–744.
- Dakin, S.C., Watt, R.J., 1997. The computation of orientation statistics from visual texture. *Vis. Res.* *37* (22), 3181–3192.
- Donner, T.H., Siegel, M., Fries, P., Engel, A.K., 2009. Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Curr. Biol.* *19* (18), 1581–1585.
- Engel, A.K., Fries, P., 2010. Beta-band oscillations-signalling the status quo? *Curr. Opin. Neurobiol.* *20* (2), 156–165.
- Fujioka, T., Ross, B., Trainor, L.J., 2015. Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *J. Neurosci.* *35* (45), 15187–15198.
- Fujioka, T., Trainor, L.J., W., L. E., Bernhard, R., 2012. Internalized timing of isochronous sounds is represented in neuromagnetic  $\beta$  oscillations. *J. Neurosci.* *32* (5), 1791–1802.
- Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* *15* (4), 870–878.
- Gross, J., Baillet, S., Barnes, G.R., Henson, R.N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Maess, B., Oostenveld, R., Parkkonen, L., Taylor, J.R., van Wassenhove, V., Wibral, M., Schoffelen, J., 2013. Good practice for conducting and reporting MEG research. *Neuroimage* *65* (7), 349–363.
- Haberman, J., Whitney, D., 2007. Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* *17* (17), R751–R753.
- Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* *32* (1), 35–42.
- Iversen, J.R., Repp, B.H., Patel, A.D., 2009. Top-down control of rhythm perception modulates early auditory responses. *Ann. N. Y. Acad. Sci.* *1169* (1), 58–73.
- Jasper, H., Penfield, W., 1949. Electroencephalograms in man: effect of voluntary movement upon the electrical activity of the precentral gyrus. *Arch. F. ür. Psychiatr. Und Z. Neurol.* *183* (1-2), 163–174.
- Jazayeri, Mehrdad, Shadlen, Michael N., 2010. Temporal context calibrates interval timing. *Nat. Neurosci.* <https://doi.org/10.1038/nn.2590>.

- Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16 (12), 606–617.
- Kononowicz, T.W., van Rijn, H., 2015. Single trial beta oscillations index time estimation. *Neuropsychologia* 75, 381–389.
- Kononowicz, T.W., Sander, T., van Rijn, H., 2015. Neuroelectromagnetic signatures of the reproduction of supra-second durations. *Neuropsychologia* 75, 201–213.
- Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*. <https://doi.org/10.1126/science.1154735>.
- Li, B., Chen, Y., Xiao, L., Liu, P., Huang, X., 2017. Duration adaptation modulates EEG correlates of subsequent temporal encoding. *Neuroimage* 147, 143–151.
- Li, B., Jia, J., Chen, L., Fang, F., 2021. Electrophysiological correlates of the somatotopically organized tactile duration aftereffect. *Brain Res* 1762, 147432.
- Macar, F., Vidal, F., Casini, L., 1999. The supplementary motor area in motor and sensory timing: evidence from slow brain potential changes. *Exp. Brain Res.* 125 (3), 271–280.
- Macar, F., Vidal, F., 2004. Event-related potentials as indices of time processing: a review. *J. Psychophysiol.* 18 (2-3), 89–104.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164 (1), 177–190.
- Miniussi, C., Wilding, E.L., Coull, J.T., Nobre, A.C., 1999. Orienting attention in time. Modulation of brain potentials. *Brain A J. Neurol.* 122 (8), 1507 ( Pt 8).
- Morillon, B., Baillet, S., 2017. Motor origin of temporal predictions in auditory attention. *Proc. Natl. Acad. Sci. USA* 114 (42), E8913–E8921. <https://doi.org/10.1073/pnas.1705373114>.
- Morillon, B., Schroeder, C.E., Wyart, V., Arnal, L.H., 2016. Temporal Prediction in lieu of Periodic Stimulation. *J. Neurosci.* 36 (8), 2342–2347.
- Nobre, A.C., Correa, A., Coull, J.T., 2007. The hazards of time. *Curr. Opin. Neurobiol.* 17 (4), 465–470.
- Nobre, A.C., Van, F.E., 2017. Anticipated moments: temporal structure in attention. *Nat. Rev. Neurosci.* 19 (1), 34.
- Nolte, G., Ziehe, A., Nikulin, V.V., Schlögl, A., Krämer, N., Brismar, T., Müller, K.R., 2008. Robustly estimating the flow direction of information in complex physical systems. *Phys. Rev. Lett.* 100 (23), 234101.
- Obleser, J., Kayser, C., 2019. Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences* 23 (11), 913–926.
- 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.
- Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10 (4), 437.
- Pfeuty, M., Ragot, R., Pouthas, V., 2003. When time is up: CNV time course differentiates the roles of the hemispheres in the discrimination of short tone durations. *Exp. Brain Res.* 151 (3), 372–379.
- Pfurtscheller, G., Neuper, C., Brunner, C., Silva, F.L.D., 2005. Beta rebound after different types of motor imagery in man. *Neurosci. Lett.* 378 (3), 156–159.
- Piazza, E.A., Sweeny, T.D., David, W., Silver, M.A., David, W., 2013. Humans use summary statistics to perceive auditory sequences. *Psychol. Sci.* 24 (8), 1389–1397.
- Plastira, M.N., Avraamides, M.N., 2021. Music tempo and perception of time: musically trained vs nontrained individuals. *Timing Time Percept.* 10, 1–16.
- Praamstra, P., Kourtis, D., Kwok, H.F., Oostenveld, R., 2006. Neurophysiology of implicit timing in serial choice reaction-time performance. *J. Neurosci.* 26 (20), 5448–5455.
- van Rijn, H., Kononowicz, T.W., Meck, W.H., Ng, K.K., Penney, T.B., 2011. Contingent negative variation and its relation to time estimation: a theoretical evaluation. *Front. Integr. Neurosci.* 5, 91.
- Rohenkohl, G., Cravo, A.M., Wyart, V., Nobre, A.C., 2012. Temporal expectation improves the quality of sensory information. *J. Neurosci. Off. J. Soc. Neurosci.* 32 (24), 8424.
- Ruchkin, D.S., McCalley, M.G., Glaser, E.M., 1977. Event related potentials and time estimation. *Psychophysiology* 14, 451–455.
- Ruchkin, D.S., Sutton, S., Mahaffey, D., Glaser, J., 1986. Terminal CNV in the absence of motor response. *Electroencephalogr. Clin. Neurophysiol.* 63 (5), 445–463 (v).
- Schweickert, R., Han, H.J., Yamaguchi, M., Fortin, C., 2014. Estimating averages from distributions of tone durations. *Atten. Percept. Psychophys.* 76 (2), 605–620.
- Sedley, W., Gander, P.E., Kumar, S., Kovach, C.K., Oya, H., Kawasaki, H., Howard, M.A., Griffiths, T.D., 2016. Neural signatures of perceptual inference. *eLife* 5, e11476.
- Spitzer, B., Haegens, S., 2017. Beyond the status quo: a role for beta oscillations in endogenous content (Re)activation. *eNeuro* 4 (4).
- Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* 2011 (2011), 879716. 2011,(2011-02-23).
- Tark, Kyeong-Jin, Min-Suk Kang, Sang Chul Chong, Won Mok Shim, 2021. Neural representations of ensemble coding in the occipital and parietal cortices. *NeuroImage* 245 (October), 118680.
- Uusitalo, M.A., Ilmoniemi, R.J., 1997. Signal-space projection method for separating MEG or EEG into components. *Med. Biol. Eng. Comput.* 35 (2), 135.
- Walter, W.G., Cooper, R., Aldridge, V.J., Mccallum, W.C., Winter, A.L., 1964. Contingent negative variation: an electric sign of sensorimotor association and expectancy in the human brain. *Nature* 203 (4943), 380–384.
- Wearden, J.H., 2003. Applying the scalar timing model to human time psychology: progress and challenges. *Time Mind*.
- Wiener, M., Parikh, A., Krakow, A., Coslett, H.B., 2018. An intrinsic role of beta oscillations in memory for time estimation. *Sci. Rep.* 8 (1).
- Wiener, M., Thompson, J.C., 2015. Repetition enhancement and memory effects for duration. *Neuroimage* 113, 268–278.
- Zeng, H., Chen, L., 2019. Robust temporal averaging of time intervals between action and sensation. *Front. Psychol.* 10 (March), 511.
- Zhao, S., Chait, M., Dick, F., Dayan, P., Furukawa, S., Liao, H.I., 2019. Pupil-linked phasic arousal evoked by violation but not emergence of regularity within rapid sound sequences. *Nat. Commun.* 10 (1), 4030.
- Zhao, T.C., Lam, H.T.G., Sohi, H., Kuhl, P.K., 2017. Neural processing of musical meter in musicians and non-musicians. *Neuropsychologia* 106, 289–297.