Title

How do you feel the rhythm: Dynamic motor-auditory interactions are involved in the imagination of hierarchical timing

Authors

Cheng, Tzu-Han Zoe
Creel, Sarah C.
Iversen, John R.

Publication Date

2021-11-30

Peer reviewed

Title: How do you feel the rhythm: Dynamic motor-auditory interactions are involved in the imagination of hierarchical timing

Abbreviated title: Motor-auditory interactions in rhythm imagination

Tzu-Han Zoe Cheng (□□□)[1,2*], Sarah C. Creel[1], John R. Iversen[2]

[1]Department of Cognitive Science, UCSD, La Jolla, CA 92093

[2]Institute for Neural Computation and Swartz Center for Computational Neuroscience, UCSD, La Jolla, CA 92093

*Corresponding author: tzcheng@ucsd.edu

Number of pages: 40

Number of figures: 10

Number of words for abstract: 250, significant statement: 118, introduction: 641, and discussion: 1497

Acknowledgments: This work was supported by the National Science Foundation (BCS1460885).

Conflict of interest: The authors declare no competing financial interests

Abstract

Predicting and organizing patterns of events is important for humans to survive in a dynamically changing world. The motor system has been proposed to be actively, and necessarily, engaged in not only the production but the perception of rhythm by organizing hierarchical timing that influences auditory responses. It is not yet well understood how the motor system interacts with the auditory system to perceive and maintain hierarchical structure in time. This study investigated the dynamic interaction between auditory and motor functional sources during the perception and imagination of musical meters. We pursued this using a novel method combining high-density EEG, EMG and motion capture with independent component analysis to separate motor and auditory activity during meter imagery while robustly controlling against covert movement. We demonstrated that endogenous brain activity in both auditory and motor functional sources reflects the imagination of binary and ternary meters in the absence of corresponding acoustic cues or overt movement at the meter rate. We found clear evidence for hypothesized motor-to-auditory information flow at the beat rate in all conditions, suggesting a role for top-down influence of the motor system on auditory processing of beat-based rhythms, and reflecting an auditory-motor system with tight reciprocal informational coupling. These findings align with and further extend a set of motor hypotheses from beat perception to hierarchical meter imagination, adding supporting evidence to active engagement of the motor system in auditory processing, which may more broadly speak to the neural mechanisms of temporal processing in other human cognitive functions.

Significance statement

Humans live in a world full of hierarchically structured temporal information, the accurate perception of which is essential for understanding speech and music. Music provides a window into the brain mechanisms of time perception, enabling us to examine how the brain groups musical beats into, for example a march or waltz. Using a novel paradigm combining measurement of electrical brain activity with data-driven analysis, this study directly investigates motor-auditory connectivity during meter imagination. Findings highlight the importance of the motor system in the active imagination of meter. This study sheds new light on a fundamental form of perception by demonstrating how auditory-motor interaction may support hierarchical timing processing, which may have clinical implications for speech and motor rehabilitation.

1. Introduction

Perceiving temporally hierarchical events is critical for humans to extract meaning from speech and music. In music, perceiving time as various rhythms such as march or waltz is ubiquitous. This process is known as *meter perception*, which is when the basic units, *beats,* are organized into *meters,* nested hierarchical structures of strong and weak position (Large, 2008). Humans are able to willfully subjectively impose hierarchical meters on ambiguous rhythms such that even physically identical series of events can be grouped into either twos or threes like a march or waltz rhythm (Bolton, 1894; Nozaradan et al., 2011, 2012, 2016; Fujioka et al., 2015). These studies showed that meters can be induced not only by bottom-up physical stimulus features such as loudness, but also by intrinsically interpretive or imaginative processes, which are the primary focus of this study to understand how time perception is shaped by endogenous neural activity.

While the underlying mechanisms of how humans organize beats into hierarchical meters are still not clear, numerous studies and theories suggest the importance of motor regions when interpreting rhythmic sounds (see Cannon & Patel, 2021; Merchant et al., 2015; Proksch et al., 2020 for reviews). The motor cortico-basal ganglia-thalamo-cortical (mCBGT) circuit is involved in beat listening tasks even when participants were asked not to move (Grahn and Brett, 2007; Bengtsson et al., 2009; Grahn and Rowe, 2013; Vaquero et al., 2018), and functional connectivity between motor and auditory regions is higher for rhythms with stronger metricality (Chen et al., 2006, 2008; Zatorre et al., 2007; Kung et al., 2013). These studies, together with M/EEG studies of beat processing (Iversen et al., 2009; Fujioka et al., 2012), inspired motor theories of beat perception such as the Action Simulation for Auditory Prediction (ASAP) hypothesis (Patel and Iversen, 2014; see also Schubotz, 2007; Arnal, 2012; Patel and Iversen,

2014; Morillon et al., 2015) which propose that the motor system plays an essential role in the active perception of time, influencing temporal expectation by exerting top-down influence on the auditory system to shape the temporal processing of sound events.

Critical to motor theories is not merely that auditory information influences motor regions (indeed, the influence of auditory information in shaping movement is well-known) but that there is influence in the reverse direction, of the motor system on auditory processing. Most prior studies have examined connectivity between auditory and motor regions in a non-directed way and thus were not able to speak to this distinction. Prior studies have primarily studied simple beat perception, without hierarchical metrical patterns, showing auditory-(sensori)motor phase coupling at the delta band (2 - 4 Hz) when subjects tapped to beats (Nozaradan et al., 2015), and in the beta band (18 - 24 Hz) during a passive listening task of isochronous rhythm (Fujioka et al., 2012). Using a directed connectivity method, Morillon & Baillet (2017) first observed that temporal attention is revealed by a directed-phase transfer entropy from the sensorimotor cortex to the associative auditory cortex in the beta band. These studies broadly support the importance of the motor system at the *beat level*, and while this could parsimoniously be assumed to apply as well for multiple hierarchical levels present in real-life music or speech, how human brains encode and maintain hierarchical organization at the *meter level* has not been tested.

In the current study, we recorded high-density EEG while subjects imagined binary and ternary meters without overt movements (with appropriate control conditions). We then directly tested how the motor and auditory system together respond to imagined meters by isolating the two functional sources to observe their individual neural responses and connectivity between them. Specifically, based on the ASAP hypothesis and other motor theories, we predicted that not only auditory but also motor sources will reflect imagined meters, and that directional

connectivity will show top-down motor-to-auditory flow in beat listening and, critically, during meter imagination.

2. Materials and Methods

2.1 Ethics statement

The IRB office of the University of California, San Diego (UCSD) approved this study. Participants signed informed consent forms to participate. The participants were compensated $15 USD/hour for participating in the experiment.

2.2 Participants

Twenty-six participants (15 females, 11 males, 0 non-binary, mean = 23.08 years old) were recruited. All except for one of them (ambidextrous) were right-handed. All participants reported having normal hearing and no history of psychological disorders, neurodegenerative diseases, or brain trauma. Of these 26 participants, 6 were excluded for the following reasons: one of them (the subject who is ambidextrous) did not finish the task due to technical issues, two could not perform the main task correctly (see more details in Results 3.1), and three had very few usable data after preprocessing. After exclusion, the final sample size was 20. We chose our final sample size to be consistent with recently published works in general and in specific, the most relevant previous study included 21 subjects (Morillon and Baillet, 2017).

2.3 Stimuli

At the beginning of each trial, we presented a double warning-tone (60-ms 400-Hz sine tones), then a series of unaccented and accented bass drum strokes were presented 1 second after

the warning-tone. The drum strokes were modified from GarageBand SoCal drum kit and were delivered in MATLAB (MathWorks) using the Psychophysics Toolbox software (Brainard, 1997). Each drum stroke was modified to be 416.67 ms long to include the attack and decay, played with a 2.4 Hz period (i.e. an inter-onset-interval of 416.67 ms). The stimuli were presented using earbud headphones sealed in the ear (Sony MDRXB50AP). In certain phases of a trial, accented (+10 dB rms) stimuli were presented every two or three drum strokes, creating a binary or ternary meter. To label the onset time of the stimuli, we created a time-locked square pulse for each drum sound as trigger events. Then the sound events and trigger events were combined and output in a wav file. During the experiment, the left channel of the wav file (i.e. sound events) was converted from mono to stereo to play on both left and right channels of the earbuds. The right channel of the wav file (i.e. trigger events) was recorded by the BioSemi Analog Input Box time locked to EEG data recording. In this way, we had a precise (within 0.5 ms) record of sound onsets.

2.4 Procedure

Participants signed informed consent documents and received experiment instruction. As the purpose of the experiment was to investigate metrical perception and production, participants were required to do a metrical tapping practice session before the setup to ensure they could perform the task by accurately recognizing and tapping binary and ternary meters. All participants passed the practice session. Two participants who had the most difficulty grasping the task did not, however, perform the task correctly in the main experiment, and were excluded from further analysis (see more details in Results 3.1).

Figure 1. Localizer trials used to identify auditory and motor cortical sources included spontaneous tapping without sound stimulus, listening to auditory stimuli (unaccented drum strokes) with inter-onset-interval randomized between 1 to 1.5 seconds, and synchronous tapping along with isochronous drum strokes (inter-onset-interval = 600 ms). Each localizer trial took 3 mins to complete. Participants then continued to the main task.

The current experiment included two tasks, a localizer task (Figure 1) and the main task (Figure 2). In the localizer task, participants performed three tasks. First, they performed spontaneous tapping at their preferred tempo without hearing any sound stimulus for 3 minutes (i.e. Tap localizer). They were instructed to tap as stably as possible. Then they listened to unaccented drum stroke stimuli with inter-onset-interval randomized between 1 to 1.5 seconds for 3 minutes (i.e. Listen localizer). Finally, they tapped along with regular drum strokes (inter-onset-interval = 600 ms). They were asked to synchronize with the sound as precisely as possible (i.e. Sync localizer). The Sync localizer was not analyzed and discussed in this study.

In the main task, there were 90 trials. Each trial was divided into 4 conditions in order, Baseline, Physical Meter, Imagined Meter and Tap. Trials were preceded by a warning tone. In the Baseline condition, they were presented with 12 unaccented bass drum strokes played with a 2.4 Hz period. In the Physical Meter condition, the drum strokes were accented every two or three beats, creating either a binary march or a ternary waltz meter. In the Imagined Meter condition, they were instructed to subjectively impose the meter from the Physical Meter

condition onto unaccented sounds that were identical to the Baseline condition. Finally, in the Tap condition, the sound stopped and they tapped the imagined meter, both the strong and weak beats, for verification.

During the whole experiment, subjects wore sealed earbuds and were comfortably seated in a chair with their right hands resting on a support. They were instructed to stay still during the listening part and tap *only* with their index finger during the tapping part of each trial.

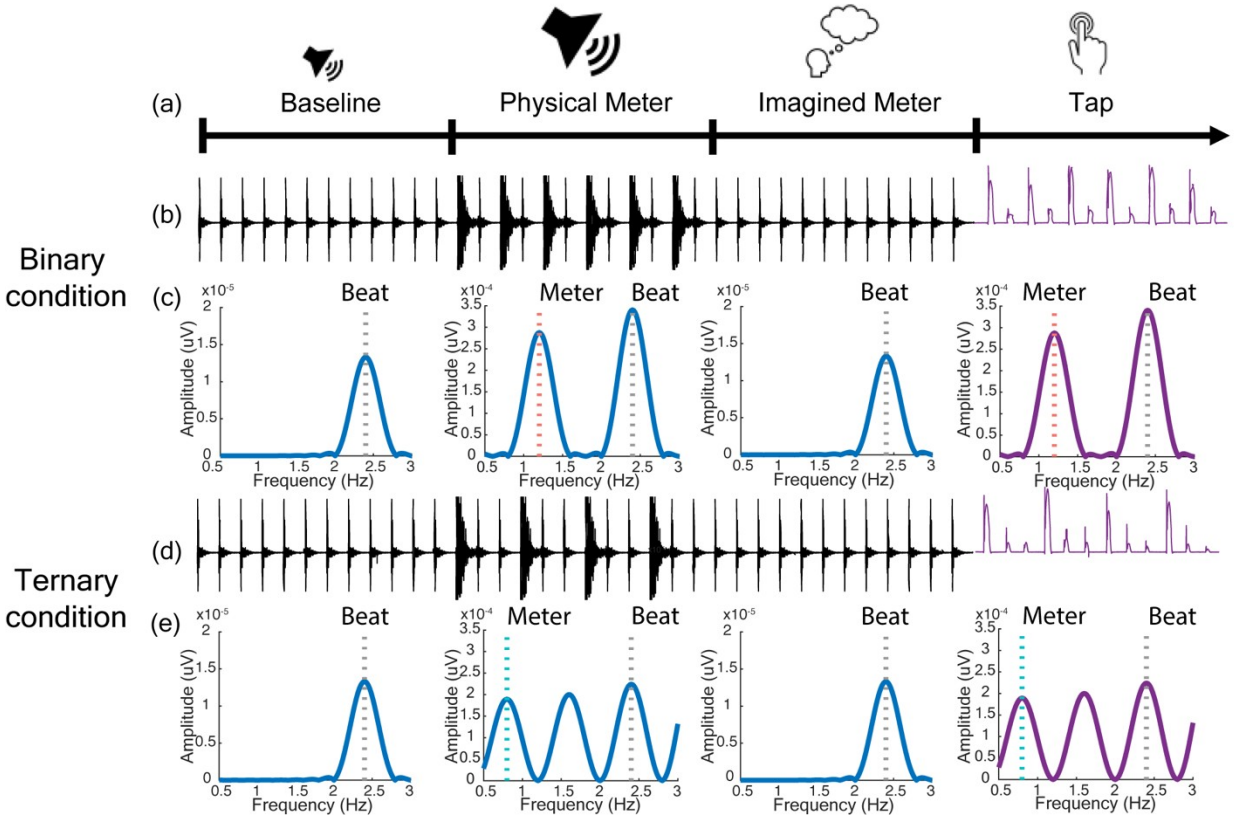

Figure 2. Design of main task (perceptual imagery). (a) One trial consisted of four conditions: Baseline condition (unaccented stimuli), Physical Meter condition (unaccented and accented stimuli), Imagined Meter condition (unaccented stimuli), and imagery-verification Tap condition. Time series of the auditory stimuli of binary meter trial (b) and the ternary meter trial (d). Following the stimulus, a representative tapping pressure signal (purple) is shown. The

amplitude of Fourier transform of the sounds are shown for binary (c) and the ternary (e) stimuli, showing the presence of a meter signal only in the physically accented condition. Dashed lines indicate the spectral peaks at the beat rate (2.4 Hz), and the meter rates for binary (1.2 Hz; pink) and ternary (0.8 Hz; cyan).

2.5 Experiment design

A within-subject 2 (Meter: binary and ternary) × 4 (Conditions: Baseline, Physical Meter, Imagined Meter and Tap) repeated-measures design was used in this study. Each trial was either in a binary or a ternary meter. We had 45 binary trials and 45 ternary trials in total. Each trial, including the warning tone with a gap of 1 second and four conditions (each 5 seconds long), took 21 seconds. The inter-trial interval was randomized between 0 to 5 seconds to avoid anticipation of the next trial. The experiment, including the localizer and the main tasks, took about 45 minutes to complete. The total experiment, with setup, took about 2 hours to complete.

2.6 Data recording

High-density EEG was recorded on a BioSemi Active II system (BioSemi, Amsterdam, Netherlands) using a custom montage cap with 205 active electrodes chosen to fit the participant's head. Data was amplified and sampled at 2048 Hz while recording through the BioSemi ActiveTwo AD-box. The sound and tap trigger were recorded through a BioSemi Analog Input Box sent to the EEG amplifier and became a time locked part of the EEG stream. We first checked offset (< 40) and signal quality in BioSemi ActiView, then used Lab Streaming Layer (Christian Kothe, https://github.com/sccn/labstreaminglayer) to output the EEG data

stream for LabRecorder. Prior to EEG recording, the electrode locations were recorded with a 3-D digitizer (Zebris Medical GmbH, Isny im Allgäu, Germany).

Participants tapped with their right index finger on a custom force-sensing button box (FX1901, 50 lb, TE.com). The participants were instructed to tap on the center (i.e. the bump) of the button and not to rest on or push the button when they weren't required to tap. We monitored during the experiment to ensure that their tapping pattern followed the meter structure (i.e. binary: Stong-Weak-Strong-Weak…; ternary: Stong-Weak-Weak-Strong-Weak-Weak...). The tapping data was recorded through the BioSemi Analog Input Box. The latencies of the moments the finger touched the button were extracted from the data by an in-house script. We attenuated the sound feedback of tapping by using sealed earbuds and putting a layer of soft cloth on the button box. Simultaneously, to monitor that participants did not make unwanted movements during the tasks (especially during imagining), we recorded muscle activity using two surface EMG sensors placed on the first dorsal interosseous muscle on the right hand. Other body parts were also monitored by experimenters in real-time and recorded at 480 Hz using a custom 10 LED marker motion capture configuration (Figure 4; PhaseSpace Inc., San Leandro, CA), measuring the movements of head, elbows, left hand, right index finger, knees and toes.

2.7 EEG preprocessing

Throughout the preprocessing and analysis, in-house scripts with EEGLAB v2019.1 (Delorme and Makeig, 2004) running under Matlab 2019a (The MathWorks, Inc., Natick, MA, USA) were used. The code will be publicly accessible on Open Science Framework. The data were downsampled to 512 Hz and high-pass filtered with cutoff frequency at 0.1 Hz. We performed average referencing on EEG channels, then did artifact removal using artifact

subspace reconstruction (ASR) with the following parameters: Flatline Criterion = -1; Highpass = -1; ChannelCriterion = 0.6; LineNoiseCriterion = -1; BurstCriterion = 5; WindowCriterion = 0.25 (Christian Kothe, https://github.com/sccn/labstreaminglayer; see Chang et al., 2020). Independent component analysis (ICA; infomax binica) was then applied on the data to extract brain EEG sources as well as outside-brain artifact sources like blink, eye movement, and facial and neck muscle activation (Bell and Sejnowski, 1995; Makeig et al., 1996). We localized equivalent dipole locations of independent components (ICs) scalp maps using DIPFIT v3.0 plug-in (Oostenveld et al., 2011). The ICLabel v0.3.1 plugin was used to generate probabilistic labels based on a machine learning algorithm using response features extensively trained by field experts (Pion-Tonachini et al., 2019). Only ICs with probability more than 0.4 as brain ICs went through further analysis. For the frequency analysis and phase analysis, continuous data were epoched into trial conditions (i.e. Baseline, Physical Meter, Imagined Meter and Tap) using a 0 to 5s window without baseline correction. For the directional connectivity analysis, we resampled the data to 128 Hz, and epoch with a -0.6 to 5s window to avoid edge effects.

2.8 Data analysis

2.8.1 Identification of motor and auditory ICs using localizer trials

We applied data-driven ICA to Listen and Tap localizer data to identify one IC in each subject that best captures primary auditory and motor neural activities. ICA was proposed as a solution to this mixing problem by decomposing the recorded EEG into a collection of maximally independent components or independent components (Bell and Sejnowski, 1995; Makeig et al., 1996). The reason for using ICA is to overcome the inherent mixing of cortical sources in the recorded EEG channels. While traditionally a single channel, such as Cz, might be

examined for auditory responses, in a sensory-motor paradigm, this will be contaminated by motor activity. Auditory and motor ICs were identified by selecting the ICs that explained the most variance of sound and movement-related evoked responses in localizer trials, using the following procedures: After applying logistic infomax ICA (i.e. binica of the runica function), we used ICLabel to identify and exclude artifactual, non-brain ICs (threshold > 0.4). After back-projecting all identified brain ICs to EEG channels, we assessed which of the putative brain ICs accounted for the most variance in the auditory and motor event-related potentials (ERPs) computed across all channels in the Listen and Tap localizer trials, using the pvaf (percent variance accounted for) measure. [$pvaf_i$ = 100-100*mean(var(data - $back\_proj_i$))/mean(var(data))], where i indexes the IC number.

For both sound and tap events we used an epoch window from -300 ms to 500 ms relative to event onset, with a baseline from -100 ms to -50 ms. For the auditory ICs, we chose the IC with the highest pvaf of the N1-P2 complex (50 ms - 250 ms time locked to the sound onset) in our Listen localizer trials. The N1-P2 complex was chosen because it is a typical auditory evoked response widely found in auditory relevant tasks (Näätänen and Picton, 1987; Korzyukov et al., 1999). For the motor ICs, we selected the ICs with the highest pvaf of the movement-related potential (-50 ms to 100 ms time locked to the tap onset) (Gerloff et al., 1998; Pollok et al., 2003; Shibasaki and Hallett, 2006). Note that the right hand tapping has focused attention on the left hemisphere, so we expect to observe a left hemisphere lateralization showing on the topography of our motor ICs as observed in other studies (Krause et al., 2012; Ross et al., 2018).

It is important to note that, although consistent with localized cortical activity, the precision of localization is limited and an IC may include multiple highly temporally dependent

processes, and thus each 'motor' IC could potentially involve multiple processes including temporal prediction, motor planning, movement execution and sensory feedback if they were precisely correlated in time and spatially overlapping. The same logic could be applied to auditory ICs as well given that previous research has found that at least six anatomical sources together contribute to auditory N1 ERPs (Näätänen and Picton, 1987). ICA accomplishes the dissection of auditory from motor activity, but distinguishing the multiple functions within the motor and auditory ICs is beyond the scope of this current study.

2.8.2 Frequency analysis

We conducted frequency transforms on the tapping data and the EEG signals in the main task in order to detect characteristic frequencies related to beat and meter, measuring peaks at the beat frequency (i.e. 2.4 Hz) and meter frequency (i.e. 0.8 and 1.2 Hz for ternary and binary meters). We used the frequency analysis of tapping performances in the main task to verify that participants paid attention during the task and could imagine and tap the beats with the correct metrical organization. Time-series tapping data (a continuous trace of tapping pressure as measured by a force sensor, marked in red in Figure 4a) was averaged across trials, then converted to the frequency domain (FFT; 5s data).

EEG time series were first averaged across epochs to enhance the signal-to-noise ratio by reducing trial-by-trial noise contributed by the task-irrelevant activities. We then transformed the average time-series data for each meter and trial condition (5s) to the frequency domain using a discrete Fourier transform (Frigo and Johnson, 1998). The output was the amplitude frequency spectrum (μV) ranging from 0 to 256 Hz with a frequency resolution of 0.2 Hz. We focused on beat frequency (i.e. 2.4 Hz) in both binary and ternary conditions as well the meter frequency in

binary meter (i.e. 1.2 Hz) and ternary (i.e. 0.8 Hz) meter in order to assess the presence of the appropriate meter-related peak for each experiment condition and meter. Standard baseline-subtraction peak-detection methods (e.g. Nozaradan et al., 2011) can verify that there are peaks at 1.2 Hz and 0.8 Hz in binary and ternary conditions, but because of the closeness of the peaks in relation to the frequency resolution due to the short trial length (i.e. 5s in our study vs. 32s in Nozaradan et al., 2011), such measures are misleadingly influenced by neighboring task-driven activity. Instead, to more directly highlight the contrast of interest that is at the core of this project, we subtract the amplitude at 1.2 Hz by the amplitude at 0.8 Hz in all four conditions, two meters and two ICs. This binary-ternary metric acts as our indicator for metrical processing, contrasting the relative strength of meter response that is expected from the condition with the meter response for the other condition. Such a measure will be larger if the neural response to the binary meter predominates and smaller if responding to the ternary meter.

2.8.3 Phase Synchronization between sound and brain signals (non-directional connectivity measurement)

The synchronization between sound and EEG signals was measured by calculating the phase-locking value (PLV) between the sound envelope and both auditory and motor ICs. All signals were filtered between 2 to 3 Hz (FIR filter calculated by matlab function fir1, order 1024). PLV was computed using the formula (Lachaux et al., 1999):

$$PLV\,n = \frac{1}{T}\left|\sum_{t=1}^{T} e^{i(\theta_1(t,n)-\theta_2(t,n))}\right|$$

where t is each time point, T is the total number of time points, n is each trial. The sound envelope was calculated using the Hilbert transform from the original 2.4 Hz sound stimuli

without any accents or meter structure concatenated to match the maximum trial length (25s) (corresponding to 60 drum strokes). The sound envelope was then downsampled from 44100 Hz to 512 Hz to match the sampling rate of EEG. PLV was computed for each condition by calculating the phase difference consistency across time points in each given condition, then averaged across trials for each subject. PLV ranges between 0 and 1 and a higher value means that the phase difference between the two signals is consistent across time, and thus may suggest that EEG signals precisely track the sound stimuli. On the other hand, low PLV indicates that the phase difference is inconsistent across time, and may suggest that EEG signals are not phase-locked nor precisely tracking the sound stimuli.

2.8.4 Information flow analysis (directional connectivity measurement)

The information flow analysis was carried out by using EEGLAB plugin SIFT v1.52. EEG data were downsampled to 128 Hz. We then averaged the signal across trials at each time point, ensemble mean equals to zero and the variance equals to one, to ensure local stationarity. We also detrended the signal with a segment window of 1s and stepsize of 0.1s. The key parameters we used for fitting the multivariate autoregressive (MVAR) model were model order ($p = 30$), window length (1 s) and window steps (0.1s). The target frequency of interest is the beat frequency (i.e. 2 - 3 Hz). Segment window determines if we can get the frequency of interest (i.e. decide how many cycles of the oscillation are included), while model order is influencing the frequency resolution by deciding how many spectral peaks can be modeled (Schlögl and Supp, 2006). As a rule of thumb, a model order of 30 can fit 15 distinct frequency peaks. The model order of 30 was chosen to provide sufficient degrees of freedom to model the frequencies of interest as well as additional frequencies present in the brain response. It is the

simplest model that would allow us to observe the connectivity at the meter and beat frequencies using simulated data. Segment window of 1s enables us to analyze the neural signals down to 1 Hz without violating the signal stationarity. We used the Vieira-Morf lattice algorithm to fit the MVAR model. The selected parameters yield a parameter to data points ratio of ~.02, which indicates a sufficient amount of data to fit the model (Schlögl and Supp, 2006; Korzeniewska et al., 2008). The direct directed transfer function (dDTF08), the product of the full frequency direct transfer function and the partial coherence to remove spurious links in a multivariate design, was used as the index of effective connectivity (Korzeniewska et al., 2003, 2008). This connectivity measure is based on the information flow and does not imply a direct, unmediated, auditory-motor functional or anatomical connection.

2.8.5 Statistical analysis

Each dependent measure—frequency spectrum, phase-locking value, and information flow—was analyzed in repeated-measures analysis of variance (ANOVA) with different focuses. **For the frequency analysis**, we focused on the difference between meter frequencies (1.2 Hz - 0.8 Hz). The independent variables for the ANOVA test were the two meters (binary, ternary) and the four conditions (Baseline, Physical Meter, Imagined Meter and Tap). We further analyzed whether there is a peak in the corresponding meter conditions by a one-sample t-test. **For the phase-locking value**, we focused on the beat frequency (2 - 3 Hz), keeping an eye toward the comparison among stimulus-auditory and motor-auditory coupling. An ANOVA test was performed on the same independent variables. **For the information flow**, we focused on the directed connectivity at the beat frequency (2 - 3 Hz), analyzing the effect of the two

directionalities (auditory-to-motor flow, motor-to-auditory flow) and the four conditions (Baseline, Physical Meter, Imagined Meter and Tap). In cases where there were sphericity violations, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. We conducted post hoc pairwise comparisons whose p values were adjusted using the Bonferroni correction method (marked as p_b).

2.9 Hypothesis and predictions

We made predictions from the perspective of the ASAP and other motor theories, which propose that the motor system plays a necessary role in beat perception. Specifically, we hypothesized that both auditory and motor ICs will reflect imagined meters, and that there will be information flow not only from auditory to motor but motor to auditory during meter imagination. The frequency analysis is our first step to verify this hypothesis such that we contrast two conditions, Baseline vs. Imagined Meter, which have identical stimulus input but varying in imagined meter. **We predict neural responses in *meter frequency* in the Imagined Meter but not in the Baseline in both auditory and motor ICs.** We also predict stronger neural responses at the *beat frequency* in the Imagined Meter than the Baseline in both ICs since the maintenance of meter frequency may result in larger amplitude at the beat frequency due either to harmonics of the meter oscillation, or a general upregulation of top-down control to actively maintain the rhythm across metrical levels.

An important question is if the neural response in auditory ICs during the Imagined Meter is influenced by the motor system (as opposed to being generated entirely within the auditory system). The presence of meter responses in the motor system during imagining alone does not demonstrate this, so our non-directional and directional connectivity analysis were designed to

further investigate this question. Our non-directional connectivity analysis contrasts phase locking of the auditory ICs with the stimulus on the one hand and with the motor ICs on the other, providing a measure of the relative correlation of auditory activity with the stimulus (bottom-up) and motor activity (top-down). **We predict that the relative balance will depend on condition, with auditory ICs phase-locked more to external stimuli in the Physical Meter and phase-locked more with motor ICs in the IM.** Finally for the directional connectivity, based on ASAP's central hypothesis that motor to auditory influence is present and necessary during beat perception, **we predict directional motor-to-auditory causal flow present in all beat-perception conditions (i.e. Baseline, Physical Meter, Imagined Meter conditions), and we further predict stronger motor-to-auditory flow than auditory-to-motor flow especially in the Imagined Meter than other conditions**, due to the need to stabilize the metrical percept in the absence of physical meter cues.

3. Results

3.1 Tapping data to verify imagery

In the last segment of each trial, participants were asked to tap out their imagined pattern of strong and weak beats as a verification of the correct tempo and meter of imagery. The frequency transform of their tapping data shows that most participants tapped with the 2.4 Hz beat rate, 1.2 binary meter rate and 0.8 ternary meter rate, consistent with correct imagery (Figure 3). We analyzed the peaks nearest to the meter rate (i.e. 0.8 Hz and 1.2 Hz) in each meter condition across participants. In the binary condition, the mean of the peak frequency around 1.2 Hz was 1.200. In the ternary condition, the mean of the peak frequency around 0.8 Hz was 0.808. Tapping at the correct rate and meter is essential for our frequency analysis methods and so

participants who tapped at the wrong tempo, suggesting improper imagery, were excluded if their peaks were beyond 1.5 interquartile ranges (IQR) around the meter frequency mean. One participant was excluded from both meter conditions, while a second was excluded for the binary condition (pink lines in Figure 3).

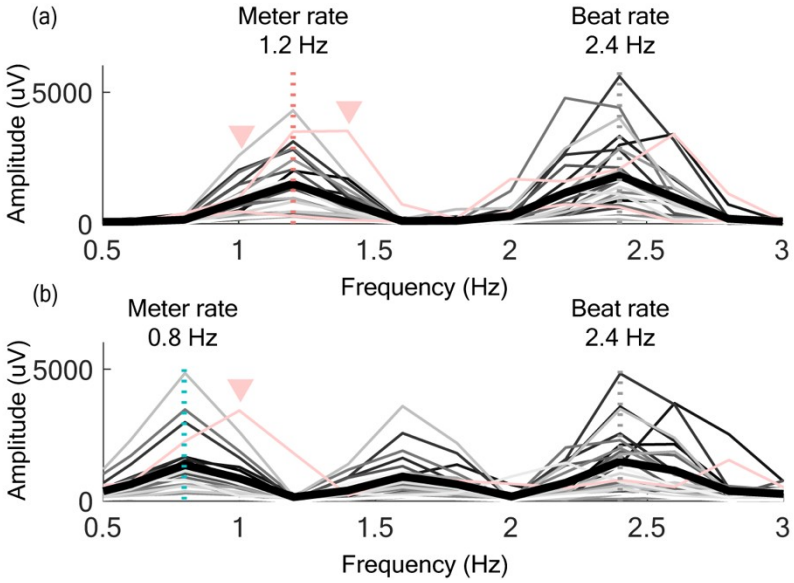

Figure 3. Spectral amplitude of tapping data in the binary (a) and ternary (b) condition. Participant means are shown in gray scale, with the across-participant average shown by the thick black line. Behavioral peaks at the meter rate indicate that the imagined meter was correctly remembered and was a precondition for study inclusion, with outliers (pink lines, arrows point to the peaks at the tapping rates) tapping at rates beyond 1.5 quartiles of the data.

3.2 EMG and motion capture data to confirm no overt movement during meter imagining

We analyzed EMG and motion capture data to confirm that participants did not move rhythmically in any condition other than the tapping condition. This is important, as subtle, even unintentional movements at the meter rate could masquerade as a truly "imagined" meter

response; in order to be certain that the imagined meter response relates to endogenous processes and not movement, it is critical to verify the absence of movement objectively, and not simply trust that participants were capable of following instructions to not move. Figure 4 summarizes the results of the movement analysis, showing EMG and finger position for the right, tapping, index finger. Time series of tapping finger EMG envelope and motion capture of finger movement (Figure 4b) confirmed there was periodic, meter-related rhythmic muscle activity and finger movements only in the Tap condition (purple curves). Frequency analysis confirmed beat and meter power was not present in any of the listening conditions. The only finger movement present in the Imagined Meter condition was a finger lift preparatory to the following tapping condition (visible as slowly-ascending yellow lines in Figure 4b and slightly higher yellow bar than the blue bar of the marker 6 right index finger in Figure 4d) -- there was no rehearsal or shadowing of the meter. To rule out the presence other covert body motions such as head bobbing and foot tapping during in listening conditions, Figure 4c and 4d shows a comprehensive statistical analysis of rhythmic movement across all ten body markers, again confirming no significant power was evident at the beat (2.4 Hz) and meter rate (averaged across 0.8 Hz and 1.2 Hz) by paired t-tests between the Baseline and the Imagined Meter conditions (p_b > .05). Note that the Imagined Meter condition is slightly higher than the Baseline condition in marker 6 in Figure 4d, which is due to the preparatory finger lifting for the following Tap condition. There was neither meter nor beat components found in the first 4.5s (i.e. before the finger lifting) of the Imagined Meter condition. In sum, we objectively demonstrated that no unwanted movements were observed in the task, confirming that our participants followed instructions and critically confirming the activities of motor ICs in meter perception and imagination cannot be explained by covert or overt body movements.

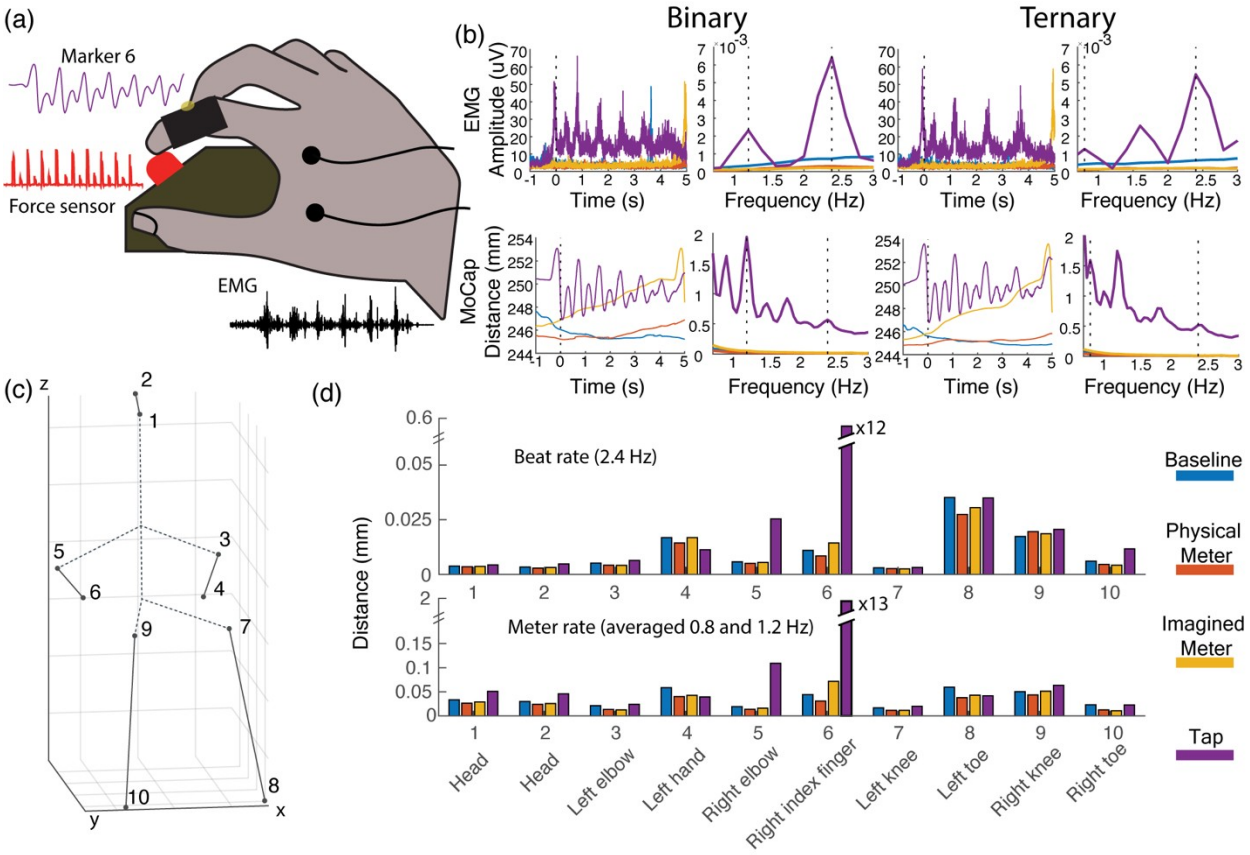

Figure 4. (a) Experimental setup with right hand tapping on a force sensor (red) while instrumented with EMG (black) and motion capture (purple). (b) Averaged time series and amplitude spectra of finger EMG (top row) and movement (bottom row) in binary and ternary meters across conditions. (c) Motion capture marker locations across the body. (d) The averaged displacement of all motion capture markers at meter and beat rate across conditions. Note that the discontinuity of the y-axis is used to show the displacements in the non-tapping conditions which are much smaller than the displacement in the tapping condition.

3.3 Identification of auditory and motor ICs from localizer trials

Auditory and motor responses were modeled as the best fitting ICs to Listen and Tap localizer trials. To extract the auditory ICs, one hundred and fifty trials were extracted from the

Listen localizer. After preprocessing, an average of 145.85 (SD = 7.78) trials per participant (out of the original 150) were available for analysis of the auditory ERP. The average percent variance accounted for (pvaf) of auditory ICs across our final sample was 40.20% (SD = 16.38%). For the motor ICs an average of 352.40 (SD = 122.89) taps available for analysis of the movement-evoked response in the Tap localizer task. The average pvaf of motor ICs across our final sample is 36.17% (SD = 18.77%). In two out of twenty subjects the same IC improbably contributed to the highest pvaf in sound and movement-related potentials and we manually selected the auditory and motor ICs for them based on the best fit to scalp maps, ERPs and spectra. Such a situation may have arisen due to individual differences in cortical geometry, which may prevent ICA from converging on functionally distinct components. The averaged scalp maps and the ERPs of auditory and motor ICs across all subjects were shown in Figure 5a and 5b. The two clusters showed distinct patterns: a N1-P2 complex for the auditory cluster (Naatanen and Picton, 1987; Korzyukov et al., 1999) and a movement-evoked potential for the motor cluster (Gerloff, 1998; Hiroshi & Hallett, 2006; Pollok et al., 2002). To further validate our identification of auditory and motor ICs, the unlabeled components were subjected to unsupervised k-means clustering based on their dipole locations and scalp maps. We did not include any time/frequency domain metric in this clustering process to avoid circularity in the later inferential statistics (Kriegeskorte et al., 2009). The results demonstrated that 37 out of 40 ICs were classified the same as our pvaf method (accuracy = 92.5 %).

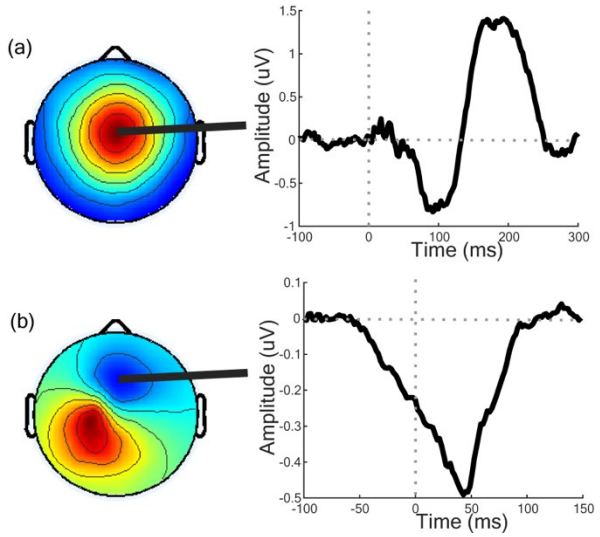

Figure 5. The averaged scalp map and the channel projection ERP time-locked to sound onset for the identified auditory ICs (a) and tap onset for the identified motor ICs (b). Both show canonical patterns consistent with left-hemisphere right-hand-tap-evoked sensorimotor and auditory ERPs. The straight thick black lines (one for each IC) point to the channel used to plot the projection ERPs.

3.4 Auditory and motor responses at the meter and beat frequencies

Our frequency analyses sought to investigate the neural signatures of imagined meters by examining the meter-frequency amplitude in auditory and motor responses and comparing this between the Baseline and the Imagined Meter conditions. This highlights the effect of meter imagining by comparing brain responses to identical, unaccented stimuli without and with metrical imagination. The frequency spectrums ranging from 0.5 to 3 Hz in both meter conditions and both ICs are shown in Figure 6 with a bar chart visualizing the results of statistical analysis of the window of interest in Figure 7.

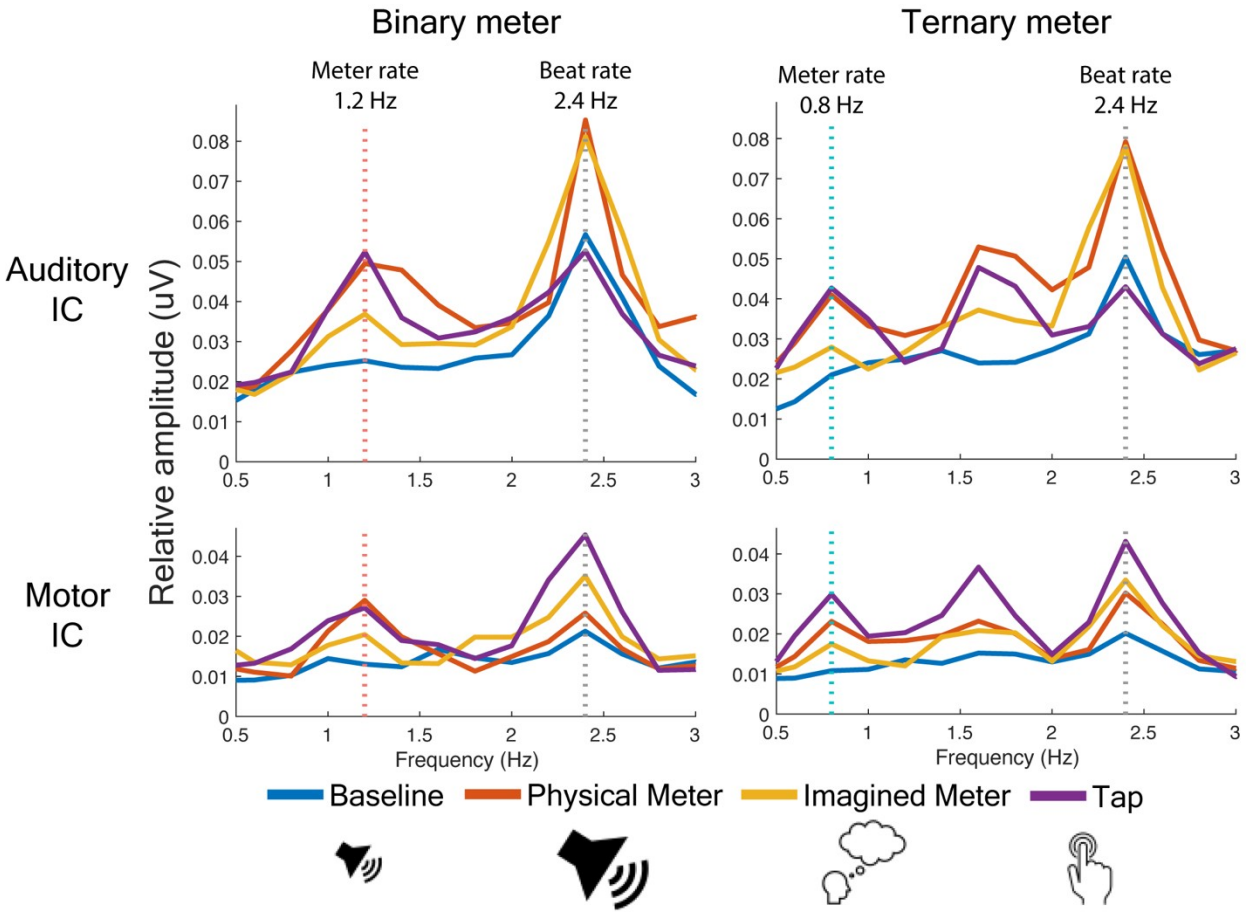

Figure 6. Frequency amplitude spectrum averaged across participants for auditory and motor ICs in binary and ternary meter trials across four conditions (Baseline, Physical Meter, Imagined Meter and Tap indicated by colors as in the legend). Beat (2.4 Hz) and binary and ternary meter frequencies (1.2 and 0.8 Hz) are indicated. Evident spectral peaks are observed at the meter frequency only for Physical Meter, Imagined Meter and Tap conditions, and not for the Baseline, for both auditory and motor ICs and both meters.

3.4.1 Responses at the Meter Frequencies

To investigate physical and imagined meter representations at 0.8 Hz and 1.2 Hz, that is, if there are peaks in corresponding meter frequency, we subtracted the spectral amplitude at 0.8

Hz from the amplitude at 1.2 Hz to provide a measure of metrical discrimination that will be positive for binary meter responses and negative for ternary. If there were no peaks at the meter frequency on the designated conditions, given that the noise level (e.g. including 1/f dynamics accounted for), the results could be close to zero. Overall Meter and interaction effects between Meter and Condition were tested with an ANOVA, where we predicted significant effects of Meter and interaction between Meter and Condition.

In auditory ICs, there was a significant main effect of Meter, $F(1, 19) = 24.540$, $p < .000$, $^2 = .564$, with larger amplitudes in binary compared to ternary meter. There was also a significant interaction effect, $F(3, 57) = 5.395$, $p = .002$, $^2 = .221$. Pairwise comparison showed a significantly larger amplitude in binary than ternary meter in the Physical Meter ($t(19) = 3.386$, $p_b = .003$, $d = 0.972$), Imagined Meter ($t(19) = 2.247$, $p_b = .037$, $d = 0.720$) and the Tap conditions ($t(19) = 3.868$, $p_b = .001$, $d = 1.174$), but not in the Baseline condition ($t(19) = .147$, $p_b = .884$, $d = 0.048$.

In motor ICs, there was a significant main effect of Meter, $F(1, 19) = 29.530$, $p = .000$, $^2 = .609$, with more positive amplitudes in binary compared to ternary meter. There was also a significant interaction effect, $F(1.797, 34.147) = 5.456$, $p = .011$, $^2 = .223$. Pairwise comparison showed a significantly larger amplitude in binary than ternary meter in the Physical Meter ($t(19) = 3.262$, $p_b = .004$, $d = 1.177$), Imagined Meter ($t(19) = 3.876$, $p_b = .001$, $d = 1.210$) and Tap conditions ($t(19) = 4.866$, $p_b = .000$, $d = 1.250$), but not in the Baseline condition ($t(19) = .037$, $p = .968$, $d = 0.012$.

These results highlighted that both auditory and motor ICs reflect the imagined binary and ternary meters by contrasting identical stimuli *without* and *with* metrical imagination (i.e. the

Baseline vs. the Imagined Meter condition), suggesting the endogenous generation of meter timing in both auditory and motor regions.

3.4.2 Response at the Beat Frequency

For the analysis at the 2.4 Hz beat frequency, we focused on the main effect of Condition, keeping an eye toward the patterns of auditory and motor ICs according to the listening and the tapping tasks. We did not expect to see a main effect of Meter, nor the interaction between Condition and Meter. The amplitude at the beat frequency (2.4 Hz) was significantly larger than zero in all conditions based on one-sample t-tests ($p_b < .0001$).

In auditory ICs, there was a significant main effect of Condition, $F(2.258, 42.901) = 6.589$, $p = .002$, $^2 = .257$, with marginally larger amplitude in the Physical Meter compared with the Baseline ($t(19) = 2.818$, $p_b = .066$, $d = 0.692$) and Tap conditions ($t(19) = 3.249$, $p_b = .025$, $d = 0.897$), and a larger amplitude in the Imagined Meter compared with the Baseline condition ($t(19) = 3.257$, $p_b = .025$, $d = 0.587$). Although not expected, there was a significant main effect of Meter, $F(1, 19) = 5.221$, $p = .034$, $^2 = .216$, with slightly higher amplitude in binary compared to ternary meter. There was no significant interaction effect. In motor ICs, we also observed a significant main effect of Condition, $F(1.866, 35.449) = 5.757$, $p = .008$, $^2 = .233$, with the higher amplitudes in the Tap compared with the Baseline condition ($t(19) = 3.492$, $p_b = .015$, $d = 0.989$) and marginally higher amplitude in the Physical Meter compared with the Baseline condition ($t(19) = 2.864$, $p_b = .060$, $d = 0.541$).

These results showed that our selected auditory and motor ICs respond the most to the corresponding task types (i.e. auditory IC peaked at Physical Meter; motor IC peaked at Tap). More importantly, even though the explicit sound stimuli are identical in the Baseline and Imagined Meter conditions, only the Imagined Meter condition showed meter responses.

Furthermore the beat responses in the Imagined Meter condition were also distinct from the Baseline condition, and more similar to the ones in the Physical Meter condition (see beat responses at 2.4 Hz in Figure 6). These findings demonstrate an effect of intrinsic imagination on neural responses, creating endogenous organization at the meter rate as well as modulating stimulus response at the beat rate.

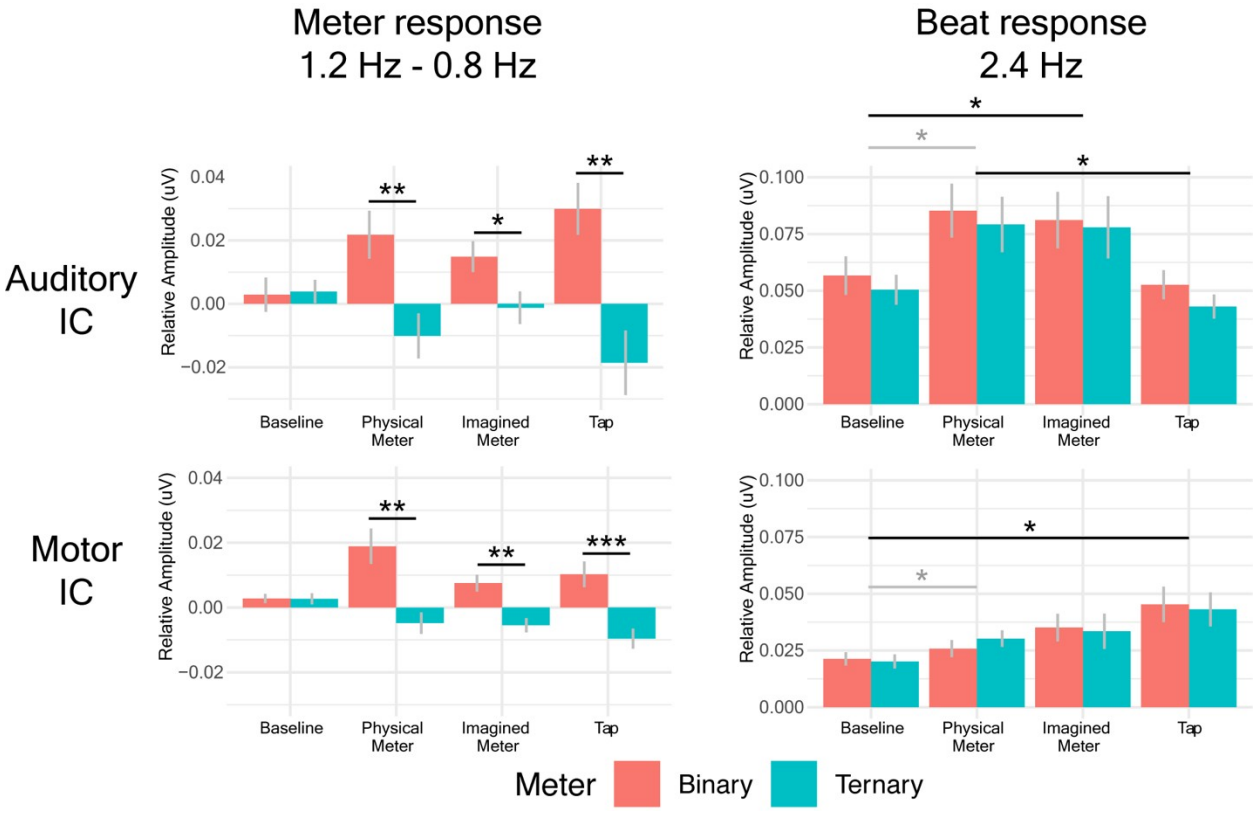

Figure 7. Summary of spectral amplitude differences between the binary and ternary meter frequency (1.2 Hz - 0.8 Hz) and amplitudes at the beat frequency (2.4 Hz) of auditory and motor ICs across four conditions (Baseline, Physical Meter, Imagined Meter and Tap). The results show clear differentiation of metrical responses in both auditory and motor ICs not only for physically accented meters, but also imagined meters. *** $p < .0005$, ** $p < .005$, * $p < .05$, gray * marginal significance.

3.5 Non-directional connectivity analysis among sound, auditory and motor responses

To more closely test if the brain signals follow the external stimuli across time, we calculated PLV between the sound envelope and the selected motor and auditory ICs across time at 2 - 3 Hz (Figure 8).

3.5.1 Sound Envelope and Brain Signal Coupling

Consistent with the frequency spectrum of beat frequency, the coupling between sound stimuli and auditory ICs had a significant main effect of Condition, $F(3, 57) = 3.141$, $p = .032$, $^2 = .142$, with the highest PLV in the Physical Meter compared with the Baseline condition ($t(19) = 3.126$, $p_b = .033$, $d = 0.806$). There was no main effect of Meter nor the interaction effect. To be complete, we also analyzed the coupling between sound stimuli and motor ICs. There was a significant main effect of Condition ($F(3, 57) = 11.980$, $p = .000$, $^2 = .387$). Across the conditions, PLV of the Tap condition was significantly higher compared with the Baseline ($t(19) = 4.152$, $p_b = .003$, $d = 1.363$), Physical Meter ($t(19) = 4.033$, $p_b = .004$, $d = 1.336$) and the Imagined Meter conditions ($t(19) = 3.166$, $p_b = .031$, $d = 1.080$). There was no interaction effect. Consistent with the responses at the beat rate (see 3.4.2), the PLV between the sound envelope and the selected motor and auditory ICs at 2 - 3 Hz also showed that auditory and motor ICs respond the most to the corresponding listening task (i.e. the Physical Meter condition) and the tapping task (i.e. the Tap condition).

3.5.2 Auditory-motor ICs Coupling

The non-directional coupling between auditory and motor ICs had a significant main effect of Condition, $F(3, 57) = 5.363$, $p = .003$, $^2 = .220$, with the highest PLV in the Tap

compared with the Baseline condition ($t(19) = 3.532$, $p_b = .013$, $d = 1.179$). There was no main effect of Meter nor an interaction effect. This finding indicated strong non-directional communication between motor and auditory ICs during the Tap condition, which requires both auditory inputs and motor outputs.

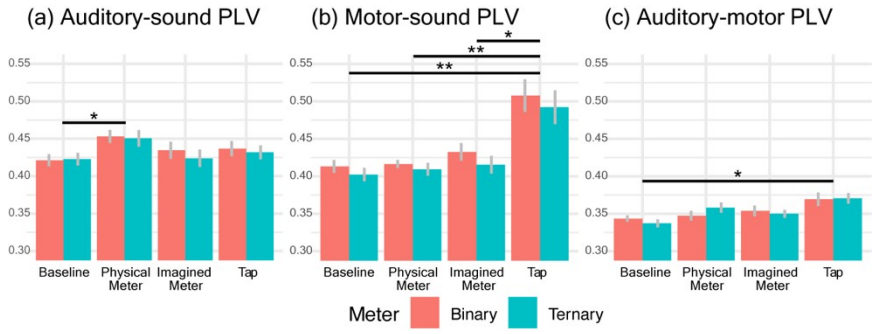

Figure 8. Non-directional connectivity analysis. (a) Phase-locking values of sound envelope and auditory IC across time. (b) Phase-locking values of sound envelope and motor IC across time. (c) Phase-locking values of auditory and motor IC across time. ** $p < .005$, * $p < .05$, gray * marginal significance.

3.6 Information flow between auditory IC and motor IC

We analyzed the direction of information flow (auditory to motor vs. motor to auditory) across conditions at the beat frequency (i.e. delta band 2 - 3 Hz). We averaged the data across two meters to increase the sample size for testing the Direction by Condition effects since the comparison between binary and ternary meters is not the focus of this study (Figure 9 and 10). For thoroughness, we analyzed the causal flow in the alpha band (8 - 12 Hz), which also shows visible connectivity activities across time.

At the beat rate, auditory-to-motor and motor-to-auditory flows are significantly larger than zero in all conditions and directions ($p_b < .000$). There was a significant main effect of

Condition, $F(1.271, 24.142) = 5.424$, $p = .022$, $^2 = .222$, with a significantly stronger information flow in the Physical Meter compared with the Baseline condition ($t(19) = 3.482$, $p_b = .015$, $d = 0.498$) and a marginally stronger information flow in the Tap compared with the Baseline condition ($t(19) = 2.732$, $p_b = .080$, $d = 0.768$). There was no main effect of Direction nor an interaction effect. Interestingly, pairwise comparison showed a marginally significant result such that a stronger information flow from motor to auditory ICs than auditory to motor ICs was seen only in the Imagined Meter condition ($t(19) = 1.957$, $p_b = .065$, $d = 0.198$) , while not in other three conditions. In the alpha band, there were no significant main effects nor the interaction effect. No pairwise differences were observed across the experimental conditions.

In sum, the directional connectivity analysis showed that not only auditory-to-motor but also motor-to-auditory information flow is present at the beat rate in all conditions, demonstrating a tight reciprocal informational coupling during rhythm listening, imagination and production. Bidirectional connectivity was equivalent in the Imagined Meter and Physical Meter conditions, suggesting similar motor-to-auditory communication is present during both meter imagination/maintenance and meter perception. Importantly, this is consistent with previous studies hypothesizing a top-down modulation of auditory regions by motor regions during beat processing. During meter imagery, top-down information flow from motor to auditory ICs was marginally stronger than bottom-up auditory to motor, which could suggest that top-down flow may be slightly relatively more important during imagery. Interestingly, the strongest connectivity was found in the Tap condition, consistent with our PLV results, suggesting that our participants may rely on strong auditory-motor interactions to actively maintain the metrical pattern during tapping without a driving stimulus.

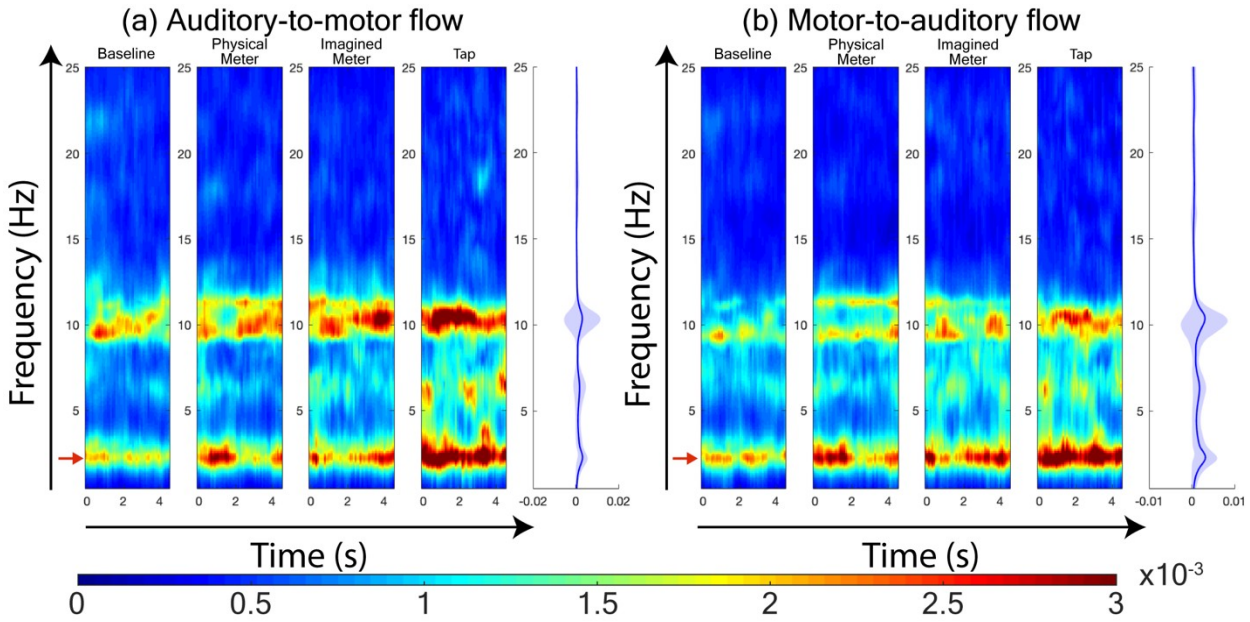

Figure 9. Time-Frequency plots of directional auditory-motor coupling defined by direct Directed Transfer Function causal flow. Color indicates the magnitude of causal flow. (a) Information flow from auditory to motor ICs and (b) from motor to auditory ICs are similar, with significant flow at the beat rate (red arrow). The spectrum of information flow, averaged across time for each frequency bin is shown in the marginal plots (with +/- SE shaded).

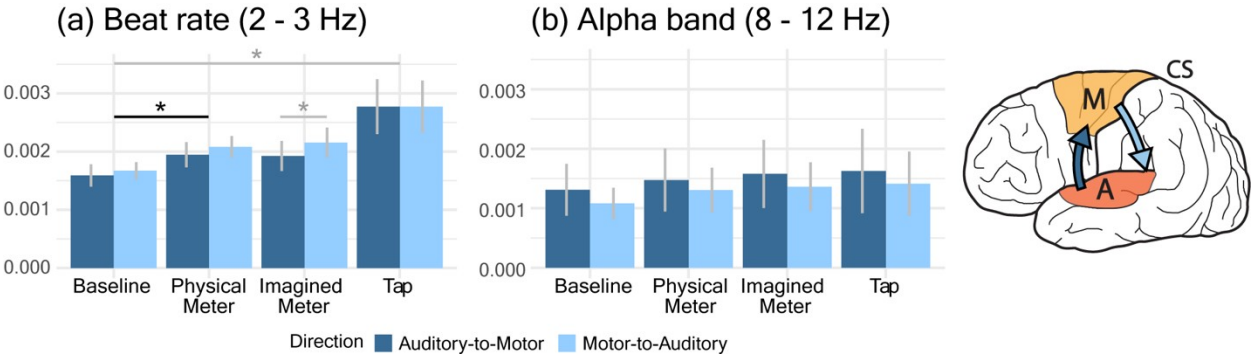

Figure 10. Quantification of information flow value averaged across time at the (a) beat rate (2 - 3 Hz, delta band) and (b) alpha band (8 - 12 Hz). The dark color indicates auditory to motor and the light color indicates motor to auditory as shown in the schematic cortical diagram. M, motor system; A: auditory system; CS, central sulcus. * $p < .05$, gray * marginal significance.

4. Discussion

Principal findings of the current study

The goal of this study was to understand how humans process hierarchically structured temporal information in music rhythm, testing the hypothesis that it involves the motor system. Using a meter imagination paradigm and a novel method to separate auditory and motor source activities with high-density EEG and ICA, we found evidence for motor involvement as hypothesized: Firstly, endogenous neural responses to imagined meters were present in both auditory and motor sources in the absence of corresponding acoustic cues. Secondly, motor-to-auditory information flow was found at the beat rate in all listening conditions without overt movements. These findings confirm and extend predictions of motor theories of rhythm perception, suggesting that the motor system actively maintains hierarchical information and exerts a top-down influence on auditory processing and metrical imagery of rhythms.

Neural substrates of meter imagination

Auditory and motor source activities and their functional connectivity were contrasted between two conditions with identical stimulus but varying in the presence of an imagined meter: In the Baseline the stimuli were presented prior to cueing a specific meter; in the Imagined Meter the participants were instructed to imagine a previously cued binary or ternary meter. Our indicator for metrical processing was spectral activity at the meter rate, which is not physically present in the input signal. No meter response was observed in the non-metrical Baseline in either auditory or motor ICs. In contrast, the meter response was clearly present during meter imagining in the Imagined Meter, demonstrating the endogenous generation of meter timing in both auditory and motor ICs. The spectral power findings extend prior studies using similar

methods (e.g. Nozaradan et al., 2011) by separating auditory and motor activity using a unimodal localizer design and ICA, a critical precondition to investigate the auditory-motor interactions underlying meter imagining.

The presence of meter response during meter imagining in both ICs is suggestive that meter imagining involves the motor system, but a critical question is if the auditory response to the imagined meter is influenced by the motor system. Directional connectivity analysis showed not only highly significant auditory-to-motor connectivity, but equally significant motor-to-auditory flow in all conditions, at the beat rate, demonstrating the presence of bidirectional auditory-motor causal connectivity during listening, imagination and production of rhythms. The main effect of experiment condition and post-hoc pairwise tests revealed that overall connectivity (combining across both auditory-motor and motor-auditory directions) was highest in the Physical Meter and was significantly greater than in the Baseline. Overall connectivity in the Imagined Meter was not significantly different from that observed in the Physical Meter, while it was also not significantly different from the Baseline. Although there was no main effect of connectivity direction nor an interaction effect, we used pairwise comparison to test our proposed motor hypothesis corollary that motor-to-auditory flow would be greater during imagined meters than physically realized meters and found marginal significance ($p=0.065$) only in the Imagined Meter.

Our findings suggests that motor-auditory interaction is essential in hierarchical meter imagery, which may apply to meaningful nested structure of imagined speech as well (Tian and Poeppel, 2012; Proix et al., 2021). We analyzed connectivity at the beat rate, assuming that the beat rate is a reasonable proxy for the importance of motor to auditory flow, since it turned out that we did not have sufficient data to robustly fit an MVAR model capable of capturing causal

flow at the meter frequencies. Future studies will need to collect more data to have the power to describe the meter rate connectivity, which would have been more conclusive regarding meter imagery. Future studies will also examine and extend our findings to other brain areas involved in meter processing (Cannon and Patel, 2021). The observed causal flow in our study was greatest around the beat rate (2-3 Hz), within the delta band, which is proposed to be a top-down temporal constraint of cognitive processes of rhythmic patterns (Morillon et al., 2019). We did not, however, observe beta oscillations and connectivity in the beta range as in previous studies (Iversen et al., 2009; Fujioka et al., 2012, 2015; Morillon and Baillet, 2017), leaving open the question of beta band modulation during meter imagery.

Motor-auditory interaction to physically accented explicit meters

While the Imagined Meter-Baseline contrast in frequency analysis highlights the internal imaginative maintenance of an established, implicit meter, the Physical Meter-Baseline contrast examines response to explicit meter. In addition to an expected meter response in the auditory IC, the motor ICs also showed a significant meter response, even though there were no overt movements required in the Physical Meter (their absence was confirmed via analysis of EMG and motion capture). There was no difference between the Physical Meter and the Imagined Meter in either beat or meter rate, consistent with Iversen et al. (2009) and Vlek et al. (2011) but in a more generalized population that is not restricted to musicians, showing that imagining the meters increases the meter responses in a similar way as if the meter is physically presented. Directional connectivity was significantly stronger in both directions in the Physical Meter than the Baseline, consistent with prior findings that stronger metricality elicits stronger (non-

directional) auditory-motor functional connectivity (Zatorre et al., 2007; Chen et al., 2008; Kung et al., 2013).

One ancillary finding is that neural responses to the binary meter are stronger than the ternary meter in all metrical conditions (i.e. Physical Meter, Imagined Meter and Tap), matching previous findings (Pablos Martin et al., 2007; Celma-Miralles et al., 2021). This may be due to the advantage of binary meter in perception and sensorimotor synchronization (Celma-Miralles et al., 2021; see also Creel, 2012, 2020 for behavioral evidence) or due to a lower signal-to-noise ratio in ternary condition, which had fewer accented positions than binary condition. This can be addressed in future studies by using longer trials.

On the role of the motor system during meter perception and imagination: Revisiting motor hypotheses and other alternative hypotheses

Our results are in line with the motor system role during beat perception posited by the ASAP and other motor hypotheses (Schubotz, 2007; Patel and Iversen, 2014; Morillon et al., 2015; Cannon and Patel, 2021). They extend it in three ways: *Firstly*, current motor hypotheses do not make explicit predictions about how hierarchically structured meters (as opposed to single-level beats) might be represented in the interaction between auditory and motor regions. We found stronger *bidirectional* auditory-motor connectivity in the Physical Meter than the Baseline, consistent with prior findings of stronger non-directional connectivity in fMRI for more strongly metrical rhythms (Chen et al., 2006). *Secondly*, we found marginally stronger motor-to-auditory in the Imagined Meter compared to the Physical Meter, consistent with previously found stronger activations from putamen during beat maintenance, which requires actively regenerating and predicting the beats, compared to beat finding (Grahn and Rowe,

2013). *Finally*, covert movement is a potential confound for this type of studies, and has generally only been dealt with by instructing experimental participants not to move, which is far from definitive. Notably, this study explicitly and objectively verified the absence of motor activity during meter imagining by thorough analyses of finger force-sensor, EMG and 10 whole-body motion capture markers. Thus we can be confident that the endogenous activity at the meter rate in the Imagined Meter was not simply due to covert movement. Ideally, such verification would become standard.

While a unified auditory-motor account of metrical imagery is supported by the evidence, in principle other modality-specific mechanisms could be involved in metrical imagery, for example a) auditory echoic memory of accented sounds from the Physical Meter and b) motor preparation. *Auditory echoic memory* predicts a gradually decreasing meter-frequency amplitude in auditory ICs during the five seconds of the Imagined Meter since memory of acoustic features has been found to decline across a few seconds (Cowan et al., 1997; Snyder and Weintraub, 2013; Cheng and Creel, 2020). Such a decline was not observed (see Figure 7). A purely auditory mechanism would also not predict the observed meter activity in motor ICs nor motor to auditory connectivity. *Motor preparation* during rhythm listening is known to affect motor activity (Chen et al., 2008) and would predict, as observed, meter response in the motor ICs. However, it is not clear how the observed meter activity in the auditory ICs would be explained. Our findings thus appear most parsimoniously explained by an auditory-motor interaction view of meter imagery.

Conclusion

In conclusion, our study demonstrated that the endogenous oscillations in both auditory and motor functional sources reflected the imagination of binary and ternary meters in the absence of

corresponding acoustic cues or overt movements. We found clear evidence for motor-to-auditory information flow at the beat rate in all conditions suggesting a top-down influence of the motor system on auditory processing of beat-based rhythms and reflecting an auditory-motor system with tight reciprocal informational coupling. These findings align with and further extend the ASAP hypothesis and other motor theories from beat perception to meter imagination, adding supporting evidence for active sensing in auditory processing.

References

Arnal LH (2012) Predicting "when" using the motor system's beta-band oscillations. Front Hum Neurosci 6:1–3.

Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7:1129–1159 Available at: https://pubmed.ncbi.nlm.nih.gov/7584893/ [Accessed May 30, 2021].

Bengtsson SL, Ullén F, Henrik Ehrsson H, Hashimoto T, Kito T, Naito E, Forssberg H, Sadato N (2009) Listening to rhythms activates motor and premotor cortices. Cortex 45:62–71.

Bolton TL (1894) Rhythm. Am J Psychol 6:145 Available at: https://www.jstor.org/stable/1410948?origin=crossref [Accessed May 28, 2021].

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436 Available at: https://pubmed.ncbi.nlm.nih.gov/9176952/ [Accessed May 24, 2021].

Cannon JJ, Patel AD (2021) How Beat Perception Co-opts Motor Neurophysiology. Trends Cogn Sci 25:137–150.

Celma-Miralles A., Kleber BA., Toro JM., Vuust P (2021) Neural entrainment facilitates duplets : Frequency-tagging differentiates musicians and non-musicians when they tap to the

beat. bioRxiv.

Chang CY, Hsu SH, Pion-Tonachini L, Jung TP (2020) Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. IEEE Trans Biomed Eng 67:1114–1121.

Chen JL, Penhune VB, Zatorre RJ (2008) Listening to musical rhythms recruits motor regions of the brain. Cereb Cortex 18:2844–2854.

Chen JL, Zatorre RJ, Penhune VB (2006) Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. Neuroimage 32:1771–1781.

Cheng THZ, Creel SC (2020) The interplay of interval models and entrainment models in duration perception. J Exp Psychol Hum Percept Perform 46:1088–1104 Available at: /record/2020-47026-001 [Accessed May 24, 2021].

Cowan N, Saults JS, Nugent LD (1997) The role of absolute and relative amounts of time in forgetting within immediate memory: The case of tone-pitch comparisons. Psychon Bull Rev 4:393–397 Available at: https://link.springer.com/article/10.3758/BF03210799 [Accessed May 24, 2021].

Creel SC (2012) Similarity-based restoration of metrical information: Different listening experiences result in different perceptual inferences. Cogn Psychol 65:321–351 Available at: http://dx.doi.org/10.1016/j.cogpsych.2012.04.004.

Creel SC (2020) Metrical Restoration From Local and Global Melodic Cues: Rhythmic Patterns and Overall Melodic Form. Music Percept An Interdiscip J 38:106–135.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21 Available at: www.sccn.ucsd.edu/eeglab/ [Accessed May 30, 2021].

Frigo M, Johnson SG (1998) FFTW: An adaptive software architecture for the FFT. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pp 1381–1384.

Fujioka T, Ross B, Trainor LJ (2015) Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. J Neurosci 35:15187–15198.

Fujioka T, Trainor LJ, Large EW, Ross B (2012) Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. J Neurosci 32:1791–1802.

Gerloff C, Uenishi N, Hallett M (1998) Cortical activation during fast repetitive finger movements in humans: Dipole sources of steady-state movement-related cortical potentials. J Clin Neurophysiol 15:502–513.

Grahn JA, Brett M (2007) Rhythm and beat perception in motor areas of the brain. J Cogn Neurosci 19:893–906.

Grahn JA, Rowe JB (2013) Finding and feeling the musical beat: Striatal dissociations between detection and prediction of regularity. Cereb Cortex 23:913–921.

Iversen JR, Repp BH, Patel AD (2009) Top-down control of rhythm perception modulates early auditory responses. Ann N Y Acad Sci 1169:58–73.

Korzeniewska A, Crainiceanu CM, Kuś R, Franaszczuk PJ, Crone NE (2008) Dynamics of event-related causality in brain electrical activity. Hum Brain Mapp 29:1170–1192 Available at: https://pubmed.ncbi.nlm.nih.gov/17712784/ [Accessed May 27, 2021].

Korzeniewska A, Mańczak M, Kamiński M, Blinowska KJ, Kasicki S (2003) Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method. J Neurosci Methods 125:195–207.

Korzyukov O, Alho K, Kujala A, Gumenyuk V, Ilmoniemi RJ, Virtanen J, Kropotov J, Näätänen

R (1999) Electromagnetic responses of the human auditory cortex generated by sensory-memory based processing of tone-frequency changes. Neurosci Lett 276:169–172.

Krause V, Bashir S, Pollok B, Caipa A, Schnitzler A, Pascual-Leone A (2012) 1Hz rTMS of the left posterior parietal cortex (PPC) modifies sensorimotor timing. Neuropsychologia 50:3729–3735 Available at: http://dx.doi.org/10.1016/j.neuropsychologia.2012.10.020.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: The dangers of double dipping. Nat Neurosci 12:535–540.

Kung SJ, Chen JL, Zatorre RJ, Penhune VB (2013) Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. J Cogn Neurosci 25:401–420.

Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. Hum Brain Mapp 8:194–208 Available at: /pmc/articles/PMC6873296/ [Accessed May 27, 2021].

Large EW (2008) Resonating to Musical Rhythm: Theory and Experiment. New Scholasticism 39:241–246.

Makeig, S., Bell, A., Jung, T.-P., & Sejnowski, T. (1996) Independent component analysis of electroencephalographic data. Advances in Neural Information Processing Systems 8, 145–151.

Merchant H, Grahn J, Trainor L, Rohrmeier M, Fitch WT (2015) Finding the beat: A neural perspective across humans and non-human primates. Philos Trans R Soc B Biol Sci 370.

Morillon B, Arnal LH, Schroeder CE, 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 Available at: https://doi.org/10.1016/j.neubiorev.2019.09.012.

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, Hackett T, Kajikawa Y, Schroeder C (2015) Predictive motor control of sensory dynamics in Auditory Active Sensing. Curr Opin Neurobiol 176:139–148.

Näätänen R, Picton T (1987) The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. Psychophysiology 24:375–425.

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, Schönwiesner M, Caron-Desrochers L, Lehmann A (2016) Enhanced brainstem and cortical encoding of sound during synchronized movement. Neuroimage 142:231–240.

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.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011 Available at: https://pubmed.ncbi.nlm.nih.gov/21253357/ [Accessed May 30, 2021].

Pablos Martin X, Deltenre P, Hoonhorst I, Markessis E, Rossion B, Colin C (2007) Perceptual biases for rhythm: The Mismatch Negativity latency indexes the privileged status of binary vs non-binary interval ratios. Clin Neurophysiol 118:2709–2715.

Patel AD, Iversen JR (2014) The evolutionary neuroscience of musical beat perception: The

Action Simulation for Auditory Prediction (ASAP) hypothesis. Front Syst Neurosci 8:1–14.

Pion-Tonachini L, Kreutz-Delgado K, Makeig S (2019) ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. Neuroimage 198:181–197.

Pollok B, Müller K, Aschersleben G, Schmitz F, Schnitzler A, Prinz W (2003) Cortical activations associated with auditorily paced finger tapping. Neuroreport 14:247–250.

Proix T, Delgado Saa J, Christen A, Martin S, Knight RT, Tian X, Poeppel D, Doyle WK, Arnal LH, Mégevand P, Giraud A-L (2021) Imagined speech can be decoded from low-and cross-frequency features in perceptual space. bioRxiv:2021.01.26.428315 Available at: https://doi.org/10.1101/2021.01.26.428315.

Proksch S, Comstock DC, Médé B, Pabst A, Balasubramaniam R (2020) Motor and Predictive Processes in Auditory Beat and Rhythm Perception. Front Hum Neurosci 14.

Ross JM, Iversen JR, Balasubramaniam R (2018) The Role of Posterior Parietal Cortex in Beat-based Timing Perception: A Continuous Theta Burst Stimulation Study. J Cogn Neurosci:139 Available at: https://www.apa.org/ptsd-guideline/ptsd.pdf%0Ahttps://www.apa.org/about/offices/directorates/guidelines/ptsd.pdf.

Schlögl A, Supp G (2006) Chapter 9 Analyzing event-related EEG data with multivariate autoregressive parameters. Prog Brain Res 159:135–147.

Schubotz RI (2007) Prediction of external events with our motor system: towards a new framework. Trends Cogn Sci 11:211–218.

Shibasaki H, Hallett M (2006) What is the Bereitschaftspotential? Clin Neurophysiol 117:2341–2356.

Snyder JS, Weintraub DM (2013) Loss and persistence of implicit memory for sound: Evidence from auditory stream segregation context effects. Attention, Perception, Psychophys 75:1059–1074 Available at: https://pubmed.ncbi.nlm.nih.gov/23653411/ [Accessed May 24, 2021].

Tian X, Poeppel D (2012) Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation. Front Hum Neurosci 6:1–11.

Vaquero L, Ramos-Escobar N, François C, Penhune V, Rodríguez-Fornells A (2018) White-matter structural connectivity predicts short-term melody and rhythm learning in non-musicians. Neuroimage 181:252–262.

Vlek RJ, Schaefer RS, Gielen CCAM, Farquhar JDR, Desain P (2011) Shared mechanisms in perception and imagery of auditory accents. Clin Neurophysiol 122:1526–1532 Available at: http://dx.doi.org/10.1016/j.clinph.2011.01.042.

Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: Auditory-motor interactions in music perception and production. Nat Rev Neurosci 8:547–558.

Figure 1. Localizer trials used to identify auditory and motor cortical sources included spontaneous tapping without sound stimulus, listening to auditory stimuli (unaccented drum strokes) with inter-onset-interval randomized between 1 to 1.5 seconds, and synchronous tapping along with isochronous drum strokes (inter-onset-interval = 600 ms). Each localizer trial took 3 mins to complete. Participants then continued to the main task.

Figure 2. Design of main task (perceptual imagery). (a) One trial consisted of four conditions: Baseline condition (unaccented stimuli), Physical Meter condition (unaccented and accented

stimuli), Imagined Meter condition (unaccented stimuli), and imagery-verification Tap condition. Time series of the auditory stimuli of binary meter trial (b) and the ternary meter trial (d). Following the stimulus, a representative tapping pressure signal (purple) is shown. The amplitude of Fourier transform of the sounds are shown for binary (c) and the ternary (e) stimuli, showing the presence of a meter signal only in the physically accented condition. Dashed lines indicate the spectral peaks at the beat rate (2.4 Hz), and the meter rates for binary (1.2 Hz; pink) and ternary (0.8 Hz; cyan).

Figure 3. Spectral amplitude of tapping data in the binary (a) and ternary (b) condition. Participant means are shown in gray scale, with the across-participant average shown by the thick black line. Behavioral peaks at the meter rate indicate that the imagined meter was correctly remembered and was a precondition for study inclusion, with outliers (pink lines, arrows point to the peaks at the tapping rates) tapping at rates beyond 1.5 quartiles of the data.

Figure 4. (a) Experimental setup with right hand tapping on a force sensor (red) while instrumented with EMG (black) and motion capture (purple). (b) Averaged time series and amplitude spectra of finger EMG (top row) and movement (bottom row) in binary and ternary meters across conditions. (c) Motion capture marker locations across the body. (d) The averaged displacement of all motion capture markers at meter and beat rate across conditions. Note that the discontinuity of the y-axis is used to show the displacements in the non-tapping conditions which are much smaller than the displacement in the tapping condition.

Figure 5. The averaged scalp map and the channel projection ERP time-locked to sound onset for the identified auditory ICs (a) and tap onset for the identified motor ICs (b). Both show canonical patterns consistent with left-hemisphere right-hand-tap-evoked sensorimotor and

auditory ERPs. The straight thick black lines (one for each IC) point to the channel used to plot the projection ERPs.

Figure 6. Frequency amplitude spectrum averaged across participants for auditory and motor ICs in binary and ternary meter trials across four conditions (Baseline, Physical Meter, Imagined Meter and Tap indicated by colors as in the legend). Beat (2.4 Hz) and binary and ternary meter frequencies (1.2 and 0.8 Hz) are indicated. Evident spectral peaks are observed at the meter frequency only for Physical Meter, Imagined Meter and Tap conditions, and not for the Baseline, for both auditory and motor ICs and both meters.

Figure 7. Summary of spectral amplitude differences between the binary and ternary meter frequency (1.2 Hz - 0.8 Hz) and amplitudes at the beat frequency (2.4 Hz) of auditory and motor ICs across four conditions (Baseline, Physical Meter, Imagined Meter and Tap). The results show clear differentiation of metrical responses in both auditory and motor ICs not only for physically accented meters, but also imagined meters. *** $p < .0005$, ** $p < .005$, * $p < .05$, gray * marginal significance.

Figure 8. Non-directional connectivity analysis. (a) Phase-locking values of sound envelope and auditory IC across time. (b) Phase-locking values of sound envelope and motor IC across time. (c) Phase-locking values of auditory and motor IC across time. ** $p < .005$, * $p < .05$, gray * marginal significance.

Figure 9. Time-Frequency plots of directional auditory-motor coupling defined by direct Directed Transfer Function causal flow. Color indicates the magnitude of causal flow. (a) Information flow from auditory to motor ICs and (b) from motor to auditory ICs are similar, with

significant flow at the beat rate (red arrow). The spectrum of information flow, averaged across time for each frequency bin is shown in the marginal plots (with +/- SE shaded).

Figure 10. Quantification of information flow value averaged across time at the (a) beat rate (2 - 3 Hz, delta band) and (b) alpha band (8 - 12 Hz). The dark color indicates auditory to motor and the light color indicates motor to auditory as shown in the schematic cortical diagram. M, motor system; A: auditory system; CS, central sulcus. * $p < .05$, gray * marginal significance.