

Macaque monkeys and humans sample temporal regularities in the acoustic environment

Antonio Criscuolo^a, Michael Schwartze^a, Luis Prado^b, Yaneri Ayala^b, Hugo Merchant^b, Sonja A. Kotz^{a,c,*}¹

^a Department of Neuropsychology & Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, the Netherlands

^b Instituto de Neurobiología, UNAM, Campus Juriquilla, Boulevard Juriquilla No. 3001, 76230 Queretaro, QRO, Mexico

^c Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ARTICLE INFO

Keywords:

Neural oscillations
Rhythm processing
Macaque monkeys
EEG
Temporal regularities

ABSTRACT

Many animal species show comparable abilities to detect basic rhythms and produce rhythmic behavior. Yet, the capacities to process complex rhythms and synchronize rhythmic behavior appear to be species-specific: vocal learning animals can, but some primates might not. This discrepancy is of high interest as there is a putative link between rhythm processing and the development of sophisticated sensorimotor behavior in humans. Do our closest ancestors show comparable endogenous dispositions to sample the acoustic environment in the absence of task instructions and training? We recorded EEG from macaque monkeys and humans while they passively listened to isochronous equitone sequences. Individual- and trial-level analyses showed that macaque monkeys' and humans' delta-band neural oscillations encoded and tracked the timing of auditory events. Further, mu- (8–15 Hz) and beta-band (12–20 Hz) oscillations revealed the superimposition of varied accentuation patterns on a subset of trials. These observations suggest convergence in the encoding and dynamic attending of temporal regularities in the acoustic environment, bridging a gap in the phylogenesis of rhythm cognition.

1. Introduction

“The perception, if not the enjoyment, of musical cadences and of rhythm is probably common to all animals and no doubt depends on the common physiological nature of their nervous system” – Charles Darwin.

Research in non-human animal species is considered a test case for unravelling the evolutionary origin(s) of human rhythm cognition (Honing, 2018; Kotz et al., 2018; Ravignani et al., 2019; Patel, 2014). Here, we conceptualize ‘rhythm’ as any pattern of events re-occurring over time. The *temporal regularity* of an acoustic event sequence (from isochronous tones to music) and complex behavior (e.g., walking and speaking) can, thus, be seen as a form of *rhythm*.

What makes us capable of *detecting* temporal regularities in the environment? And in turn, what allows us to *produce* rhythmic behaviors and *synchronize* our movements to an external rhythm? *Detection* requires the *encoding* of a rhythm, i.e., the neurophysiological processing

of temporal regularity. In turn, this allows producing and synchronizing with external rhythms. However, are the capacities to *detect*, *produce*, and *synchronize* rhythms (from now on, ‘DPS’) innate and shared across species?

Comparative studies have shown that DPS is found in vocal learning animals (e.g., parrots; (Patel et al., 2009)), but only partially in others (Schachner et al., 2009). Macaque monkeys performed similarly to humans in tapping tasks (Zarco et al., 2009), showed sensitivity to temporal (ir-)regularities in auditory sequences (Selezneva et al., 2013), and chimpanzees spontaneously aligned their tapping to task-irrelevant auditory rhythms (Hattori et al., 2013). However, the macaque monkey seems unable to detect and synchronize with salient periodicities from complex rhythms such as human-made (musical) stimuli (Honing et al., 2012, 2018). Such results partially supported the notion that many species share a basic capacity for detecting temporal regularities, but some cognitive processes underlying DPS might be species-specific (Patel, 2008, 2006; Fitch, 2013), and further dependent upon neuro-anatomical differences (Merchant and Honing, 2014; Patel and Iversen,

* Corresponding author at: Department of Neuropsychology & Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, the Netherlands.

E-mail address: Sonja.Cimon@gmail.com (S.A. Kotz).

¹ Lead contact.

<https://doi.org/10.1016/j.pneurobio.2023.102502>

Received 15 March 2023; Received in revised form 6 July 2023; Accepted 10 July 2023

Available online 12 July 2023

0301-0082/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2014).

However, other studies have shown that the capacity to process and synchronize behavior with rhythms also depends on voluntary control, cognitive state, attention, and motivation, beyond the actual capacity to control behavior (Wilson and Cook, 2016). When exposed to music, children do not automatically synchronize with the musical beat, but do so in a social context (in presence of adults; (Kirschner and Tomasello, 2009)). Similarly, macaque monkeys spontaneously synchronize behavioral displays (Nagasaka et al., 2012), and bonobos tend to spontaneously synchronize rhythmic behavior with a human experimenter (Large and Gray, 2015).

Thus, a more fundamental question in comparative rhythm cognition is: do nonhuman animals have the neurophysiological predisposition to *encode* temporal regularities in the environment? Once we have addressed this question we can probe their ability for rhythmic displays, which, in isolation, do not necessarily inform about their sensory capacities to process temporal regularities (Wilson and Cook, 2016).

Accumulating evidence shows that premotor cortex cells in macaque monkeys resemble a ‘neural chronometer’ encoding time intervals (Merchant et al., 2011, 2013a) with a dynamic time-varying representation (Crowe et al., 2014). These cells predict regularly timed stimuli (Bartolo and Merchant, 2015) and allow predictive tapping to tempo changes (Gámez et al., 2018). Macaque monkeys’ neural activity indicates the encoding and synchronization with temporal regularities in the sensory environment (Lakatos et al., 2008) and dynamic attending (Large and Jones, 1999) to sensory streams, similarly to humans (for overviews: Obleser and Kayser, 2019; Schroeder and Lakatos, 2009).

In the auditory as well as in other sensory domains, fluctuations of neural activity instantiate a ‘rhythmic mode’ of sensing (Lakatos et al., 2013) and attending (Large and Jones, 1999) the environment that can bias subjective perception (Iemi and Busch, 2018; Zoefel and VanRullen, 2017). Thus, the alternation of high- and low-salience may influence how we perceive sensory input and further shape behavioral performance (e.g., perceptual tasks). For instance, humans show a disposition to perceive *subjective* accentuations (the *tic-toc illusion*) when listening to isochronous equitone sequences (Brochard et al., 2003). Even though all tones are physically identical, the human brain tends to accentuate two or three equidistant tones according to a binary (on-/off-beat; or strong-weak (S-w)) or ternary (S-w-w) pattern (Brochard et al., 2003; Abecasis et al., 2005; Schmidt-Kassow et al., 2011; Poudrier, 2020; Baath, 2015). As the phenomenon emerges during passive listening, the superimposition of accentuations may represent a spontaneous tendency to sample (or attend) the acoustic environment beyond the encoding of single event onsets or time-intervals. As such, subjective accentuation may represent an unbiased marker of endogenous rhythm processing in a quasi-naturalistic context as it does not depend on training nor task demands.

Human event-related potentials (ERPs) mirror these subjective accentuation patterns, showing amplitude differences for tones in S-w positions (Brochard et al., 2003; Abecasis et al., 2005). Neural oscillations in the alpha- and beta-band (8–20 Hz) not only coincide with but precede the onset of expected tones (Fujioka et al., 2012; Arnal, 2012; Fujioka et al., 2009; Snyder and Large, 2005), and their amplitude can be modulated by trial-level accentuation patterns (Criscuolo et al., 2023). These oscillatory brain dynamics likely reflect the active generation of temporal predictions, whereby the first element in a series of two or three might be more salient (S) than others (w).

Does our closest ancestor show a comparable endogenous disposition to sample the acoustic environment in the absence of task instructions and/or training?

In the current study, we recorded EEG in two macaque monkeys who passively listened to isochronous equitone sequences. Similar to a recent study with human participants (Criscuolo et al., 2023), we investigated the monkey’s endogenous tendencies to (i) internalize the timing of external sound events, (ii) track tone onsets, and (iii) parse equitonal sequences with superimposed binary accentuation. Lastly, we directly

compared human and macaque monkey EEG to (iv) test for (dis-)similarities in basic rhythm processing.

The current findings suggest that macaque monkeys have an adequate neural outfit to go beyond simple isochrony processing. The unexplored parallels between macaque monkeys and humans bridge a critical gap in the phylogenesis of rhythm cognition, potentially lending support to Darwin’s notion of a shared neurophysiological predisposition for rhythm processing in non-human primates.

2. Materials and methods

2.1. Experimental procedure

We tested two macaque monkeys. Monkey 1 (M1) is a 12-year-old male, Monkey 2 (M2) a 9-year-old female. Both monkeys had normal hearing and were previously trained in spatial and temporal categorization tasks (M1) (Mendoza et al., 2018) and a synchronization tapping task (M2) (Gámez et al., 2019). Both monkeys were awake (i.e., were not sedated) while EEG was recorded, sitting in a quiet room [3 (l) × 2 (d) × 2.5 (h) m] with dimmed lighting and two loudspeakers placed at a ~50 cm from their ears. The animals were seated comfortably in a monkey chair, where they could freely move their head, hands, and feet. No head fixation was used and the EEG electrodes were attached to the monkey’s scalp using tape (see EEG data acquisition below). To ease the fixation of the electrodes, the monkey’s hair on the scalp and reference ear was shaved. Detailed information about human participants, EEG data collection, and analysis procedures can be found elsewhere (Criscuolo et al., 2023). Briefly, we randomly selected 4 human participants from a dataset of 20 individuals (21–29 years of age, mean age 26.2 years). EEG was recorded from 59 Ag/AgCl scalp electrodes (Electrocap International), amplified using a PORTI-32/MREFA amplifier (DC to 135 Hz), and digitized at 500 Hz.

2.2. Ethics statement

Animal care and experimental procedures were approved by the National University of Mexico Institutional Animal Care and Use Committee and conformed to the principles outlined in the Guide for Care and Use of Laboratory Animals (NIH, publication number 85–23, revised 1985). The Mexican standards on research ethical protocols with non-human primates (NHP) are in the ‘NORMA Oficial Mexicana NOM-062-ZOO-1999’, and are in line with regulations from 13 other countries (Hartig et al., 2023).

2.3. Audiogram

The animal’s hearing capacity was recorded with a scalp-recorded audiogram in partially sedated states. Sedation was induced and maintained with Ketamine (Aranada, Mexico).

2.3.1. Stimuli materials

For the audiogram, click sounds were produced through TTL pulses generated with a TDT-RZ6 signal processor (Tucker-Davis Technologies, system 3, Florida, USA) at a digitization rate of 97656.24 Hz. Clicks lasted 0.5 ms and were delivered at a rate of 15.1 Hz, in random polarities and sound intensities (from 20 to 90, steps of 10 dB SPL). Twenty blocks of 4000 stimuli were presented in a single recording session to collect 1000 repetitions per click intensity. Clicks were binaurally delivered through open-field speakers (KRK 5-G3, USA) located 85 cm from the animal’s ears. Sound intensity was calibrated using a free-field condenser microphone (426B03), a sensor signal conditioner (480C02, PCB Piezotronics, NY USA), and the TDT RPvdsEx circuits.

2.3.2. EEG recording

For the audiogram, continuous EEG was recorded from three Grass gold-plated electrodes (Natus Neurology, #FS-E5GH-60; Fig. 1) located

Experimental condition

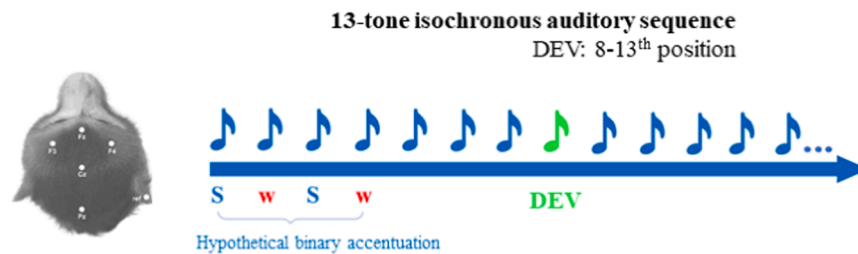


Fig. 1. Electrode positions and stimulus sequence. Electrode positions on the macaque monkey scalp (left) and the 13-tone isochronous equitone sequence (right). The hypothesized superimposition of binary accentuations would parse the auditory sequence in alternating "strong" (S) and "weak" (w) accents. Deviant tones (DEV) occurred from the 8th position onward. Accordingly, they could occur on S and w accentuated positions with equal probability.

at Fz, Cz, and Pz according to the 10/20 system. A reference electrode was located on the right earlobe, while the ground electrode was located on the central forehead. Scalps were shaved and cleaned with a mild abrasive gel (Nuprep, Weaver and Company, USA) before the recording session to reduce scalp impedances. The signal was amplified by means of a medusa preamplifier (RA16PA, TDT systems) and digitized at 24414.06 Hz with an on-line filter from 3 Hz to 6000 Hz. Additionally, a notch filter at 60 Hz was applied to remove the line frequency.

2.3.3. Signal processing

Channel Cz was selected for the analyses of the audiogram based on the known higher signal-to-noise ratio of the vertex signal. The signal was further band-pass filtered (150–3000 Hz, Butterworth 4th-order filter) and epoched relative to stimulus onset (−10 to 66 ms). EEG epochs were sorted into positive and negative polarity click presentations and sub-averages were computed for each polarity condition. An added polarity grand average was obtained and used for further analysis to avoid any transduction stimulus artifact and to minimize the cochlear and microphonic potentials. ABR waves of similar latency as those reported in anesthetized or sedated animals (Laughlin et al., 1999) were observed (Suppl. Fig. 1). A significant evoked response (2-tailed t-test, 10 ms window size) was observed at 60- and 50-dB SPL for monkey 1 and 2, respectively. A previous study in fuscata macaques reported that at a level of 60 dB SPL level, the monkeys were able to hear tones in a 28 Hz to 37 kHz range (Jackson et al., 1999), frequencies that are elicited by the click broad-band stimulus. Three ABR components were clearly identified at 90 dB SPL in both monkeys. These components with positive peaks had latencies of \approx 3, 4.6, and 8 ms. The shape and latency of the first two waves agrees with previous reports (Suppl. Fig. 1, (Lasky et al., 1995)). The likely neural generators of the observed waves are the cochlear nucleus, lateral lemniscus, and inferior colliculus (Lasky et al., 1995; Uno et al., 1993, 1991). The observed peak amplitudes (\approx 3–14 μ V) were larger than those reported in awake head-fixed *Macaca fuscata* (Uno et al., 1993) (tenths of μ V) using similar level and rate parameters but close-field and monaural stimulation. Although it is known that anaesthesia or sedation might considerably diminish or even abolish evoked responses (Uno et al., 1993), the audiogram ensured that both monkeys could perceive the tones employed in the experimental paradigm as both standard (STD) tones and amplitude-deviant (DEV) tones were above the individual hearing threshold in both monkeys.

2.4. Experimental paradigm

Monkeys listened to 13-tone (440 Hz, 85 dB, 50 ms duration) isochronous equitone sequences (Fig. 1, right). On < 5% of the trials, amplitude-deviant tones (DEV; 66 dB) could fall on the 8–9–10–11th position (Fig. 1, bottom row). The inter-stimulus-interval between tones was fixed at 0.6 s, corresponding to a constant stimulus rate of

(1.6667 Hz). The entire trial sequence lasted 7.8 s and was followed by a random inter-trial silent period between 3.5 and 5.5 s. Critically, no accentuation pattern (strong-weak sounds) was imposed on the auditory sequence. M1 underwent 21 recording sessions and M2 25 sessions during which both animals listened to 100 13-tone sequences each. Out of 1300 total events, 1240 were STD tones (>95%) and 60 were DEV (4,6%). The stimulus materials in use were nearly identical for humans and monkeys (Criscuolo et al., 2023).

2.4.1. EEG data acquisition

The EEG was recorded from electrodes (Grass gold-plated electrodes) attached to five scalp positions (Fz, Cz, Pz, F3, F4) according to the 10–20 system (Fig. 1). Both monkeys previously underwent surgery procedures where the head fixation posts were implanted during aseptic surgery and under gas anesthesia. Importantly, the temporal maxillary muscles of the two monkeys were retracted during the surgery, thus leaving the upper skull surface free of muscular or eye-induced artifacts. A second surgery was performed and head holding devices were removed prior to data collection. All electrodes were attached to the scalp using Ten20 Conductive EEG paste and medical tape and were referenced to the right ear (fleshy part of the pinna). The electrodes were connected to a Tucker-Davis Technologies (TDT) head stage (#RA16LI) for low impedance electrodes. This head stage was connected to a TDT RA16PA preamplifier, which in turn was connected to a TDT RZ2 processor. RZ2 was programmed to acquire the EEG signals with a sampling rate of 610.35 Hz and the bandpass filters were set at 0.01–100 Hz.

2.4.2. Data analysis

2.4.2.1. Preprocessing. Data were pre-processed with a combination of custom Matlab scripts/functions and the Matlab-based FieldTrip toolbox (Oostenveld et al., 2011). Data were band-pass filtered with a 4th order Butterworth filter in the frequency range of 0.5–50 Hz (*ft_preprocessing*). Next, data segmentation was conducted separately for 'rhythm-tracking', event-related potentials (ERP) and time-frequency representation (TFR) analyses.

2.4.2.2. Rhythm tracking analyses. Rhythm-tracking analyses were time-locked to encompass the whole equitone sequence. 100 sequences (per experimental session) were created, starting from the third tone onset and including up to the 13th tone (6.6 s). Next, we computed a fronto-central channel cluster encompassing 'Fz', 'F3', 'F4', 'Cz'. Data from this fronto-central cluster were used for Fast-Fourier transform (FFT) and phase-locking analyses.

2.4.2.3. Fast-Fourier transform. Single-trial data from the fronto-central cluster were submitted to a FFT ("FFT data") with an output frequency resolution of 0.15 Hz ($1/6.6 \text{ s} = 0.15 \text{ Hz}$). Spectral power was

calculated as the squared absolute value of the complex Fourier output. Data in each frequency bin were normalized by the frequency-specific standard deviation across trials. Lastly, we averaged the frequency-domain data across channels and trials. For illustration purposes, we restricted the Fourier spectrum to 1–7 Hz in Fig. 2 A, top for M1 and bottom for M2.

2.4.2.4. Phase-locking analyses. A time-resolved phase-locking analysis was performed to estimate the phase relationship between neural activity at the stimulation frequency and the sequential tone onsets. Sequence-level data from the fronto-central cluster were bandpass-filtered with a 4th order Butterworth filter around the stimulation frequency (1.1–2.1 Hz, considering a 1.67 Hz center frequency; *ft.preprocessing*) and underwent Hilbert transform to extract the analytic signal. Next, we plotted the time-course of the real part of the analytic signal (Fig. 2B, top for M1 and bottom for M2) as a function of the onsets of STD tones preceding (blue) and following (red) the DEV (green; this plot is for illustrative purposes only). Phase-locking analyses were performed at sequence- and channel-levels by means of circular statistics (circular toolbox in Matlab (Berens, 2009)) based on the circular mean phase-angles estimated in the 60 ms (proportional to the stimulation frequency: $1/1.67 \text{ Hz}/10$) preceding individual tone onsets. Next, the sequence- and channel-levels mean vector length (MVL; (Berens, 2009)) were calculated for pre-DEV STD tones and the values averaged across

channels. MVL for pre-DEV STD tones were statistically assessed against the MVL from a random distribution (random uniform distribution of phase-angles) by means of 1000 permutation tests. A p-value lower than .05 was considered statistically significant. In Suppl. Fig. 2, we also provide session-, channel-, and sequence-level ‘relative phase angles’. These were expressed as the absolute phase difference between phase-angles for each tone position (e.g., three to eight) and the most common phase-angle in the sequence (the one with the highest probability, as obtained from the *histogram* function in MATLAB, with ‘probability’ as input). The pooling over sessions and channels is displayed in Figs. 2D and 3D.

2.4.2.5. ERP and TFR data. Data were segmented into 4-s-long epochs symmetrically time-locked to every tone onset. Next, we employed a data-driven channel-by-channel and trial- and monkey-level artifact suppression procedure ((Criscuolo et al., 2023); and similar to the method implemented in (Kaneshiro et al., 2020)). Artifact suppression focused on time-windows ranging from -0.4 to 0.4 s relative to each stimulus onset. Amplitude values were temporarily normalized by their standard deviation across trials and outliers (data points per epoch and channel) were defined by means of a threshold criterion (values $> \text{mean} + 4 * \text{SD}$). The identified noisy time-windows (with 50 ms symmetrical padding) then served to suppress (replace by NaNs) time-points in the non-normalized data. The missing values were replaced by means of

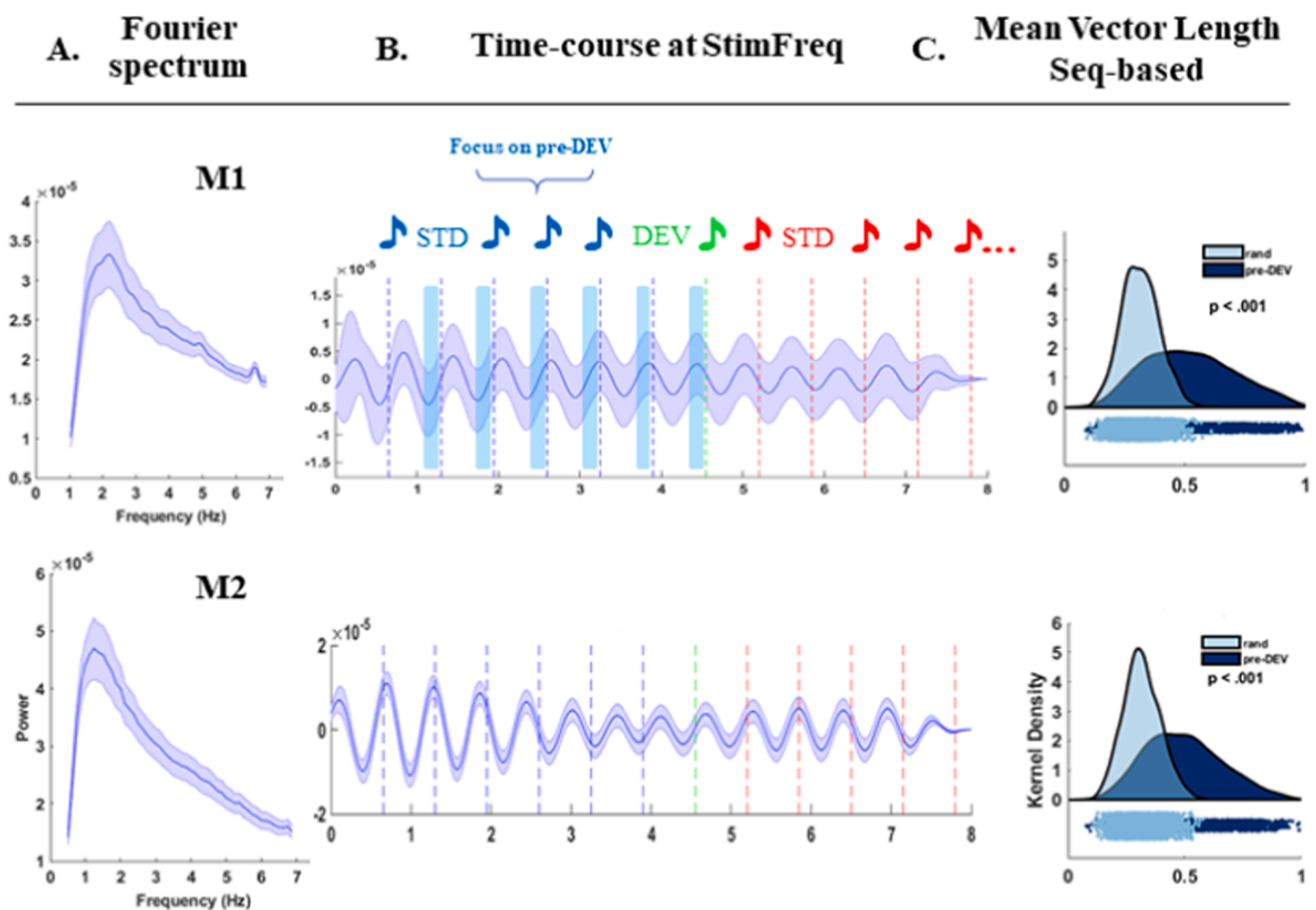


Fig. 2. Rhythm tracking analyses. A: Fourier spectrum of neural activity along the entire auditory sequence. The plot displays the grand-average power in the frequency range from 1 to 7 Hz. B: time-course of neural activity at the stimulation frequency. Vertical dotted lines indicate the onsets of STD tones prior- (blue) and post-DEV (red). The DEV onset is reported in green. Blue shades represent the standard errors. Light-blue rectangles indicate the pre-stimulus intervals of STD in which we performed phase analyses (not scaled). C: the kernel density distribution of mean vector length (MVL) calculated at the single-session and sequence-level and averaged across the fronto-central cluster of interest. These MVLs are based on the raw phase-angles for pre-DEV STD tones (blue) and are statistically compared to the MVL for random distribution of phase-angles. Single-session statistics are reported in Suppl. Table 2.

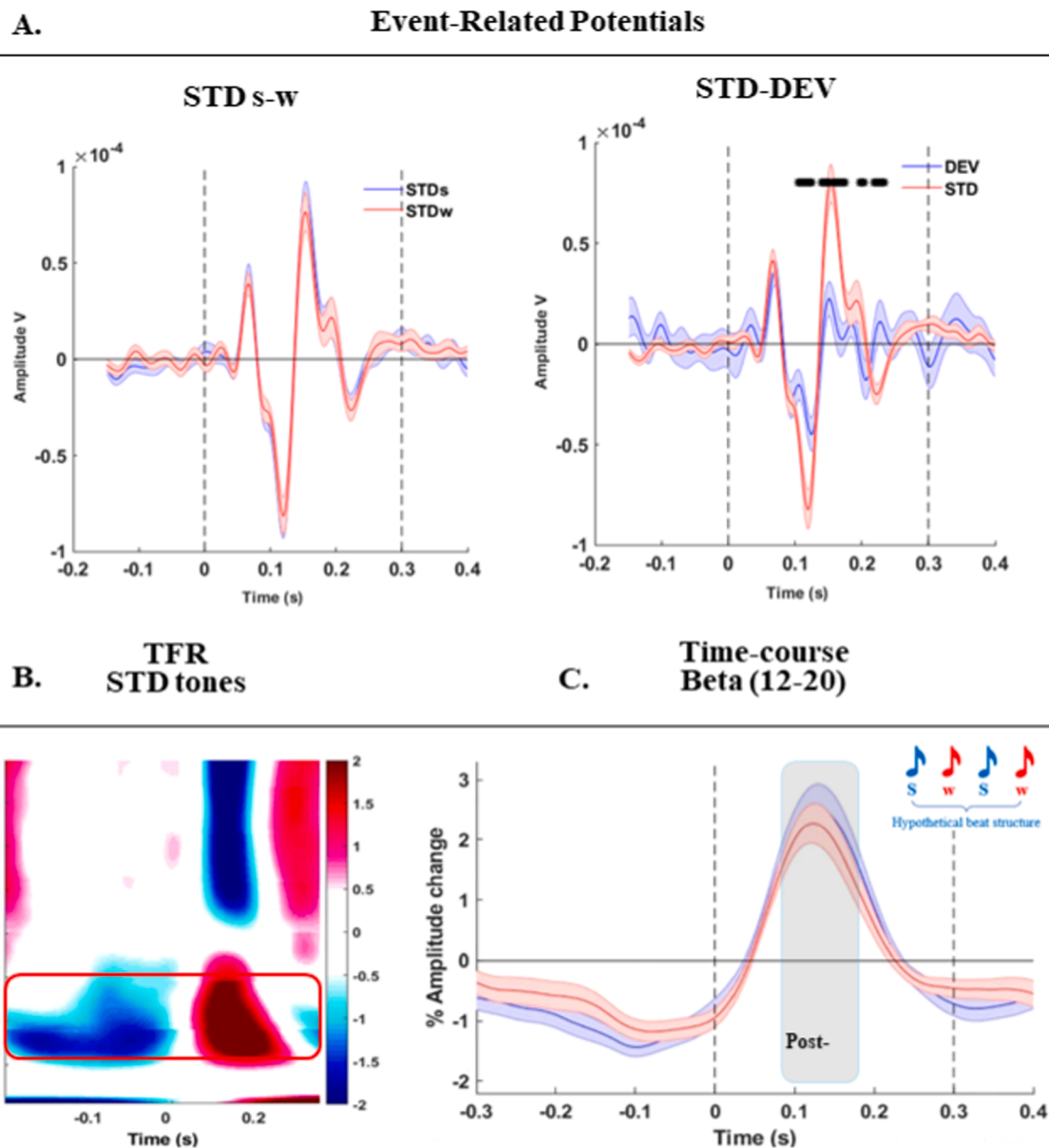


Fig. 3. ERPs and TFR data for monkey 1.

cubic temporal interpolation (using the *'pchip'* option for both the built-in Matlab and FieldTrip-based interpolation functions) considering the time-course of neighboring time-windows (extending up to 500 ms when possible, automatically reduced otherwise). The current approach is a data-driven procedure developed to minimize data loss. Rather than rejecting entire epochs only partly contaminated by artifacts (i.e., standard artifact rejection procedure), we opted for an artifact suppression approach that allowed keeping all trials. The channel-by-channel routine allowed the algorithm to flexibly adapt the outlier threshold estimates to the inherent noise varying over channels. Lastly, a standard whole-trial rejection procedure based on an amplitude criterion (85 μ V) was applied. Data selected for event-related-potential (ERP) analyses (“ERP data”) were segmented including 500 ms prior and following each tone onset (1 s in total). Data destined to time-frequency representation analyses (“TFR data”) were not further segmented at this stage. ERP data were band-pass filtered between 1 and 30 Hz and TFR data low-pass filtered at 40 Hz.

TFR data finally underwent time-frequency transformation by means of a wavelet-transform (Uno et al., 1993) with a frequency resolution of

0.25 Hz. The number of fitted cycles ranged from 3 for the low frequencies (<5 Hz) to 10 for high frequencies (>5 Hz and up to 40 Hz). TFR data were then re-segmented, so to reduce the total length to 2 s, symmetrically distributed relative to tone onsets.

2.4.2.6. Post-processing of ERP and TFR data. Single-trial ERP amplitudes were mean-corrected by a global average over epochs and 500 ms long (−0.2 to 0.3 s relative to tone onset). Similarly, single-trial TFR amplitudes were normalized by computing relative percent change with reference to the global mean amplitude across epochs and 500 ms long. This approach has been used elsewhere (Fujioka et al., 2012; Abbasi and Gross, 2020) and was preferred over baseline correction as we were interested in analyzing amplitude fluctuations in the pre-stimulus intervals. Then, we created a fronto-central channel cluster. All following analyses were performed exclusively on this channel cluster.

2.4.2.7. ERP analyses. We averaged evoked responses over trials separately for STD and DEV tones, and for odd (hypothetical “Strong” position in a binary accent; S, Figs. 3,4) and even (“weak”; w) serial

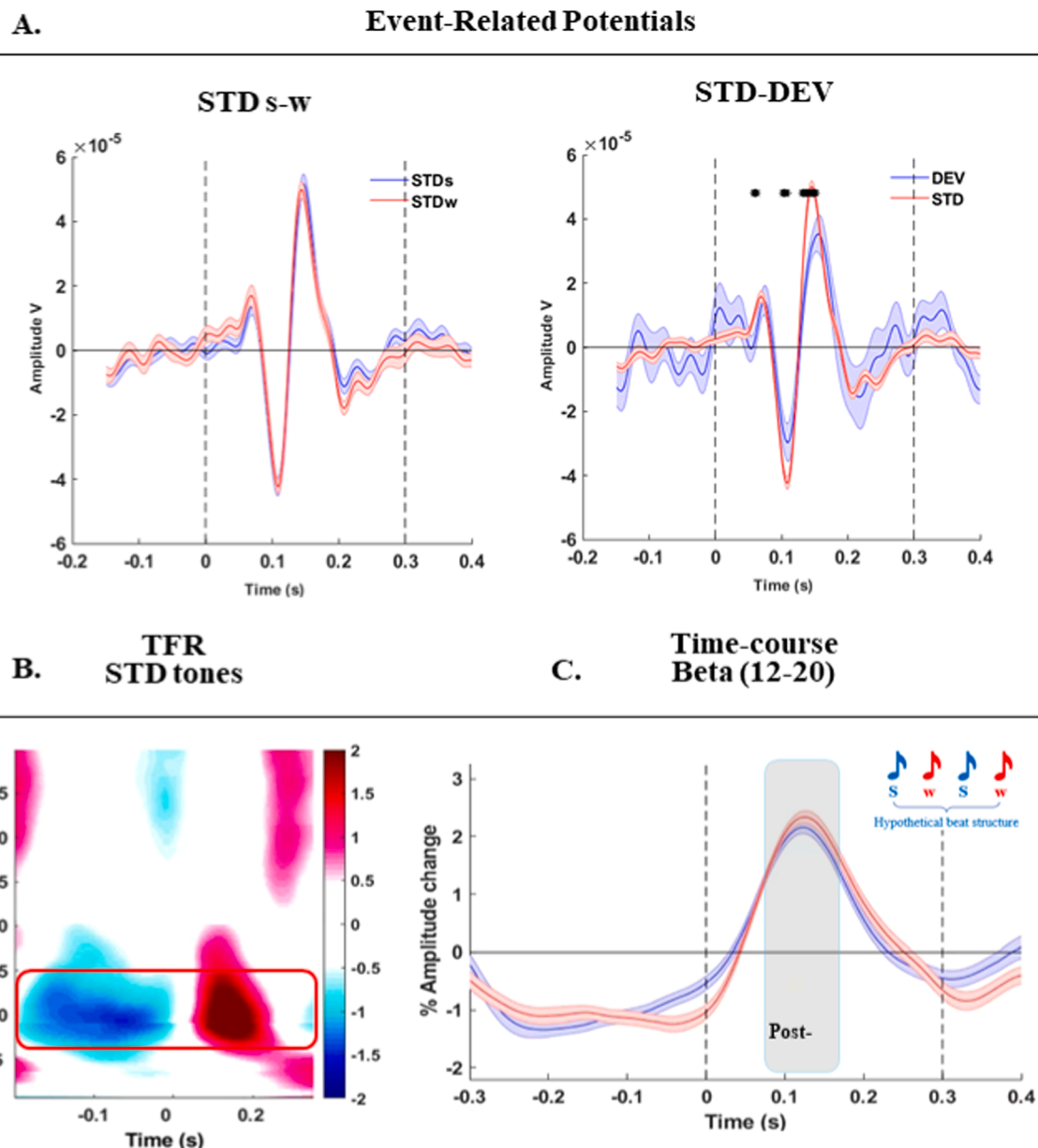


Fig. 4. ERPs and TFR data for monkey 2. **A:** On the left, ERP responses for STD tones in S (blue) w (red) positions. On the right, ERP responses for STD (blue) and DEV tones. Stars indicate significant time-windows, as assessed by means of paired-sample t-tests (FDR-adjusted $p < .05$). **B:** grand-average time-frequency spectrum time-locked to STD tones (-0.2 to 0.4 s). The frequency range spans 1–40 Hz with a frequency resolution of 0.25 Hz. The red rectangle highlights predominant responses in the mu (8–15 Hz) frequency range, on which we performed statistical comparisons. **C:** extracted time-course of mu-band activity in hypothetical S-w positions, time-locked to STD tones onsets, in blue for odd-numbered positions (Strong binary accent) and red for even-numbered positions (weak binary accent). Shaded colors indicate standard errors. On top, a grey rectangle delineates the time-window in which peak amplitude extraction is performed.

positions from the 3rd to the 11th tone. The first standard tones were not included to avoid increased responses typically observed at the beginning of an auditory sequence. Fig. 3 A and 4 A show the respective ERPs for the averaged fronto-central channel for STD tones on S-w positions (left), and for the comparison of STD and DEV tones averaged over S-w positions (right). Statistical analysis was performed by means of paired-sample t-tests. An FDR-adjusted p-value lower than .05 was considered statistically significant (Benjamini & Hochberg correction).

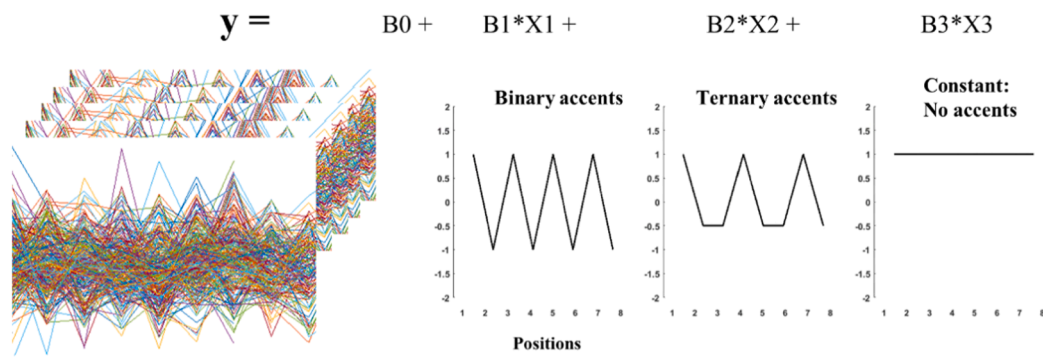
2.4.2.8. TFR analyses. We averaged time–frequency representations over STD trials, separately for odd and even positions (hypothetical strong and weak positions, respectively; Figs. 3B and 4B). Next, we quantified mean peak amplitudes in the mu-band (8–15 Hz) in the post-stimulus intervals (80 ms (proportional to the center frequency: 1/

12 Hz) and compared them for S-w positions (Fig. 3 C and 4 C).

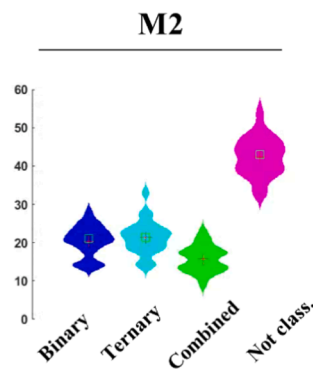
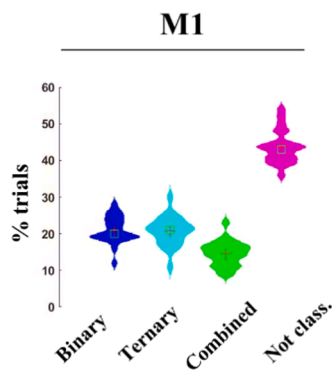
2.4.2.9. Individual classification of accents. An individual modelling approach was developed to identify binary accents. Since other accentuation patterns are possible (Abecasis et al., 2005) beyond the binary default (Brochard et al., 2003), the model further tested for the presence of ternary accents. We focused on single-subject, mu-band peak amplitudes for STD tones in the first 8 positions of the auditory sequence in 80 ms time-windows (proportional to the center frequency of interest; mu-band: $1/12$ Hz = 83 ms) following the stimulus onset.

Single-tone mu-band amplitudes were concatenated to mimic an 8-tone auditory sequence (i.e., a trial). Next, single-subject and trial-level mu-band fluctuations (8 tones) were entered into a stepwise regression model (Fig. 5 A; ‘stepwiselm’ in Matlab) with 3 predictors:

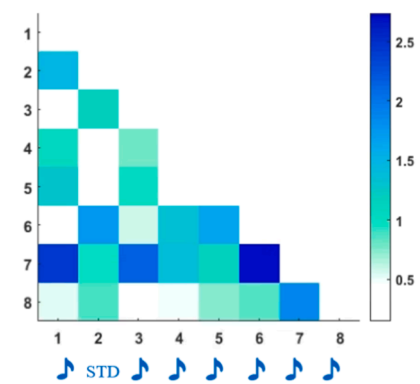
A. Identification of accents: step-wise regression modelling



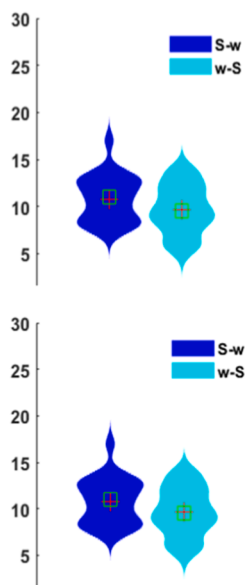
B. Preferences for accents



C. Binary pair-wise Amp difference



Binary accents

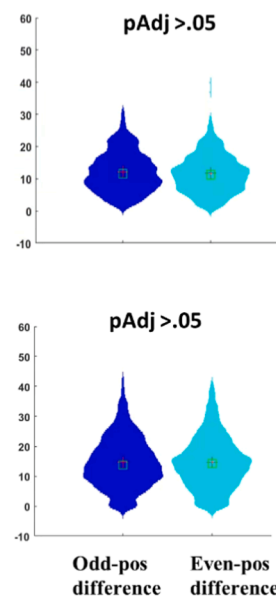


D.

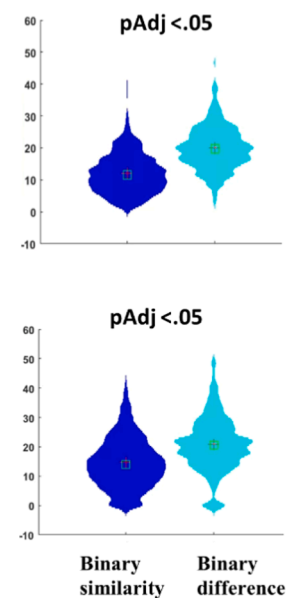
Binary similarity

M1

M2



Binary dissimilarity



(caption on next page)

Fig. 5. Modelling of individual accents and analyses on binary accents. A: The modelling of accents was performed by means of stepwise regression modelling and using mu-band post-stimulus responses as the dependent variable. The predictors were a binary (1, -1), a ternary (1, -0.5, -0.5) and a constant term (ones). B: preferences for accents, as reported from the modelling. In order, we plot the distribution of trials assigned to binary, ternary, combined (binary-ternary) accents, and 'not classified' (neither binary nor ternary) across the two monkeys. At the bottom, we zoom into binary trials and distinguish S-w accents from w-S accents based on trial-level Beta coefficients from the modelling (top for M1, M2 below). C: grand-average pair-wise difference for mu-band peak amplitudes across the first 8 positions of the auditory sequence in binary trials. D: on the left, the distribution of amplitude differences across odd-numbered positions (in blue) and even-numbered positions (cyan). The average of these two distributions forms the 'Binary similarity'. On the right, the 'binary similarity' (blue) and the mean amplitude difference of the odd- versus even-numbered position ('binary dissimilarity'; in cyan). Statistical testing was performed by means of 1000 permutations and an FDR-adjusted $p < .05$ was considered as statistically significant.

binary accents (values: 1, -1), ternary (1, -0.5, -0.5), and a constant term (ones). The winning model was chosen based on adjusted Eta-squared. Trials for which the winning model involved the binary predictor were labeled "binary". Similarly, trials for which the winning model involved the ternary predictor were labeled "ternary". The model thus allowed the combination of multiple predictors, but no interactions between terms. We accordingly interpreted (and labeled) the combination of binary and ternary terms as "combined". Lastly, trials in which grouping could not be clearly identified were labeled as "not classified". Session-level model results are provided in [Suppl. Table 2](#), designated as "Preferences for accents" and expressed as the percentage of trials relative to the full number of auditory sequences (100 per session, per monkey). The session-level goodness of fit of the model is provided in the same table. The "Preferences for accents" across sessions and monkeys are provided in [Fig. 5B](#).

2.4.2.10. Binary accent analyses. To confirm that the identified "binary" trials indeed showed binary accentuation patterns, we performed further analyses. First, we concatenated trials classified as "binary" and computed a trial-based single-tone pair-wise amplitude difference (its lower-triangle 2-D mean is provided in [Fig. 5C](#)). Namely, we calculated the amplitude difference between every tone (1–8 positions) along the auditory sequence independently for each sequence. Hence, we estimated the amplitude difference for the 1st and 2nd position, then to the third position, and so forth. Similarly, the amplitude for the 2nd position was compared to the 3rd, the 4th position, and so on. The result of this computation is a session-level pair-wise amplitude difference matrix, whose size is $N \text{ trials} \times N \text{ positions} - 1 \times N \text{ positions} - 1$. Next, we isolated the pair-wise amplitude difference for tones in odd-positions ([Fig. 5D](#); "odd-pos difference") and even-positions ("even-pos difference") and statistically compared them by means of 1000 permutations of odd-even labels. An FDR-adjusted p -value lower than .05 was considered statistically significant (Benjamini & Hochberg correction). These two variables were then combined into a distribution of "binary similarity". The binary similarity thus features the amplitude difference for tones in odd-numbered positions (1–3–5–7th) and the amplitude difference for tones in even-numbered positions (2–4–6–8th). In contrast, the binary dissimilarity was created by extracting the mean difference of tones in odd versus even positions ([Fig. 5D](#)). Finally, the binary similarity was statistically compared to the binary dissimilarity. Statistical testing was performed by means of 1000 permutations of odd-even labels, and an FDR-adjusted p -value lower than .05 was considered statistically significant (Benjamini & Hochberg correction).

2.4.2.11. Comparative analyses. In a prior study ([Criscuolo et al., 2023](#)) we investigated rhythm processing capacities in humans, using a comparable experimental paradigm and analyses. Human participants only took part in two experimental sessions, hence we only focused on the first two experimental sessions of the monkeys for comparative data analyses (see ([Criscuolo et al., 2023](#)) for further details regarding the analysis procedure adopted for the human dataset).

From the human sample, we randomly selected 4 participants (gender-balanced, as for the monkeys). In monkeys as well as in humans, we extracted the 'individual preferences for accents' as calculated in the accent modelling. Next, we isolated 'binary' trials, performed 'binary similarity' and 'binary dissimilarity' analyses as described in the 'Binary

accent analyses' section. Lastly, we inspected the time-course of neural activity time-locked to STD tones and compared it across monkeys and humans ([Fig. 6](#)).

2.4.2.12. Ternary accent analyses. Exploratory analyses zoomed in ternary trials. While the 'binary trials' could only show two accentuation patterns (S-w or w-S), ternary trials can show at least three different accentuation patterns, i.e., the accent can either fall on the first (S-w-w), second (w-S-w), or the third position (w-w-S). To disentangle these three accentuation patterns from the distribution of "Ternary trials" identified during the "Individual classification of accents", we ran a second stepwise regression model. This model featured three predictors corresponding to the respective accentuation types, implemented as: 1, - 0.5, - 0.5 (pattern 1), - 0.5, 1, - 0.5 (pattern 2) and - 0.5, - 0.5, 1 (pattern 3). The model did not allow interaction terms, and the winning model was chosen based on adjusted Eta-squared. The output of the model is provided in [Fig. 5B](#) (bottom right), as the percent distribution of three accentuation patterns across sessions and relative to the total number of auditory sequences (100 per session). Note that not all trials could be classified as pertaining to the three modelled accentuation patterns. Other accentuation patterns are possible in the ternary trials, which were not modelled here: for instance, a S-w-w pattern could be as well represented by a stair-case amplitude change (i.e., 1, -0.75, -0.25) or a shuffled version of it (i.e., 1, -0.25, -0.75). However, given that the stimulation rate in use is likely suboptimal to test ternary accentuations even in humans ([Brochard et al., 2003](#); [Abecasis et al., 2005](#); [Poudrier, 2020](#); [Baath, 2015](#); [Fujioka et al., 2012](#)), we did not build models to test all possible ternary accentuation patterns. Furthermore, it is important to note that the model in use here only uses the first 8 tones of the auditory sequence. This choice avoids the onset of DEV tones in later positions, which may disrupt ongoing accentuations, but inevitably leaves only up to two periods of a ternary accent (as compared to 4 repetitions of a binary accent). Consequently, even two small amplitude fluctuations with superimposed noise (inherent in EEG recordings) may drive the 'ternary' classification, but these trials may not necessarily reflect a true ternary accent. In turn, we expected a large proportion of 'ternary' trials to fail to be further classified as strictly reflecting the modelled patterns (1, -0.5, -0.5 (pattern 1), - 0.5, 1, - 0.5 (pattern 2) and - 0.5, - 0.5, 1 (pattern 3)). The small percentage of trials belonging to the three accentuation types (~2%) precluded further analyses due to insufficient statistical power to interpret results.

3. Resources and details

The datasets supporting the current study will be deposited in a public repository but are available from the corresponding author upon reasonable request. The code for analyses is available upon request. Further information and requests for resources should be directed to the Corresponding Author.

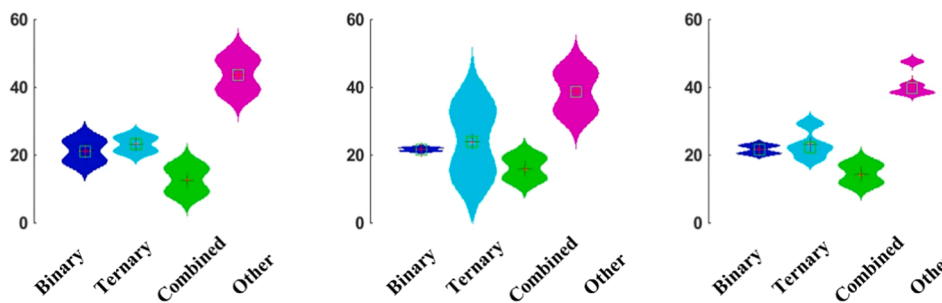
4. Results

4.1. Rhythm tracking

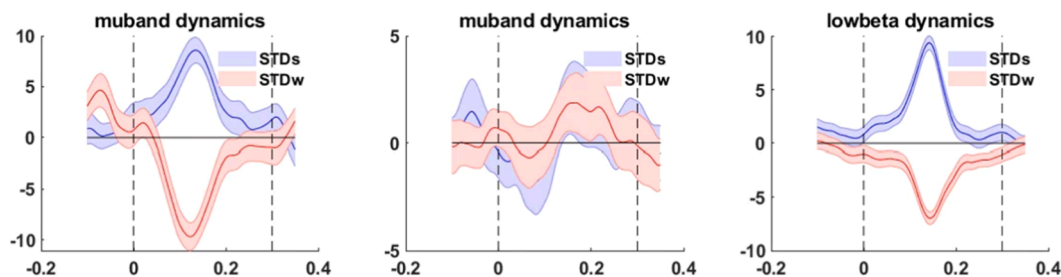
Macaque monkeys passively listened to isochronous equitone sequences presented at a stimulation rate of 1.67 Hz and containing 13-to-

A. Monkeys – Humans preferences for accents

	M1				M2				Humans				
	Binary %	Triplet %	Comb %	Other %	Binary %	Triplet %	Comb %	Other %	Binary %	Triplet %	Comb %	Other %	
Sess1	18	25	9	48	22	15	19	44	20	23	16	41	Pt. 3
Sess2	24	21	16	39	21	33	13	33	20	29	13	38	Pt. 4
									22	21	18	39	Pt. 5
									23	19	11	48	Pt. 7
Avg	22	23	12.5	43.5	21.5	24	16	36.5	21.3	23	14.5	41.5	Avg



B. Time-locked dynamics in binary trials



C. Binary accentuation effect

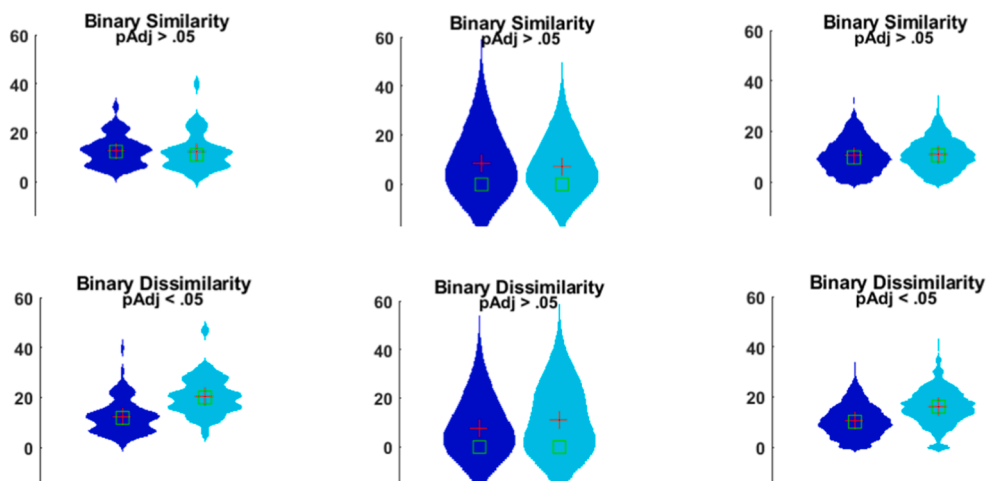


Fig. 6. Monkey-human similarities in rhythm processing. A: distribution of preferences for accents for, from left to right, M1, M2, and 4 human participants selected from a separate human dataset (Criscuolo et al., 2023). B: The time-course of time-locked neural dynamics in the mu-band for M1 and M2, and in the low-beta band for human participants. In blue, the time-locked responses to STD tones on Strong positions and red for weak positions. C: Binary accent effect quantified by means of Binary Similarity and Dissimilarity metrics. While M1 and the four human participants showed comparable binary accentuations from the beginning, M2 showed the same pattern later (Fig. 5). Of note, M2 only showed significant binary accentuations after two recording sessions. This observation differentiates the one depicted in Fig. 5, where we denoted significant binary accentuations when pooling across data from all recording sessions (>20). Thus, monkeys, similarly to humans, tend to vary in how they subjectively employ binary accentuation patterns over time.

15 frequent tones (standard; STD) and one amplitude-attenuated deviant tone (DEV). We tested *whether* and *how* their neural activity would show idiosyncratic signatures of rhythm tracking.

Macaque monkeys' neural activity encoded the timing of external events (Fig. 2 A, B top for M1 and bottom for M2). The Fourier spectrum showed a power peak at the stimulation frequency (1.67 Hz; Fig. 2 A), indicating that neural activity responded timely to tone onsets. Next, we quantified the consistency of pre-stimulus phase in delta-band (centered at 1.67 Hz) neural activity during a time-window preceding tone onsets (Fig. 2B) by means of mean vector length (MVL) analyses. This phase analysis focused on the ~60 ms (proportional to the stimulation frequency: $1/1.67 \text{ Hz}/10$) prior to tone onsets. Single-session trial- and channel-level MVL of STD tones preceding a DEV (pre-DEV) significantly differed from a random distribution (MVLs are plotted as kernel distributions in Fig. 2D; pre-DEV in blue; Suppl. Table 1–2 for statistics).

4.2. Accent processing

We tested whether participants' neural activity would sample the acoustic environment by superimposing binary accentuation patterns (S-w accents in odd-numbered versus even-numbered positions) onto the isochronous equitone sequences. Thus, we analyzed event-related responses (ERP) to STD tones in S and w positions, and further inspected the time-frequency representation of time-locked responses.

ERPs to STD tones in S-w positions did not differ statistically (Fig. 3 A, 4 A). However, STD tones elicited stronger N100 and P200 responses as compared to DEV tones (FDR-adjusted $p < .05$; Fig. 3 A, 4 A, right), confirming the processing of an unpredicted amplitude-attenuated deviant tone. The time-frequency representation plots of neural activity in response to STD tones mainly showed one event-locked response in the mu-band (8–15 Hz; Fig. 3B for M1 and 4B for M2). In this frequency band, we compared event-locked fluctuations for STD tones in odd (S) versus even (w) positions along the sequence (Fig. 3 C, 4 C), corresponding to hypothetical S-w positions (blue and red, respectively). The result of the statistical comparison of S-w positions did not survive FDR correction for multiple comparisons (FDR > 0.05).

To summarize, neither ERP nor TFR analyses revealed a binary accentuation, as STD tones elicited similar responses when they occurred in odd- and even-numbered positions along the auditory sequence. This result, however, may be associated with inter- and intra-individual differences in *when* and *how* accentuation patterns are superimposed onto the auditory sequences (Brochard et al., 2003). These hypotheses required a more adequate method to be tested: the novel accent modelling approach below.

4.3. Accent modelling

To address the questions of whether (i) monkeys accentuate in a similar way, (ii) they always accentuate, and (iii) accentuation patterns influence DEV processing (Brochard et al., 2003), we focused on trial-level data and modelled various accentuation patterns. We used a stepwise regression model to classify session-level, mu-band neural responses as best reflecting a binary, ternary, or absence of accentuation patterns. The model predicted tone-by-tone mu-band amplitude changes from three predictors: binary, ternary, and constant terms (Fig. 5 A). Resulting preferences for accents are reported in Suppl. Table 3 and summarized in Fig. 5B. Note that most trials (~60%) did not reflect either binary or ternary accents. In the absence of perceptual reports, we cannot confirm whether this observation signals lack of sensitivity of our method or whether monkeys did indeed not accentuate.

Next, we zoomed into “binary trials” and disentangled S-w from w-S accentuation patterns based on the single-trial β -coefficients obtained from the modelling (see methods). The resulting distributions are reported in Fig. 5B, bottom left. Similarly, we disentangled three possible accentuation patterns in the “ternary trials” (Suppl. Fig. 3). We

performed a separate stepwise regression modelling using S-w-w, w-S-w and w-w-S accents as predictors (see methods). Distributions are reported in Suppl. Fig. 3.

Overall, this approach allowed showing that macaque monkeys' neural activity spontaneously superimposes accentuation patterns on identical tones embedded in isochronous equitone sequences. Importantly, the monkeys' neural activity seems to switch between binary, ternary, and other accentuation patterns over trials. Notably, however, in the majority of trials no consistent accentuation pattern was identified.

4.4. Binary accents

After isolating trials showing binary accentuation patterns, we aimed to statistically test whether mu-band responses would significantly differ in S versus w positions. If so, neural responses to tones falling on odd-numbered positions should differ from those on even-numbered positions. However, there should be no differences for neural responses on the same positions: namely, tones falling on odd-numbered positions should elicit similar (i.e., non-significantly different) neural activity. In sum, we assessed whether this modelling approach delivers a meaningful classification of binary accents.

To this end, we isolated the binary trials and calculated the tone-by-tone pair-wise amplitude difference for mu-band post-stimulus activity across 8 positions in the auditory sequence and preceding the DEV tone. For visualization, the resulting matrix was averaged across trials and only the lower diagonal matrix is shown (Fig. 5 C). The original matrix (all trials) was instead used to calculate metrics of “Binary similarity” and “Binary dissimilarity” (Fig. 5D; top for M1 and bottom for M2). The Binary similarity features two distributions: the amplitude difference for tones in odd-numbered positions (1–3–5–7th; “blue”, labeled as “odd-numbered position difference”) and the amplitude difference for tones in even-numbered positions (2–4–6–8th; cyan, labeled as “even-numbered position difference”). The two distributions did not significantly differ from each other. The respective values were then combined to compute a “Binary similarity” variable. For the “Binary dissimilarity” analyses, we calculated the amplitude difference for tones in odd- versus even-numbered positions (corresponding to on-beat versus off-beat “Binary difference”) and statistically compared it to the “Binary similarity” (Fig. 5D). Statistical testing confirmed a significant difference (FDR-adjusted $p < .05$), indicating that mu-band post-stimulus amplitudes were significantly modulated according to a binary accent. The same procedure was independently repeated for both monkeys.

Taken together, these results confirmed that trials classified as ‘binary’ during the modelling did indeed show a consistent binary accentuation pattern. Mu-band amplitudes on STD tones in S positions significantly differed from those in w positions.

4.5. Comparative analyses of human and monkey data

Next, we set out to investigate similarities in rhythm processing between humans and macaque monkeys. We directly compared the two macaque monkeys with a subset of 4 datasets of participants taken from a prior human study (Criscuolo et al., 2023). Critically, these human participants underwent EEG recording while listening to similar stimulus material as the monkeys. This allowed us to reproduce the modelling of accents in use here and to directly compare the two datasets.

As human participants took part in only two experimental sessions, we focused on the first two experimental sessions of the monkeys as well. Details on the analysis procedure for the human dataset can be found in (Criscuolo et al., 2023). In Fig. 6 A, we show preferences for accents for M1, M2, and 4 human participants and below, the distribution of preferences for binary, ternary, and other accents (non-classified trials). Like humans, macaque monkeys showed the emergence of binary accentuations in 21% of the trials and ternary accentuations in 23% of the trials. From the selected binary trials, we then plotted the

time-course of event-locked activity in the mu-band for M1 and M2 and low-beta for humans, and for STD tones on S-w positions (blue and red respectively; Fig. 6B). Comparable to human participants, both monkeys showed larger amplitudes in response to STD-S tones than for STD-w tones (Fig. 5). However, while in M1 and human participants binary accentuations were evidenced by a non-significant Binary Similarity but a significant Binary Dissimilarity, M2 showed such an effect later (Fig. 6 C). We take this finding as additional evidence for the inter- and intra-individual variability across species in the *if*, *when*, and *how* accentuations occur.

5. Discussion

In this comparative EEG study, we set out to investigate the basic rhythm processing capacities of macaque monkeys and humans. All participants passively listened to isochronous equitone sequences, and we examined spontaneous neurophysiological activity that underlies the sampling of temporal regularities in the acoustic environment. We intentionally did not choose an active experimental task setting as it might enforce unspecific goal-directed behavior and potentially confound genuine endogenous rhythm processing. We further suggest that the testing of task-independent neural behavior in non-human primates might be a quintessential step in understanding the phylogenetic trajectories of *basic* rhythm cognition.

The present results show that the macaque monkey's neurophysiological responses display the encoding of temporal regularities in the acoustic environment even during passive listening and confirm prior task-active results (Zarco et al., 2009; Honing et al., 2018; Merchant et al., 2011; Crowe et al., 2014; Gámez et al., 2018; Ayala et al., 2017; Bartolo et al., 2014; Merchant et al., 2013b). We further show that these neurophysiological responses go beyond the mere encoding of isochrony: neural activity in the mu-band indicated the superimposition of binary (strong (S) - weak (w)) accentuations in a subset of trials, mirroring results in humans (Criscuolo et al., 2023). Even though all tones were physically identical, in some trials tone-locked neural responses were modulated by a binary (S-w) or ternary (S-w-w) *subjective* accentuation, resembling the well-documented *tic-toc* phenomenon observed in humans (Brochard et al., 2003; Abecasis et al., 2005; Schmidt-Kassow et al., 2011; Poudrier, 2020; Baath, 2015). As this phenomenon emerged during passive listening, the superimposition of accentuations might represent the spontaneous sampling of the acoustic environment beyond the encoding of single event onsets or time-intervals.

Standard ERP and time-frequency analyses, relying on the averaging of neural activity over hundreds of trials, failed to show these binary accentuation patterns. In comparison, our novel trial-based analysis increased sensitivity to such accentuation patterns. These combined observations also support earlier human studies that reported large inter- and intra-individual differences in *if*, *when*, and *how* participants accentuate (Poudrier, 2020; Baath, 2015). Hence, various accentuation patterns are possible: individuals may start accentuating at different time points (i.e., not necessarily from the beginning of an auditory sequence), may alternate accents over time (thus, over trials), or may not accentuate at all (if not instructed to do so). Thus, any trial-averaging procedure may inevitably overwrite these accentuation possibilities and mask out individual tendencies that influence the parsing of acoustic environmental rhythms.

While isochrony may not necessarily represent a (musical) rhythm, we propose it to be an ideal test-case for investigating the basic neurophysiology that underlies the encoding of temporal regularities. We further note that isochrony is present in a wide range of daily behavior, and its evolutionary advantage might lie in its simplicity: it allows generating temporal predictions (Ravignani and Madison, 2017). In turn, temporal predictability facilitates adaptive behaviours, rhythmic interactions, music, speech, and much more (Greenfield et al., 2021).

Complementing results from task-active settings (Lakatos et al., 2008; Schroeder et al., 2010), we show that macaque monkeys can

encode and track regularly timed event onsets with low-frequency neural oscillations, mirroring results in humans (Colling et al., 2017; Criscuolo et al., 2023; Nozaradan, 2014), see for an overview (Obleser and Kayser, 2019)). Along with alpha-/beta-band rhythms, delta-band oscillations have been associated with internalised timing and motor-to-auditory top-down predictions, even in the absence of overt motor tasks (Arnal, 2012; Fujioka et al., 2009; Abbasi and Gross, 2020; Bartolo et al., 2014; Saleh et al., 2010; Arnal and Giraud, 2012; Fujioka et al., 2015; Biau and Kotz, 2018; Engel and Fries, 2010). Both rhythms are prominently found in cortical motor (Keitel and Gross, 2016) and subcortical (striatal) brain regions (Bartolo and Merchant, 2015; Bartolo et al., 2014), and their functional coupling underlies complex auditory processing and temporal predictions in humans (Arnal et al., 2015; Morillon et al., 2019; Morillon and Baillet, 2017). Amplitude modulations in the alpha-beta frequency bands are associated with the priming of auditory brain regions via feedforward anticipation of incoming auditory input (Engel and Fries, 2010; Gehrig et al., 2012; Klimesch, 2012; Bowers et al., 2013; Liljeström, Mar et al., 2015). Thus, recurrent information flow between motor and auditory circuitries might underlie the emergence of simple to complex rhythm and beat structure in audition (Patel and Iversen, 2014).

Cross-species differences in *complex* rhythm and beat processing have been commonly associated with neuroanatomical and -functional differences in the motor system, specifically in cortico-basal-ganglia-thalamo-cortical (mCBGT) circuitry, which is more developed in humans than in non-human species (Patel and Iversen, 2014; Wilson and Cook, 2016; Mendoza and Merchant, 2014). However, the current findings might indicate that *basic* rhythm processing capacities not necessarily involve the mCBGT, or that they preceded neuroanatomical changes in the evolution of the human brain. However, given the absence of information on the neuroanatomical provenance in EEG signals, we refrain from speculations but motivate future studies to investigate this matter.

In summary, the current findings confirm that macaque monkeys have an adequate neural outfit to sample temporal regularities in the environment, and further show a human-like predisposition to parse regular acoustic input with accentuation patterns. These observations confirm that macaque monkeys have the fundamental building blocks that are necessary for DPS.

These unexplored parallels between humans and macaque monkeys motivate further cross-species investigations to advance better understanding of the phylogenesis of human rhythm cognition.

6. Conclusion

While passively listening to isochronous equitone sequences, macaque monkeys' neural oscillatory activity sampled the acoustic environment at multiple timescales. Delta- and mu-band oscillations encoded the temporal regularity in auditory sequences, tracked sound-onsets, and parsed them with a superimposed accentuation pattern. These observations mirror basic rhythm processing in humans and confirm a complementary role of low- (delta) and high- (mu) frequency bands. As these basic rhythm processing capacities are linked to the development of complex sensorimotor skills in humans (e.g., speech and music), these findings highlight a basic and fundamental steppingstone in the phylogenetic trajectories of humans' rhythm cognition.

CRediT authorship contribution statement

H.M., M.S., S.A.K. conceptualized the study. H.M., L.P., Y.A. collected data, performed, and interpreted audiogram test. A.C., M.S., S.A.K. planned and A.C. performed data analyses. A.C., H.M., M.S., S.A.K. wrote the manuscript.

Declaration of Competing Interest

The authors declare no competing interests.

Data Availability

Data will be made available on request.

Acknowledgements

This work was supported by Consejo Nacional de Ciencia y Tecnología (CONACYT) Grant CONACYT: A1-S-8430 and UNAM-DGAPA-PAPIIT IN201721 to HM. This work was further supported by the Van der Gaag Fund, Royal Netherlands Academy of Arts & Sciences to AC.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.pneurobio.2023.102502](https://doi.org/10.1016/j.pneurobio.2023.102502).

References

- Abbasi, O., Gross, J., 2020. Beta-band oscillations play an essential role in motor–auditory interactions. *Hum. Brain Mapp.* 41 (3), 656–665. <https://doi.org/10.1002/hbm.24830>.
- Abecasis, D., Brochard, R., Granot, R., Drake, C., 2005. Differential brain response to metrical accents in isochronous auditory sequences. *Music Percept.* <https://doi.org/10.1525/mp.2005.22.3.549>.
- Arnal, L.H., 2012. Predicting “when” using the motor system’s beta-band oscillations. *Front Hum. Neurosci.* 6 <https://doi.org/10.3389/fnhum.2012.00225>.
- Arnal, L.H., Giraud, A.L., 2012. Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16 (7), 390–398. <https://doi.org/10.1016/j.tics.2012.05.003>.
- Arnal, L.H., Doelling, K.B., Poeppel, D., 2015. Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cereb. Cortex* 25 (9), 3077–3085. <https://doi.org/10.1093/cercor/bhu103>.
- Ayala, Y.A., Lehmann, A., Merchant, H., 2017. Monkeys share the neurophysiological basis for encoding sound periodicities captured by the frequency-following response with humans. *Sci. Rep.* <https://doi.org/10.1038/s41598-017-16774-8>.
- Baath, R., 2015. Subjective rhythmicization: a replication and an assessment of two theoretical explanations. *Music Percept.* 33 (2), 244–254.
- Bartolo, R., Merchant, H., 2015. β oscillations are linked to the initiation of sensory-cued movement sequences and the internal guidance of regular tapping in the monkey. *J. Neurosci.* <https://doi.org/10.1523/JNEUROSCI.4570-14.2015>.
- Bartolo, R., Prado, L., Merchant, H., 2014. Information processing in the primate basal ganglia during sensory-guided and internally driven rhythmic tapping. *J. Neurosci.* 34 (11), 3910–3923. <https://doi.org/10.1523/JNEUROSCI.2679-13.2014>.
- Berens, P., 2009. CircStat: a MATLAB toolbox for circular statistics. *J. Stat. Softw.* 31 (10), 1–21. <https://doi.org/10.18637/JSS.V031.I10>.
- Biau, E., Kotz, S.A., 2018. Lower beta: a central coordinator of temporal prediction in multimodal speech. *Front Hum. Neurosci.* 12, 1–12. <https://doi.org/10.3389/fnhum.2018.00434>.
- Bowers, A., Saltuklaroglu, T., Harkrider, A., Cuellar, M., 2013. Suppression of the μ rhythm during speech and non-speech discrimination revealed by independent component analysis: implications for sensorimotor integration in speech processing. *PLoS One.* <https://doi.org/10.1371/journal.pone.0072024>.
- Brochard, R., Abecasis, D., Potter, D., Ragot, R., Drake, C., 2003. The “ticktock” of our internal clock. *Psychol. Sci.* 14 (4), 362–366. <https://doi.org/10.1111/1467-9280.24441>.
- Colling, L.J., Noble, H.L., Goswami, U., 2017. Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: an EEG study. *Front Neurosci.* <https://doi.org/10.3389/fnins.2017.00360>.
- Criscuolo, A., Schwartze, M., Henry, M.J., Obermeier, C., Kotz, S.A., 2023. Individual neurophysiological signatures of spontaneous rhythm processing. *NeuroImage.* <https://doi.org/10.1101/2022.03.14.484286>.
- Crowe, D.A., Zarco, W., Bartolo, R., Merchant, H., 2014. Dynamic representation of the temporal and sequential structure of rhythmic movements in the primate medial premotor cortex. *J. Neurosci.* <https://doi.org/10.1523/JNEUROSCI.2177-14.2014>.
- Engel, A.K., Fries, P., 2010. Beta-band oscillations—signalling the status quo? *Curr. Opin. Neurobiol.* <https://doi.org/10.1016/j.conb.2010.02.015>.
- Fitch, W.T., 2013. Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Front Syst. Neurosci.* <https://doi.org/10.3389/fnins.2013.00068>.
- Fujioka, T., Ross, B., Trainor, L.J., 2015. Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *J. Neurosci.* <https://doi.org/10.1523/JNEUROSCI.2397-15.2015>.
- Fujioka, T., Trainor, L.J., Large, E.W., Ross, B., 2009. Beta and gamma rhythms in human auditory cortex during musical beat processing. *Ann. N.Y. Acad. Sci.* <https://doi.org/10.1111/j.1749-6632.2009.04779.x>.
- Fujioka, T., Trainor, L.J., Large, E.W., Ross, B., 2012. Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *J. Neurosci.* 32 (5), 1791–1802. <https://doi.org/10.1523/JNEUROSCI.4107-11.2012>.
- Gámez, J., Mendoza, G., Prado, L., Betancourt, A., Merchant, H., 2019. The amplitude in periodic neural state trajectories underlies the tempo of rhythmic tapping. *PLOS Biol.* 17 (4), e3000054 <https://doi.org/10.1371/journal.pbio.3000054>.
- Gámez, J., Yc, K., Ayala, Y.A., Dotov, D., Prado, L., Merchant, H., 2018. Predictive rhythmic tapping to isochronous and tempo changing metronomes in the nonhuman primate. *Ann. N. Y. Acad. Sci.* <https://doi.org/10.1111/nyas.13671>.
- Gehrig, J., Wibral, M., Arnold, C., Kell, C.A., 2012. Setting up the speech production network: how oscillations contribute to lateralized information routing. *Front Psychol.* <https://doi.org/10.3389/fpsyg.2012.00169>.
- Greenfield, M.D., Honing, H., Kotz, S.A., Ravignani, A., 2021. Synchrony and rhythm interaction: from the brain to behavioural ecology. *Philos. Trans. R. Soc. B Biol. Sci.* 376 (1835) <https://doi.org/10.1098/rstb.2020.0324>.
- Hartig, R., Klink, P.C., Polyakova, Z., Dehaqani, M.-R.A., Bondar, I., Merchant, H., et al., 2023. A framework and resource for global collaboration in non-human primate neuroscience. *Curr. Res. Neurobiol.*, 100079 <https://doi.org/10.1016/J.CRNEUR.2023.100079>.
- Hattori, Y., Tomonaga, M., Matsuzawa, T., 2013. Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Sci. Rep.* <https://www.nature.com/articles/srep01566.pdf>.
- Honing, H., 2018. On the biological basis of musicality. *Ann. N. Y. Acad. Sci.* <https://doi.org/10.1111/nyas.13638>.
- Honing, H., Bouwer, F.L., Prado, L., Merchant, H., 2018. Rhesus monkeys (*Macaca mulatta*) sense isochrony in rhythm, but not the beat: additional support for the gradual audiomotor evolution hypothesis. *Front Neurosci.* 12 <https://doi.org/10.3389/fnins.2018.00475>.
- Honing, H., Merchant, H., Háden, G.P., Prado, L., Bartolo, R., 2012. Rhesus monkeys (*macaca mulatta*) detect rhythmic groups in music, but not the beat. *Dec 12 PLoS One* 7 (12), e51369. <https://doi.org/10.1371/journal.pone.0051369>.
- Iemi, L., Busch, N.A., 2018. Cognition and Behavior Moment-to-Moment Fluctuations in Neuronal Excitability Bias Subjective Perception Rather than Strategic Decision-Making. *eNeuro.* <https://doi.org/10.1523/ENEURO.0430-17.2018>.
- Jackson, L.L., Heffner, R.S., Heffner, H.E., 1999. Free-field audiogram of the Japanese macaque (*Macaca fuscata*). *J. Acoust. Soc. Am.* <https://doi.org/10.1121/1.428121>.
- Kaneshiro, B., Nguyen, D.T., Norcia, A.M., Dmochowski, J.P., Berger, J., 2020. Natural music evokes correlated EEG responses reflecting temporal structure and beat. *Neuroimage* 214, 116559. <https://doi.org/10.1016/J.NEUROIMAGE.2020.116559>.
- Keitel, A., Gross, J., 2016. Individual human brain areas can be identified from their characteristic spectral activation fingerprints. *PLoS Biol.* 14 (6), 1–22. <https://doi.org/10.1371/journal.pbio.1002498>.
- Kirschner, S., Tomasello, M., 2009. Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* 102 (3), 299–314. <https://doi.org/10.1016/J.JECP.2008.07.005>.
- Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci. Elsevier Curr. Trends* 16 (12), 606–617. <https://doi.org/10.1016/j.tics.2012.10.007>.
- Kotz, S.A., Ravignani, A., Fitch, W.T., 2018. The evolution of rhythm processing. *Trends Cogn. Sci.* 22 (10), 896–910. <https://doi.org/10.1016/j.tics.2018.08.002>.
- Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320 (5872), 110–113. <https://doi.org/10.1126/science.1154735>.
- Lakatos, P., Musacchia, G., O’Connell, M.N., Falchier, A.Y., Javitt, D.C., Schroeder, C.E., 2013. The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77 (4), 750–761. <https://doi.org/10.1016/j.neuron.2012.11.034>.
- Large, E.W., Gray, P.M., 2015. Spontaneous tempo and rhythmic entrainment in a bonobo (*Pan paniscus*). *J. Comp. Psychol.* 129 (4), 317–328. <https://doi.org/10.1037/COM0000011>.
- Large, E.W., Jones, M.R., 1999. The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106 (1), 119–159. <https://doi.org/10.1037/0033-295X.106.1.119>.
- Lasky, R.E., Maier, M.M., Snodgrass, E.B., Laughlin, N.K., Hecox, K.E., 1995. Auditory evoked brainstem and middle latency responses in *Macaca mulatta* and humans. *Hear Res.* [https://doi.org/10.1016/0378-5955\(95\)00140-7](https://doi.org/10.1016/0378-5955(95)00140-7).
- Laughlin, N.K., Hartup, B.K., Lasky, R.E., Meier, M.M., Hecox, K.E., 1999. The development of auditory event related potentials in the rhesus monkey (*Macaca mulatta*). *Dev. Psychobiol.* [https://doi.org/10.1002/\(SICI\)1098-2302\(199901\)34:13.0.CO;2-W](https://doi.org/10.1002/(SICI)1098-2302(199901)34:13.0.CO;2-W).
- Liljeström, M., Kujala, J., Stevenson, C., Salmelin, R., 2015. Dynamic reconfiguration of the language network preceding onset of speech in picture naming. *Hum. Brain Mapp.* 36 (3), 1202–1216. <https://doi.org/10.1002/hbm.22697>.
- Mendoza, G., Méndez, J.C., Pérez, O., Prado, L., Merchant, H., 2018. Neural basis for categorical boundaries in the primate pre-SMA during relative categorization of time intervals. *Nat. Commun.* <https://doi.org/10.1038/s41467-018-03482-8>.
- Mendoza, G., Merchant, H., 2014. Motor system evolution and the emergence of high cognitive functions. *Prog. Neurobiol.* <https://doi.org/10.1016/j.pneurobio.2014.09.001>.
- Merchant, H., Harrington, D.L., Meck, W.H., 2013a. Neural basis of the perception and estimation of time. *Annu Rev. Neurosci.* <https://doi.org/10.1146/annurev-neuro-062012-170349>.
- Merchant, H., Honing, H., 2014. Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Front Neurosci.* 7 <https://doi.org/10.3389/FNINS.2013.00274/ABSTRACT>.

- Merchant, H., Pérez, O., Zarco, W., Gámez, J., 2013b. Interval tuning in the primate medial premotor cortex as a general timing mechanism. *J. Neurosci.* <https://doi.org/10.1523/JNEUROSCI.5513-12.2013>.
- Merchant, H., Zarco, W., Pérez, O., Prado, L., Bartolo, R., 2011. Measuring time with different neural chronometers during a synchronization-continuation task. *Proc. Natl. Acad. Sci. USA.* <https://doi.org/10.1073/pnas.1112933108>.
- Morillon, B., Baillet, S., 2017. Motor origin of temporal predictions in auditory attention. *Proc. Natl. Acad. Sci. USA.* <https://doi.org/10.1073/pnas.1705373114>.
- Morillon, B.B., Arnal, L.H.L., Schroeder, C.C.E., Keitel, A., 2019. Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. *Neurosci. Biobehav. Rev.* 107, 136–142. <https://doi.org/10.1016/j.neubiorev.2019.09.012>. Dec 1.
- Nagasaka, Y., Chao, Z.C., Hasegawa, N., Tomonori, N., Naotaka, F., 2012. Spontaneous synchronization of arm motion between Japanese macaques. *Sci. Rep.* <https://www.nature.com/articles/srep01151.pdf>.
- Nozaradan, S., 2014. Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philos. Trans. R. Soc. B: Biol. Sci. R. Soc. Lond.* <https://doi.org/10.1098/rstb.2013.0393>.
- Obleser, J., Kayser, C., 2019. Neural entrainment and attentional selection in the listening brain. *Trends Cogn. Sci.* <https://doi.org/10.1016/j.tics.2019.08.004>.
- 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.* <https://doi.org/10.1155/2011/156869>.
- Patel, A., 2008. Music, language and the brain. *Brain* 131.
- Patel, A.D., 2006. Musical Rhythm, Linguistic Rhythm, and Human Evolution. *Music Percept* 24 (1), 99–104. <https://doi.org/10.1525/MP.2006.24.1.99>.
- Patel, A.D., 2014. The evolutionary biology of musical rhythm: was darwin wrong? *PLoS Biol.* <https://doi.org/10.1371/journal.pbio.1001821>.
- Patel, A.D., Iversen, J.R., 2014. The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Front Syst. Neurosci.* <https://doi.org/10.3389/fnsys.2014.00057>.
- Patel, A.D., Iversen, J.R., Bregman, M.R., Schulz, I., 2009. Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* 19 (10), 827–830. <https://doi.org/10.1016/J.CUB.2009.03.038>. May 26.
- Poudrier, É., 2020. The influence of rate and accentuation on subjective rhythmization. *Music Percept.* 38 (1), 27–45. <https://doi.org/10.1525/mp.2020.38.1.27>.
- Ravignani, A., Dalla Bella, S., Falk, S., Kello, C.T., Noriega, F., Kotz, S.A., 2019. Rhythm in speech and animal vocalizations: a cross-species perspective. *Ann. N. Y. Acad. Sci.* 1–20. <https://doi.org/10.1111/nyas.14166>.
- Ravignani, A., Madison, G., 2017. The paradox of isochrony in the evolution of human rhythm. *Front Psychol.* 8, 1820. <https://doi.org/10.3389/FPSYG.2017.01820/BIBTEX>.
- Saleh, M., Reimer, J., Penn, R., Ojakangas, C.L., Hatsopoulos, N.G., 2010. Fast and slow oscillations in human primary motor cortex predict oncoming behaviorally relevant cues. *Neuron.* <https://doi.org/10.1016/j.neuron.2010.02.001>.
- Schachner, A., Brady, T.F., Pepperberg, I.M., Hauser, M.D., 2009. Spontaneous motor entrainment to music in multiple vocal mimicking species. *Curr. Biol.* 19 (10), 831–836. <https://doi.org/10.1016/J.CUB.2009.03.061>.
- Schmidt-Kassow, M., Rothermich, K., Schwartz, M., Kotz, S.A., 2011. Did you get the beat? Late proficient French-German learners extract strong-weak patterns in tonal but not in linguistic sequences. *Neuroimage* 54 (1), 568–576. <https://doi.org/10.1016/j.neuroimage.2010.07.062>.
- Schroeder, C.E., Lakatos, P., 2009. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32 (1), 9–18. <https://doi.org/10.1016/j.tins.2008.09.012>.
- Schroeder, C.E., Wilson, D.A., Radman, T., Scharfman, H., Lakatos, P., 2010. Dynamics of active sensing and perceptual selection. *Curr. Opin. Neurobiol.* <https://doi.org/10.1016/j.conb.2010.02.010>.
- Selezneva, E., Deike, S., Knyazeva, S., Scheich, H., Brechmann, A., Brosch, M., 2013. Rhythm sensitivity in macaque monkeys. *Front Syst. Neurosci.* 7, 1–9. <https://doi.org/10.3389/fnsys.2013.00049>.
- Snyder, J.S., Large, E.W., 2005. Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cogn. Brain Res* 24 (1), 117–126. <https://doi.org/10.1016/j.cogbrainres.2004.12.014>.
- Uno, A., Kaga, K., Tsuzuku, T., Kuroki, M., 1991. Auditory Brainstem Response (ABR), Middle Latency Response (MLR), and Slow Vertex Response (SVR) in awake Japanese monkeys. *Audio Jpn.* <https://doi.org/10.4295/audiology.34.798>.
- Uno, A., Kaga, K., Tsuzuku, T., Kuroki, M., 1993. Middle-latency responses of awake and anesthetized Japanese macaques. *Int. J. Audio.* <https://doi.org/10.3109/00206099309072946>.
- Wilson, M., Cook, P.F., 2016. Rhythmic entrainment: why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychon. Bull. Rev.* 23 (6), 1647–1659. <https://doi.org/10.3758/s13423-016-1013-x>.
- Zarco, W., Merchant, H., Prado, L., Carlos Mendez, J., 2009. Subsecond timing in primates: comparison of interval production between human subjects and rhesus monkeys. *J. Neurophysiol.* 102, 3191–3202. <https://doi.org/10.1152/jn.00066.2009>.
- Zoefel, B., VanRullen, R., 2017. Oscillatory mechanisms of stimulus processing and selection in the visual and auditory systems: state-of-the-art, speculations and suggestions. *Front Neurosci.* 11, 1–13. <https://doi.org/10.3389/fnins.2017.00296>.