Amodal population clock in the primate medial premotor system for rhythmic tapping

Graphical abstract

Highlights
- MPC activity is recorded during isochronous tapping to visual or auditory metronomes
- The circular population dynamics form a loop for each interval and converge at the taps
- The tapping tempo is encoded in amplitude and temporal scaling of the neural trajectories
- The stimulus modality displaces neural trajectories without altering the timing mechanism

Authors
Abraham Betancourt, Oswaldo Pérez, Jorge Gámez, Germán Mendoza, Hugo Merchant

Correspondence
hugomerchant@unam.mx

In brief
Betancourt et al. investigate the neural basis of tapping entrainment to visual and auditory metronomes in the primate MPC. The population neural trajectories form a regenerating loop for each interval, converge at the taps in a similar state space, and encode tempo in their amplitude and temporal scaling.
INTRODUCTION

Humans have the natural ability to extract salient periodic events from sound sequences, called the beat or pulse, and to predictively align their movements to this pulse in music and dance. Importantly, these cognitive abilities depend on an internal brain representation of pulse that involves the generation of regular temporal expectations. Many studies have shown that the inter-representation of pulse that involves the generation of regular temporal expectations. Among them, the generation of regular temporal expectations is a key component of rhythmic entrainment. It is crucial for the brain to predict and entrain to the beat of a rhythm, especially in the context of isochronous movements.

The neural substrate for beat extraction and response entrainment to rhythms is not fully understood. Here, we analyze the activity of medial premotor neurons in monkeys performing isochronous tapping guided by brief flashing stimuli or auditory tones. The population dynamics shared the following properties across modalities: the circular dynamics of the neural trajectories form a regenerating loop for every produced interval; the trajectories converge in similar state space at tapping times resetting the clock; and the tempo of the synchronized tapping is encoded in the trajectories by a combination of amplitude modulation and temporal scaling. Notably, the modality induces displacement in the neural trajectories in the auditory and visual subspaces without greatly altering the time-keeping mechanism. These results suggest that the interaction between the medial premotor cortex’s amodal internal representation of pulse and a modality-specific external input generates a neural rhythmic clock whose dynamics govern rhythmic tapping execution across senses.

SUMMARY

The amazing human flexibility for perceiving and entraining to complex musical pieces with a sophisticated metrical structure seems to be species specific. Nevertheless, behavioral and electroencephalogram (EEG) studies in macaques have strongly suggested that monkeys possess all the audiomotor machinery to perceive and entrain to isochronous metronomes. Indeed, monkeys generate tapping movements in anticipation of the metronome and can flexibly change their movement tempo from trial to trial covering a range of 400–1,000 ms. Furthermore, monkeys can superimpose accentuation patterns onto an isochronous auditory sequence, suggesting that they can generate a simple subjective rhythm. These findings support the gradual audiomotor hypothesis that beat-based timing emerged gradually in primates, peaking in humans due to a sophisticated audiomotor circuit but also present for isochrony in macaques due to the close interaction between MPC, the basal ganglia, and the auditory cortex. Indeed, recent neurophysiological studies in monkeys indicate that the internal pulse representation during rhythmic tapping to visual metronomes depends on the neural population dynamics in the MPC. A key property of MPC neurons is their relative representation of beat timing. Cells that encode elapsed or remaining time for a tap show up-down ramping profiles that span the produced interval, scaling in speed as a function of beat duration. In addition, these cells are recruited in rapid succession, producing a progressive neural pattern of activation (called neural sequences or moving bumps) that flexibly fills the beat duration depending on the tapping tempo, thereby providing a relative representation of how far an interval has evolved. A critical aspect of the MPC beat-based clock is that it resets on every interval, providing an internal representation of pulse. The neural cyclic evolution and resetting are more evident when the time-varying activity of MPC neurons is projected into a low-dimensional state space. Population neural trajectories show
Figure 1. Rhythmic tapping behavior

(A) Synchronization task (ST). A trial started when the monkey placed its hand on a lever for a variable delay period. Then, a visual or auditory metronome was presented, and the monkey tapped on a button to produce five intervals following the isochronous stimuli of a particular duration. Correct trials were rewarded with an amount of juice that was proportional to the trial length. The instructed target durations were 450 and 850 ms. The last produced interval was not used for behavioral and neurophysiological analysis since it includes the expectation for reward delivery.

(B) Constant error ($\pm$ 5 x SEM) as a function of target interval during the auditory and visual condition of the ST. Inset: boxes indicate the ST condition; cyan, 450-ms short auditory (AS); blue, 850-ms large auditory (AL); orange, 450-ms short visual (VS); and red, 850-ms large visual (VL).

(C) Temporal variability (mean $\pm$ 5 x SEM as error bars) for each instructed interval and metronome modality, respectively (A, auditory, blue; V, visual, red).

(legend continued on next page)
the following properties: first, they have circular dynamics that form a regenerating loop for every interval produced. Second, they converge in a similar state space at tapping times, resetting the beat-based clock at this point. Finally, the periodic trajectories increase in amplitude as a function of the length of the isochronous beat.34,35

A fundamental unanswered question is whether the internal beat representation in the MPC works as a general clock across metronome modalities or whether visual and auditory periodic stimuli engage MPC neural populations with temporal processing dynamics that are modality specific. The classical notion of a common clock across timing contexts34–36 has been replaced by the hypothesis of a neural timing mechanism that includes both a core timing network rooted in the motor system and a set of areas that are selectively engaged depending on the specific requirements of a task.37–40 If the former notion is true, we should expect the neural trajectories in the MPC to have similar properties when monkeys execute rhythmic tapping sequences in synchrony with auditory or visual metronomes. In contrast, if the latter prevails, we should observe neural population dynamics with some shared properties but also temporal processing that is modality specific. To address these possibilities, we recorded the simultaneous activity of MPC cell populations while monkeys performed a synchronization tapping task guided by brief isochronous flashing stimuli or auditory tones.

RESULTS

Two monkeys were trained in a synchronization tapping task (ST). Each animal started by placing its free hand on a lever, kept it there for two stimuli (beat perception epoch), and then tapped a push button in response to the isochronous stimuli, producing five rhythmic intervals in a sequence (synchronization epoch; Figure 1A; see STAR Methods). Brief auditory or visual stimuli were used as metronomes with an interonset interval of 450 or 850 ms in blocks of 25 trials. The order of the four blocks of interval/modality combinations was random across days. We are assuming that monkeys solved this task by generating an internal representation of the metronome’s pulse that is coupled one-to-one with the tapping times.

We used the constant error (i.e., the difference between produced and instructed durations) and temporal variability (i.e., the standard deviation of produced intervals) to determine the accuracy and precision of the internal representation of the pulse based on the tapping times.18,19 A repeated-measures ANOVA on constant error revealed significant main effects for modality (F(1,87) = 15.57, p = 0.0002) and for the duration × modality interaction (F(1,87) = 4.1, p = 0.0457) (Figure 1B), but not for duration (F(1,87) = 2.33, p = 0.13) (Figure 1B). This indicates that, even when monkeys accurately produced intervals with errors close to zero, they slightly underestimated the intervals in the auditory condition, especially for the longer interval (t test for values different from zero: A450, p = 0.23; A850, p = 0.004; V450, p = 0.03; V850, p = 0.0004). The corresponding ANOVA on temporal variability showed significant main effects for target duration (F(1,87) = 39.53, p < 0.0001) and modality (F(1,87) = 146.49, p < 0.0001), as well as a significant effect for the duration × modality interaction (F(1,87) = 21.85, p < 0.0001) (Figure 1C). A Tukey honest significant difference (HSD) post hoc test showed a significantly larger temporal variability in the auditory than in the visual condition, accompanied by an increase in slope in the temporal variability as a function of interval (t test for slope different from zero: auditory, p < 0.0001; visual, p = 0.07). These results indicate that timing precision in monkeys follows the scalar property of timing with a larger slope for auditory metronomes.

Figure 1D depicts a large negative correlation for consecutively produced intervals (from serial order 1 to 4) for the target duration of 850 ms in the visual condition, indicating an error correction mechanism. Hence, we measured the autocorrelation of the inter-tap interval time series within trials and focused on its magnitude at lag 1. The lag 1 autocorrelation (Figure 1E) showed significant differences between modalities (F(1,87) = 15.08, p = 0.0009) and a marginal difference for target duration (F(1,87) = 3.4426, p = 0.078), but no statistically significant effect on their interaction (F(1,87) = 0.99, p = 0.33). Hence, these findings provide evidence of a stronger error correction mechanism for visual than auditory metronomes.

We further characterized the monkeys’ behavior during the ST by measuring the tapping movement kinematics. The speed profile of the hand movements captured by a high-speed camera was computed to determine the movement onset, the duration of each tap, and the dwell time between tapping movements in monkey M2 (see STAR Methods; Figure 1F). As previously observed in other animals during tapping tasks,3,9,41 the monkey showed a phasic stereotypic movement to push the button during the ST (Figure 1F), with two bell-shaped speed kinematics (corresponding to the downward and upward movements, respectively) divided by a low-speed point associated with the button press. Thus, using a speed threshold, we reliably identified five movements and four dwell times for each trial of our task (Figure 1F). Consistent with the notion of tapping stereotypy, movement times were similar across the four conditions (Figure 1G). In contrast, dwell times showed a large difference between shorter and longer target durations (Figure 1H). We carried out an ANOVA using the behavioral time-dependent parameter and epoch (movement vs. dwell time), duration, and modality as factors, which showed significant main effects for epoch.

(D) Correlation matrix of the produced intervals in the sequence (from serial order [SO] 1 to 4) within trials for the target duration of 850 ms in the visual condition (the Pearson-R minimum and maximum are −0.45 and 0, respectively). Note the large negative correlation for consecutive intervals suggesting an error correction mechanism.
(E) The lag 1 autocorrelation (±5 × SEM) for each instructed interval and metronome modality, respectively.
(F) Movement kinematics during ST. Speed profile of the hand movement (yellow trace) for the 450-ms interval of the auditory condition from the first to the fifth tap of a typical trial. Tap times (the instant of the button press) are represented as white dots. The monkey produced highly stereotypical tapping movements with a constant duration flanked by dwell periods whose duration scaled with the metronome’s tempo.
(G and H) Boxplots (median and interquartile values) for the movement (G) and dwell (H) times, respectively, for each instructed interval and metronome modality.
F(1,13) = 502.51, p < 0.0001), duration (F(1,13) = 4,560, p < 0.0001), and modality (F(1,13) = 20.12, p < 0.0001), and a large duration x epoch interaction (F(1,13) = 1,910, p < 0.0001). Indeed, the post hoc Tukey HSD showed a considerable increase in dwell time between the target durations of 450 and 850 ms for both modalities (p < 0.0001).

Overall, these findings support the idea that, during the ST, the monkeys used a complex rhythmic timing mechanism that includes at least four components (Figure 1): (1) a component for controlling the dwell to follow the metronome with different tempos, (2) another for triggering the internal beat signal that coincides with the tapping times, (3) a component that generates a command that initiates the stereotypic two-element tapping movement, and (4) a mechanism that serves as a mechanism for error correction within the rhythmic sequence. Furthermore, the metronome modality produced profound changes in rhythmic behavior, as timing was more precise and accurate, and tap synchronization that showed lowered error correction for visual rather than auditory isochronous stimuli.

Neural trajectories
General properties
The time-varying activity of a population of 1,189 MPC neurons that met the criteria of number of trials during the execution of the ST (see STAR Methods) was projected into a low-dimensional space using principal-component analysis (PCA) (see STAR Methods). The resulting neural trajectories in state space showed periodic dynamics that formed an elliptical regenerating loop for every produced interval with properties that were dependent on time encoding, target duration, modality, and serial order of the produced interval within the ST sequence (Figure 2A). We characterized these properties using geometric and kinematic approaches. Regarding the former, we first defined the subspaces for time encoding and each of the remaining three task

Figure 2. Neural population trajectories during ST and their oscillatory dynamic properties
(A) Projection of the neural activity in MPC (1,189 neurons) during ST onto the first three PCs. The first three PCs explained the 7.1, 4.1, and 4% of the total variance. We show the average trajectories across the four consecutively produced intervals included in the analysis. Each point in the trajectory represents the neural network state at a particular moment, where the trajectory completes an oscillatory cycle on every produced interval across conditions. Cyan, 450-ms auditory (AS); blue, 850-ms auditory (AL); orange, 450-ms visual (VS); red, 850-ms visual (VL). The green spheres indicate the tapping times across the trial sequence.
(B) Neural population trajectories projected into duration plane (white square), which explained 45.8%, 28.11%, and 25.9% of the variance for each PC. Target interval in milliseconds is color coded. A cube indicates the beginning of each trajectory, and an octahedron indicates the end.
(C) Tapping-time line separatrix, where the first three PCs explained 59.2%, 26.3%, and 14.3% of the variance. The almost orthogonal duration subspace is also shown as a gray square. The point in magenta corresponds to an arbitrary point used to compute the parameters depicted in Figure 3.
(D) Modality subspaces (two white squares). The first three PCs explained 58.4%, 38.9%, and 5.2% of the variance for auditory plane and 58.3%, 36.5%, and 5% for visual plane. Color code in (B–D) same as in (A).
(E) Serial-order subspace (white line that is a plane with an orthogonal angle) for VS condition explained 57.5%, 40.7%, and 1.7% of the variance. The color code goes from light to dark magenta for the first to fourth serial-order elements of the ST.
(F) General codification of the relative passage of time across the four conditions (white trajectories) as well as the overall state-space values for target duration (cyan and cinnamon diamonds), modality (red/blue asterisks), and tapping times (green spheres).
parameters by projecting the neural trajectories into second-level principal components (PCs) over the 3D neural trajectories. For example, the plane for the auditory modality (Figure 2D) was defined by projecting the data into the first three second-order PCs of the auditory neural trajectories in Figure 2A (blue and cyan trajectories; see STAR Methods).

Thus, the coding subspace for duration over time illustrated in Figure 2B shows periodic neural trajectories whose amplitude increased for longer target durations. This subspace (a plane defined by the second-order PCs 2 and 3) explained 96.7% of the variance of the trajectories and suggests the existence of a structured bimodal mechanism for tempo tracking where the amplitude of the rotary neural dynamics defines the time between tapping movements (Figure 2B). In addition, the neural trajectories showed some degree of temporal scaling, stretching for short and compressing for long target intervals. We computed the scaling index to measure the degree of scaling and found an index of 0.88, 0.72, and 0.78 for the first three PCs of the auditory condition, and 0.86, 0.71, and 0.72 for the visual condition. Fully scaled trajectories should show a scaling index of 1 (Figure S1). Hence, time encoding depends on a mixed strategy that combines changes in the amplitude and speed of the neural trajectories (see below for a detailed description of this phenomenon). Notably, the population neural trajectories also converged in a similar state space at tapping times, forming a line separatrix or boundary, that was defined distinctive planes for the auditory and visual conditions over the 3D neural trajectories. These subspaces had parameters that determine the four contexts of the ST, we computed their independent subspaces and the corresponding partial and mixed variances (see STAR Methods). Time encoding explains 38.6% of the total variance, forming circular regenerating loops for each produced interval (Figure 2F, white trajectories). This is a relative timing clock, giving a proportional measure of how far an interval has evolved within a circular revolution. Interestingly, a strong clustering of tapping times was observed when projected in this canonical rhythmic clock (Figure 2F green dots; mean resultant = 0.99, circular SD = 0.05; Rayleigh’s test, p < 0.001), emphasizing the notion of an attractor state for triggering the internal beat. Duration explained 2.7% of the total variance, with subspaces for short and long intervals that were slightly separated (Figure 2F, cyan and cinnamon diamonds) due to their increase in amplitude for longer target durations. Modality explained 40.8% of the variance, with distinct subspaces for visual and auditory conditions since the modality displaced the neural trajectories to a different region of the neural state space (Figure 2F, red and blue asterisks). Importantly, the mixed variance between time encoding, duration, and modality was 17.9%, indicating that the codification of time shared neural resources across durations and modalities. These findings validate the hypothesis of a strong rhythmic timing machinery in the MPC with cyclical neural trajectories that converge in a separatrix at the tapping time while counting the relative passage of time in each regenerating loop, independently of target duration, modality, or serial order. Nevertheless, the context also imprints specific signatures in the neural trajectories, as revealed next.

Kinematics of neural trajectories and modality effects
The kinematics of neural trajectories was characterized using the amplitude, angle, and relative position (Figures 3A–3C) between an arbitrary point in state space (fuchsia point in Figure 2C) and the neural trajectories across the four ST contexts shown in Figure 2A. We plotted these parameters as a function of the phase (relative timing) of each of the four produced intervals in the sequence (Figure 3). Notably, the amplitude, angle, and relative position showed cyclical patterns that were repeated across all serial-order elements of the ST sequence (Figures 3A–3C). It is important to note that similar kinematic properties were obtained when the arbitrary point was located within a contiguous state-space manifold (Figures S3A and S3B).

The amplitude of neural trajectories was significantly larger for long than short target durations in both the auditory and visual conditions as shown in Figure 3A. In fact, when we computed the area under the amplitude curve during the dwell and movement time periods, we found a significant duration-dependent increase in amplitude only in the former (Figure 3D). The autocorrelation lag-1 of the dwell amplitude showed negative values, especially for the visual 850-ms condition, as seen in the tapping behavior (Figure 3E). As expected, the angle acquired minimum values at the tapping times across conditions (Figure 3B), validating the existence of an attractor state that triggers the internal pulse signal when the trajectories reach a specific angular value (mean resultant = 0.99, circular SD = 0.11; Rayleigh’s test, p < 0.001).
Figure 3. Kinematic of neural trajectories
(A) Amplitude of trajectories computed as the Euclidean distance between an anchor point (fuchsia dot in Figure 2C) and each neural state of the trajectory (Figure 2C) across the four ST conditions. Inset: boxes indicate the ST condition as in Figure 2A. Dashed lines represent the tap phase for the four consecutive produced intervals.
(B) Angle computed from the dot product between the anchor point in (A) and the neural state of the trajectory.
(C) Position computed as the signed difference between an anchor point (fuchsia dot in Figure 2D) and neural trajectories for the four task conditions.

(legend continued on next page)

Cell Reports 42, 113234, October 31, 2023

6
This experiment had a block design, where each of the four conditions (short, long, auditory, and visual) was recorded for 25 trials before switching to the next in random order. Hence, the animals knew in which context they were performing the ST, and this knowledge could act as a tonic external input. In fact, modality information seems to displace the neural trajectories in different subspaces without greatly altering their cyclical organization, duration-dependent changes in amplitude, or tap attractor state behavior. This displacement was well captured by the relative position of the trajectories that corresponds to the signed difference of the distance between the arbitrary point and the trajectories in PC1. The position showed isomorphichanges within durations of the same modality but large and significant differences between the auditory and visual conditions (Figure 3C). In addition, the variability in the position of the trajectories followed an increase as a function of duration (Figure 3F) that is similar to the tapping variability of the monkeys (Figure 1C). Therefore, a modality-dependent tonic external input could diverge the cyclic neural trajectories to different subspaces, and the interaction between the amodal rhythmic clock and this external input could shape the differences in rhythmic timing between the auditory and visual metronomes.

To further scrutinize the role of temporal scaling in time encoding, we computed the speed of the neural trajectories across the four ST conditions. Figure 3G shows the complex temporal profile of the speed for the four intervals produced for each ST condition. Hence, the speed of neural trajectories during the ST does not work as a dial knob to encode a predicted interval using temporal scaling, as has been reported for single-interval perception. In fact, large repetitive peaks of speed occurred close to the tapping times, with a relatively steady level during the dwell between taps (Figure 3G). The overall ANOVA for speed showed significant main effects for epoch (F(1,199) = 1,368.9, p < 0.0001), duration (F(1,199) = 3,137, p < 0.0001), and modality (F(1,199) = 192.5, p < 0.0001) and a large duration × epoch interaction (F(1,199) = 728.8, p < 0.0001). Due to the large effects for epoch, we compared the speed of the neural trajectories between the movement and dwell times. During movement, the speed reached one or two peaks before the hand speed peaks around the tapping time (Figure 3H). This suggests that the speed of neural trajectories could encode the hand movement speed during the down-tap-up sequence. To test this hypothesis, we computed the Pearson correlation between the time series of the speed of both the hand and the neural trajectories at different lags (Figure 3L). We found large and significant correlations between the two measures during the movement time epoch where the speed of neural trajectories led the hand speed (A450, r = 0.68, p < 0.0001, lag = 160 ms; A850, r = 0.6, p < 0.0001, lag = 200 ms; V450, r = 0.72, p < 0.0001, lag = 140 ms; V850, r = 0.27, p = 0.01, lag = 160 ms). Therefore, these findings suggest that the speed of the neural trajectories during the movement epoch represents the speed changes in the stereotyped hand tapping movements during the ST (Figures 3H and 3L).

On the other hand, dwell speed showed a greater decrease between the short and long durations (Figures 3I and 3J). The decrease in speed as a function of duration through the dwell supports the hypothesis that time encoding during this critical task period depends on temporal scaling and, as described previously, on changes in the amplitude of the neural trajectories. Nevertheless, the scaling index of the PCs was smaller during the dwell than during the movement time. For the dwell, the scaling index was 0.66, 0.4, and 0.73 for the three PCs of the auditory condition and 0.59, 0.64, and 0.54 for the visual condition. For the movement time the scaling index was 0.9, 0.9, and 0.83 for the auditory condition and 0.72, 0.55, and 0.65 for the visual condition. This apparent contradiction is due to the geometry of the neural trajectories. Even if the speed difference between short and long durations is larger for the dwell (Figure 3I), so is the change in amplitude across target intervals during this epoch, making smaller the scaling index that focuses on the shape of the trajectories (Figures 3D and S1). Hence, to address this discrepancy, we developed an index that simultaneously determines the impact of amplitude and time scaling.
of the first three PCs of the neural trajectories, called the amplitude-modulation-time-scaling index (AMSI) (see STAR Methods and Figure S4). By definition, AMSI reached a value of −1 when the neural trajectory was fully temporal scaled and reached a value of 1 when the neural trajectories were fully amplitude modulated. We found that \( k_a \) attained large values after each tap, while \( k_v \) showed peak values before each tap (Figure 3K, top-middle). The AMSI was close to −1 at the beginning of the produced interval, increased monotonically within the interval, and reached values close to one just before the next tap (Figure 3K, bottom). Remarkably, AMSI was around zero in the middle of the interval, during dwell time. This behavior was similar between modalities, although the \( k_a \) showed a bimodal behavior within intervals for the auditory condition (see Figure S5 for the analysis on each monkey). These results indicate that the encoding of the passage of time during the dwell depended on an almost perfect balance between amplitude modulation and time scaling across modalities.

Finally, to gain insights into the properties of the trajectories related to an amodal clock and those related to the modality, we performed a canonical correlation using the elapsed time (fixed number of bins across durations), modality, and target duration as independent parameters and the trajectory position, amplitude, angle, and velocity as dependent parameters. The results showed that the three canonical variate pairs (independent versus dependent) were significantly correlated and dependent on one another, with a \( p < 0.0001 \) on the Wilks lambda, and the following squared canonical correlations: canonical-1 = 0.91, canonical-2 = 0.86, and canonical-3 = 0.3. The correlations between each parameter and the corresponding canonical variates revealed a clear structure between task and trajectory variables. The first canonical variate was linked to both modality (\( r = −0.86 \)) and position (\( r = −0.97 \)); the second was associated with elapsed time (\( r = −0.86 \)), angle (\( r = −0.93 \)), and speed (\( r = −0.47 \)); and the third canonical variate was associated with interval duration (\( r = −0.96 \)), amplitude (\( r = −0.96 \)), and speed (\( r = −0.96 \)) of the trajectories. These results confirmed that the main effect of the metronome’s modality was to displace the neural trajectories, acquitting different positions in the PC space without greatly altering their rhythmic clock properties.

**Individual session analysis of trajectories and behavior**

We carried out individual session analyses in the 14 sessions of monkey 2, as we had both enough simultaneously recorded cells and video analysis to characterize the tapping kinematics. This approach allows a detailed correlation between the trial-by-trial execution of the ST and the properties of the neural trajectories. We found a significant and large correlation (\( r > 0.5; p < 0.0001 \)) between the behavioral dwell times and the amplitude of the trajectories during the dwell in the 14 sessions (Tables S1 and S2; Figure S3G). In addition, the speed of the neural trajectories during dwell also showed significant correlations with the dwell times produced by the monkey (Table S1; Figure S3I). These findings corroborate the hypothesis that the amplitude and speed of the state-space trajectories in the MPC during the dwell constitute the mechanism that controls the pauses between movements to define the tempo in the ST.

The trial-by-trial variability of the relative position was highly correlated (Figure S3J) with the temporal variability in 13 of the 14 analyzed sessions of monkey 2 (Table S1), suggesting that the increase in slope of the scalar property for the auditory metronomes depended on the variability of the neural trajectories within the auditory subspace in accordance with our previous findings. Furthermore, the preference for visual metronomes in monkeys and the small associated temporal variability depended on the restrained variability of the neural trajectories within the visual subspace.

We also were interested in finding the neural correlates of error correction during the ST. The individual session analysis revealed that the lag 1 autocorrelation of the amplitude of the trajectories during the dwell was negative, especially for the 850-ms condition with a visual metronome. We found a large correlation in 12 of the 14 analyzed sessions with the autocorrelation of the tapping behavior across the four conditions (Figure S3L for session 2 of monkey 2, \( r > 0.2; p < 0.0001 \); Table S1). Next, we conducted a more refined analysis, where we first identified the recording sessions with a significant correlation between the produced interval and the amplitude of the trajectories for each of the four duration × modality conditions (Figure S6A). We found many sessions where the amplitude correlated with the slight changes in duration of the produced intervals within each condition (Figure S6A). These significant correlations showed a strong tendency to occur in time bins of the second half of the dwell periods (Figures S6C and S6D). Then, on these sessions, we computed the lag 1 autocorrelation of the amplitude of the neural trajectories and found significant correlations with the behavior (Figure S6B). Notably, these correlations were mainly observed in the first half of the dwell periods (Figure S6D). Similarly, many sessions showed significant correlations between the autocorrelation lag 1 of the produced interval and the autocorrelation lag 1 of the speed, with a similar dwell-related temporal profile (Figures S6E–S6H). Consequently, these results support the notion that the error correction for tap synchronization depends on the adjustments in the amplitude and speed of the trajectories during the dwell of consecutively produced intervals. They also suggest that the error correction signal occurred before the encoding of the dwell duration.

**Single-cell neural encoding**

We were interested in testing whether different aspects of the time-varying activity of single cells were related to the key parameters of the task and the properties of population trajectories. Initially, we carried out a four-way ANOVA using the discharge rate of a cell as the dependent variable, and the elapsed time (ET; 20 bins for each produced interval), the duration (Dur; 450 and 850 ms), modality (Mod; auditory vs. visual), and serial order (SO; 1 to 4) of the ST as factors (see STAR Methods). Notably, 86.5% of cells (\( n = 881 \)) exhibited significant main effects on elapsed time and/or the interaction of elapsed time with the other three factors (see Figure 4A).

In addition, Figure 4B depicts the Venn diagram for cells with significant effects for duration, modality, and/or serial order (see Figure S7 for individual monkey data). Since many neurons showed response modulations for multiple parameters (\( n = 967 \)), we tested whether MPC neurons showed mixed selectivity.
Figure 4. Properties of neuronal sequences
(A) Pie plots of the proportion of neurons with significant main effects (ANOVA, see main text) on duration, modality, and serial order that also had significant (blue) or nonsignificant effect of elapsed time(yellow).
(B) Venn diagram for number of cells with significant effects for duration (green), modality (magenta), and/or serial-order duration (purple).

(legend continued on next page)
Based on the ANOVA, we divided the neurons into three categories: (1) neurons with classical selectivity (namely no mixed selectivity), which exhibited only a main effect on one of four factors; (2) neurons with linear mixed selectivity, which exhibited main effects between at least two factors but nonsignificant interactions between them; and (3) neurons with nonlinear mixed selectivity, which showed significant interaction terms.

We found only 7.3% of classical selectivity (n = 75: 48 ET, 9 Dur, 13 Mod, 5 SO) and 5.5% of linear mixed selectivity (n = 56), while 82% of cells were nonlinear mixed selective (n = 836). Next, we used mutual information (MI) on the binned activity (20 bins for each produced interval) to determine the strength with which the cells represented elapsed time, duration, modality, and serial order on the cells with significant effects on the ANOVA. Most cells (94.5%) showed significant MI (n = 914/967; permutation test, p < 0.01) for at least one bin and one parameter, with many cells showing large MI above random (Figure 4C) for multiple parameters (Figure 4D). Prototypic examples of neurons with different selective profiles to modality and duration are shown in Figure 4E. In general, these findings reveal a large and strong mixed representation for the passage of time with the other task parameters in the MPC.

Sequential patterns of neural activation
Figure 4F shows the average normalized firing rate of the cell population with a significant MI for at least one ST parameter aligned to the bin of peak. Across the four task conditions, there was a dynamic activation of cells throughout the four produced intervals in a sequence, with smaller flanking peaks occurring at successive serial-order elements of the task. These cells were recruited in rapid succession, thus producing a progressive neural pattern of activation that flexibly filled each beat duration depending on the tapping tempo and providing a relative representation of how far an interval had evolved. The neural patterns are similar across subpopulations of cells (Figures S6A and S8B) but were not present when the activity maps were built using cells with no significant MI for the four task parameters (Figure S8C). To investigate which properties of the evolving patterns of activation were associated with the internal pulse-like representation of time and with duration, modality, and serial order, we determined the onset and extent of the activation periods for each cell using Poisson-train analysis (see STAR Methods). The activation periods of the cells were sorted by their time of peak activity with respect to the previous tap, producing a moving bump for each produced interval and defining four regenerating loops of activation patterns (Figure S5A). We divided the neural sequences into four quarters, starting after a tap with a group of cells, migrating during the timed interval for the second and third quarters, stopping with the last population quarter before the next tap, and simultaneously resetting to the initial set of cells for the next interval. Next, we used a polar plot representation to determine the number of neurons activated every 22.5° of a circumference representing the duration of the produced interval. The number of recruited cells significantly changed within an interval (Rayleigh's test, p < 0.0001, all four conditions), with progressively increasing cell numbers in the last two quarters of each cycle. These numbers peaked before every tap and dropped significantly for the initial segment of the next regenerating loop (Figure 5B). Therefore, the cyclical cell recruitment during a produced interval and the resetting of neural sequences on the next element of the ST sequence seems to be a neural population fingerprint of the rhythmic neural clock in the MPC across modalities and tempos.

A crucial question is whether this neural clock used a temporal scaling or an absolute encoding strategy. Under the temporal scaling scenario, the activation profile of a neuron is the same between target durations but shrinks for short and elongates for longer temps, with short and long activation periods, respectively. Thus, under this setting, a mean activation period of 200 ms for the target duration of 450 ms should produce an activation period of 377.7 ms for the 850-ms duration. Under the absolute timing strategy, the activation periods are the same across durations, but additional neurons are recruited for longer durations, so that the new neurons are active in the last portion of the interval. Hence, if 300 neurons were forming a neural sequence in the 450-ms target duration, we could expect an extra recruitment of 266 neurons in the 850-ms duration. Based on what we learned from the neural trajectories, it was not surprising to observe a mixed encoding strategy on the progressive neural patterns. The activation periods increased as a function of duration but not with full temporal scaling.

The mean durations for the auditory condition were 232.7 ± 33 ms (mean ± SEM) and 316.7 ± 54 ms for the 450- and 850-ms target durations, respectively. For the visual condition, the mean durations were 214.9 ± 23 ms for 450- and 300.6 ± 24 ms for 850-ms target durations. Therefore, the scaling indexes for the auditory and visual conditions were 0.72 and 0.74, respectively. On the other hand, the number of neurons recruited in the neural sequences was larger for longer durations, with 335 ± 22 and 398 ± 13 for the short and long intervals in the auditory condition, and 304 ± 10 and 377 ± 5 for the short and long intervals in the visual condition. Around 40% fewer neurons than expected comprised the neural sequences in the long durations if an absolute timing strategy was used. In addition to the shared mixed rhythmic timing strategy, the task modality also imposed changes in the properties of the evolving patterns of neural activity. The number of cells within the circumference of a produced interval showed statistically significant main effects

(C) Boxplot (median and interquartile values) of the MI for cells with significant permutation test on duration, modality, serial order, and elapsed time. The permuted values (median and interquartile values) are shown at the bottom of each parameter.
(D) Pie plots of the proportion of cells with significant MI for one to four task parameters. Note that most of the cells showed significant MI values for three or four task parameters.
(E) Prototypic examples of neurons with different selective profiles to modality (left) and duration (middle and right).
(F) Average normalized firing rate of cells (y axis) with a significant MI on at least one of the ST parameters displayed as a function of trial time for the fours task conditions (AS, AL, VS, VL). The four vertical black lines represent the tapping times. The cells were aligned to the bin of peak activity. Note the dynamic activation of cells throughout the four produced intervals in the ST sequence, with smaller flanking peaks occurring at successive serial-order elements of the task.
Figure 5. Dynamics of neural sequences

(A) Neuronal sequences within the progression of each interval in the rhythmic sequence for the short (450 ms, left) and long (850 ms, right) of the auditory condition. Each horizontal line depicts the beginning and end of the activation period of a cell. The color code divides in quarters the neural sequences during a produced interval (dark green, first; light green, second; yellow-green, third; yellow, fourth).

(B) Polar plots of the number of cells recruited at different phases within an interval for each of the four ST task conditions.

(C) Duration of the activation periods was larger for auditory than visual metronomes, and greater for longer tempos (ANOVA main effect modality, $F(1,51) = 36.01, p < 0.0001$; main effect duration, $F(1,51) = 541.09, p < 0.0001$; main effect quarter, $F(3,51) = 19.92, p < 0.0001$; nonsignificant modality x duration interaction, $F(1,51) = 2.62, p = 0.112$; modality x quarter interaction, $F(3,51) = 5.85, p < 0.001$; duration x quarter interaction, $F(3,51) = 9.03, p < 0.001$).

(D) Standard deviation of the peak time was larger for the 850-ms intervals of both modalities, with lower values around the tapping times (ANOVA main effect modality, $F(1,51) = 46.06, p < 0.0001$; main effect duration, $F(1,51) = 504.01, p = 0.0001$; main effect quarter, $F(3,51) = 19.58, p = 0.0001$; nonsignificant modality x duration interaction, $F(1,51) = 0.04, p = 0.84$; modality x quarter interaction, $F(3,51) = 2.81, p < 0.04$; duration x quarter interaction, $F(3,51) = 11.83, p < 0.001$).

(E) Discharge rate of the activation periods was slightly larger for the visual than the auditory condition, especially in the last two-quarters of the produced intervals (ANOVA main effect modality, $F(1,51) = 16.43, p < 0.0001$; nonsignificant main effect duration, $F(1,51) = 1.09, p = 0.301$; nonsignificant main effect quarter, $F(3,51) = 0.18, p = 0.90$).

(legend continued on next page)
for modality (chi-squared [2] = 50.2, p < 0.0001, Harrison-Kanji two-way circular ANOVA) and duration (chi-squared [2] = 90.1, p < 0.0001), as well as for the modality × duration interaction (chi-squared [1] = 6.7, p = 0.009). Thus, more neurons were recruited for longer intervals and for visual rather than auditory metronomes.

**Dynamics of neural sequences**

We determined how different properties of the neuronal sequences changed within the progression of each interval in the rhythmical sequence, using the same number of bins for each produced interval to compare task conditions. These parameters are the duration, intertrial standard deviation of the peak time, discharge rate, and Fano factor of the activation periods, as well as the neural recruitment lapse and number of neurons. Importantly, these parameters showed a cyclical variation within each produced interval that was repeated across the four serial-order elements of the rhythmical sequence (Figures 5C–5F). In fact, no statistically significant effects of serial order were found across all of them (ANOVA with duration, modality, and serial order as factors, p > 0.5). This is a remarkable phenomenon that corroborates that the neural clock is resetting for each interval, with the pulse of the metronome as the unit of measurement, not the absolute time across all the trials.

We carried out ANOVAs for each parameter using duration, modality, and the quarter of the produced interval as factors. We found that the duration of the activation periods was larger for auditory than visual metronomes and greater for longer tempos, with extended periods of activation in the middle two quarters of the 850-ms intervals (Figure 5G). Accordingly, the standard deviation of the peak time was larger for the 850-ms intervals of both modalities, with lower values around the tapping times (Figure 5D). Importantly, no evidence of an increase in peak response variability as a function of absolute time was obtained, rejecting the possibility of a Weber law scaling of response variability over absolute time.44 The discharge rate of the activation periods was slightly larger for the visual than for the auditory condition, especially in the last two quarters of the produced intervals (Figure 5E). The Fano factor, a coefficient of variation in the neural responses during the activation periods, was larger for the auditory than for the visual condition in the last two quarters of the produced interval (Figure 5F).

The neural recruitment lapse, which is the time between pairs of consecutively activated cells, showed a larger increase in the first quarter that plateaued until the third quarter and acquired lower values in the last quarter. The magnitude of this cyclical pattern was larger for the longer target interval of both metronome modalities (Figure 5G). Finally, the number of cells showed an initial decrease followed by a rebound that was steeper for the visual condition and a peak in the fourth quarter followed by a sharp decrease at the end of the produced interval. These results confirm a mixed coding strategy for rhythmic timing with time scaling, absolute timing, and modality-specific components.

We also performed a canonical correlation between the task parameters and the properties of the neural sequences to obtain an integrated notion of the relations between them. Specifically, we used the elapsed time, modality, and target duration as independent parameters, and the activation period duration, discharge rate, and Fano factor, as well as the neural recruitment lapse and number of neurons, as dependent parameters. The results showed a significant dependence between variate pairs (independent versus dependent) for the three canonical variates (p < 0.0001 for the Wilks lambda on the three canonical variates), with large squared canonical correlations: canonical-1 = 0.88, canonical-2 = 0.62, and canonical-3 = 0.43. Again, the correlations between the independent and dependent parameters with the canonical variates showed close and significant relationships. The first canonical variate was associated with both the target duration (r = −0.96) and the duration of the activation periods (r = −0.86); the second was associated with the elapsed time (r = 0.99), number of neurons (r = 0.95), and recruitment lapse (r = −0.8); and the third canonical variate was associated with the modality (r = 0.95), discharge rate (r = 0.45), and the Fano factor (r = −0.77). Hence, these findings confirm the strong relation between the tapping behavior and the neural sequences.

It is important to mention that the neurons within the neural sequences showed instantaneous activity changes that correspond to different types of ramping patterns that have been
reportd previously (see Figure S9). The neurons in the first quarter of the moving bumps showed the ramping profile of the swinging that encodes the beginning of each produced interval. The neurons of the second and third quarters showed an up-and-down profile of activation that is characteristic of the ramps that encode elapsed time in the width of the ramp in a rhythmic tapping task. Finally, the neurons of the last quarter showed ramping activity that is distinctive of cells encoding the time-remaining for an action, reaching a peak of activity at a particular time before the tap across conditions, with larger slopes for shorter durations (Figure S9).

**Generalizability of neural sequences**
We plotted the normalized activity of the 918 neurons that showed significant effects in the encoding across the four task conditions, sorting each condition according to the latency of peak activity for every unit of one task and using this sorting across tasks. Clear gradual patterns of activation were observed for self-sorted moving bumps (diagonal panels in Figure 6A), while cross-sorted sequences were quite different, with complex patterns of activity suggesting small generalizability of neural sequences between durations and modalities (off-diagonal panels in Figure 6A). We developed the distance index and the diagonal asymmetry index to determine how stable were the patterns of activation of neural sequences across task conditions (see STAR Methods). These two measures depend on the Euclidean distance matrix of the activity maps between pairs of task conditions as in Figure 6B. The distance index is a measure of the overall distance between the neural sequences of a pair of task conditions, whereas the diagonal asymmetry index is a cumulative measure of how congruent over time the two neural sequences within each produced interval of the ST are. The results showed that both the distance and the diagonal asymmetry indexes were larger between the modalities than within them, and also larger between the durations than within them (Figure 6B; see Figures S9A and S9B), suggesting strong effects of both task parameters in the cell configuration and dynamics of the neural sequences. Since we found larger diagonal asymmetry for intermediate bins than for the initial and final bins (Figure 6B), we also computed this index for the two periods. Indeed, the diagonal asymmetry index for intermediate bins was greater and statistically different from the corresponding index for initial-final bins (Figure 6D; F1,19 = 5.02, p < 0.048; Watson-Williams circular test), indicating that the neurons that were active within the interval were more context dependent than the cells after or before the taps. To further test this notion, we first computed the correlation of the activation profile between pairs of cells sorted as in Figure 6A. Interestingly, the values for the correlation matrix of self-sorted cells for the auditory short condition were not only large around the diagonal but also for the pairs of cells that were active at the beginning and end of the neural sequence (Figure S10A). Then, we computed the probability that pairs of cells showed large correlation values (r > 0.5) in their response profiles by dividing the complete neural sequence into quarters (Figure S10B). We found large probability values for the last two quarters of the moving bump across all the possible comparisons of task conditions, supporting the existence of a neural population that is similarly active near the next tap across durations and modalities (Figure S10B). Notably, we found 124 neurons with similar activation profiles across the four conditions and four serial-order elements of the ST, which were active mainly at the end of the produced interval (Figure S10C).

We also computed the distance and diagonal asymmetry indexes for neural sequences sorted by the serial-order elements within each ST condition. Figure 6C depicts the Euclidean distance matrices for neural sequences in the long auditory condition sorted by serial-order elements, with the minimum distance values as white dots. There is a systematic degradation in the neural sequence organization as the difference in serial order increases, accompanied by an increase in both indexes. In addition, the distance and diagonal asymmetry indexes were smaller for serial-order comparison than task conditions (see Figures 6D and S10D). We carried out a two-factor circular ANOVA using the diagonal asymmetry index as a dependent variable and the bin-epoch (intermediate vs. initial-final bins) and distance configuration (task vs. serial order) as factors (see Figure 6C). The results showed no significant main effects but a significant bin-epoch x configuration interaction (Figure 6D; chi-squared (1) = 10.2, p < 0.0014; Harrison-Kanji two-way ANOVA for circular data). In addition, the distance index was significantly greater for the task condition than for serial order but similar between bin epochs (ANOVA, distance configuration F(1,20) = 0.22, p = 0.88; bin-epoch F(1,20) = 13, p < 0.001; distance configuration x bin-epoch interaction F(1,20) = 0.35, p = 0.55). These findings suggest that there was a stable and quite repetitive pattern of activation in the moving bumps across the sequential structure of the task, which is different from the strong reconfiguration of the neural sequences between modalities and durations that especially occurred within the produced interval.

Overall, these findings support the theory that neural sequences during the ST form a congruent wave of neural activation that regenerates for each rhythmically produced interval, thereby engaging similar evolving patterns of activity across consecutive intervals. Furthermore, the moving bumps recapitulate many properties of the neural trajectories, including (1) cyclical patterns that are renewed across the sequential structure of the ST; (2) the generation of relative instead of absolute time representation; (3) the activation of larger groups of cells close to the taps signaling an internal pulse and then resetting the moving bump (most of these cells are active across durations, modalities, and serial-order elements); (4) the combination of an increase in the number of engaged neurons, greater recruitment lapses between neurons, and a rise in the duration of their activation periods as a neural foundation for the mixed representation of tempo in the amplitude modulation and temporal scaling of the neural trajectories; (5) the small generalizability between modalities and durations that suggests different inputs to the MPC for auditory and visual metronomes.

**Relationship between neural trajectories and neural sequences**
We simulated evolving patterns of population responses with specific profiles of activation and evaluated their conversion to state space by projecting the activity in three PC dimensions. This allowed testing different hypotheses regarding the effect of the key properties of the moving bumps (mentioned above)
on the AMSI and separatrix behavior for taps in the neural trajectories.

Initially, we simulated neural sequences with an increase in the number of neurons for the longer duration but a similar duration and magnitude of response between the two target intervals and no resetting in the moving bumps for taps. The resulting moving bumps corresponded to an absolute neural clock that the encodes elapsed time from the beginning to the end of the trial in the final position of the neural trajectories.49,50 There is no rotatory component since there is no resetting in the neural sequences,

Figure 6. Generalizability of neural sequences
(A) Gradual patterns of activation of self-sorted (diagonal panels) and cross-sorted (non-diagonal panels) moving bumps. Cross-sorted sequences show complex patterns of activity suggesting small generalizability of neural sequences between durations and modalities. Task conditions are depicted at the edges: AS, auditory short; AL, auditory long; VS, visual short; VL, visual long. The neural activity is depicted as discharge rate over time with 20 bins for each produced interval.
(B) Euclidean distance matrix between all possible pairs of task conditions using the sorted activity of (A). For the self-sorted conditions (diagonal panels) the Euclidean distance is zero. The minimum distance per bin is depicted as a white circle. Same conventions as in (A).
(C) Euclidean distance matrix between serial-order elements of the auditory long condition. Same conventions as in (B).
(D) Boxplot of the diagonal asymmetry index for initial intermediate and final bins using the Euclidean distance across task conditions (left, from B) and serial-order elements of the auditory large condition (right, from C).
Figure 7. Neural sequence simulations and their resulting neural trajectories
First two left columns: heatmaps of the normalized activation profiles for short and long intervals. Middle: neural trajectories for the short (cyan) and long (dark blue) conditions. Right: AMSI as a function of time or tapping phase. (A) Neural sequences for a single interval with the double number of neurons for the long interval but the same duration and magnitude on the responses. (B) Same as in (A) but for two resetting neural sequences simulating two chains of activation for two rhythmic intervals. (C) As in (B) but with the same number of neurons, a temporally scaled activation period, and same response magnitude. (D) As in (C) with
(legend continued on next page)
and the corresponding AMSI showed values above 0, indicating changes in amplitude (right column in Figure 7A) but not in temporal scaling. Next, we simulated absolute and resetting moving bumps for two produced intervals, with cyclical trajectories that showed no temporal scaling just changes in amplitude (AMSI with values around 1) (Figure 7B). Conversely, relative resetting neural sequences, with the same number of neurons and a scaled increase in the response duration, produced cyclical neural trajectories with AMSI around −1, indicating pure temporal scaling and no changes in amplitude (Figure 7C). If, instead of changing the duration, we changed the magnitude of the response on the longer interval in the previous simulation, the AMSI would show values around 1, indicating amplitude modulation (Figure 7D).

Hence, changes in the duration of the activation period produce temporal scaling, while changes in the number of cells and response magnitude produce shifts in the amplitude of neural trajectories, as seen in the empirical data (Figures 3 and 5).

We found that sharing neurons at the beginning and end of each produced interval across target intervals, emulating the observed moving bump organization (Figures 6 and S10), induced a convergence of the neural trajectories at tap times and, hence, a separatrix behavior (Figure 7E). Indeed, the distance of the trajectories at tapping times is zero or very small when simulations include a shared population of cells at the beginning and end of the interval (Figure S11). These findings confirm the notion that the neural internal pulse representation as the tap separatrix depends on the activation of a group of neurons whose activity flanks the interval produced, responding similarly across durations, modalities, and serial-order elements. Adding to this simulation, an increment in the duration of the activation periods, as well as a larger number of neurons in the intermediate epoch of the interval for longer durations, produced neural trajectories that were similar to the original population dynamics (Figures 7E and 7F). These neural trajectories were circular, converged at tapping times, and showed cyclical variations of AMSI with values close to zero in the middle of the produced interval, indicating a mixture of temporal scaling and amplitude modulation for time encoding. This combined simulation supports the deep relationship between the properties of the original neural trajectories and their corresponding neural sequences. Finally, simulating neural sequences with the same set of neurons (Figure 7E), partially overlapping cells (Figure 7F), or duration-selective populations (Figure 7G) produced an angular difference in the circular trajectories that was between zero (same neural populations) and 90° (completely different sets of cells). These results suggest that the angular difference in the subspaces of the auditory and visual conditions was mainly due to the partially overlapping bimodal cells of our database.

**DISCUSSION**

The parametric account of the rhythmic behavior of the animals during the ST, combined with the geometric and kinematic assessment of the neural trajectories and the detailed analysis of the properties of the neural sequences, revealed a series of fundamental principles governing the neural rhythmic clock in the MPC. On one side, there is an amodal representation of beat-based timing that includes at least three components. First, the trajectories converge in a similar state space at tapping times, while the moving bumps restart at this point, resetting the beat-based clock. The tap separatrix and the neural resetting is a neural correlate of the internal pulse that coincides with the tapping times, providing a phasic representation of a cyclic time event, as well as a continuous and relative reading of how much time has passed within each produced interval. Second, the tempo of the tap synchronization depends on the dwell between stereotyped movements. This dwell is encoded by a combination of amplitude modulation and temporal scaling in the neural trajectories, which, at the moving bump level, correlate with a mixture of an increase in the number of engaged neurons, larger recruitment lapses between neurons, and a rise in the duration of their activation periods. Third, the mechanism for error correction that maintains tap synchronization with the metronome depends on the within-trial changes in amplitude and speed of the trajectories during the dwell period. Conversely, the modality of the metronome produced profound changes in the monkeys’ rhythmic behavior, with timing that is more precise and accurate, and a tap synchronization that showed larger error correction when using visual rather than auditory isochronous stimuli. Therefore, the modality imprints specific signatures in the neural trajectories, with a large displacement in state space that does not greatly alter their cyclical organization, duration-dependent changes in amplitude and temporal scaling, or tap separatrix behavior. These findings suggest the existence of a modality-dependent tonic external input that produces a divergence in the cyclic neural trajectories to different subspaces. Concordantly, we found modality selective cells and a lack of generalizability in the neural sequences between modalities.

The notion of an internal representation of pulse implies the existence of a phasic signal that is a cognitive construct of regular temporal expectations based on the properties of the input sequence. The dynamic attending theory has been popular in the literature. This theory postulates that rhythmic temporal expectancy depends on pulses generated by coupled oscillators. Models using coupled oscillators can explain beat perception in humans for a wide variety of rhythms with complex metric structures. Nevertheless, this approach is built on the strong assumption that the brain works as a generator of long-lasting oscillations that can be coupled in time to represent the pulse. Instead, cortical and subcortical oscillatory activity usually occurs in short bursts (200–1,000 ms) and depends on local inhibitory mechanisms that generate alternating temporal windows of enhanced and decreased excitability. Importantly, this inhibition-based mechanism produces natural temporal windows to group neuronal activity into cell populations and neural sequences. Here, we offer a neurophysiological account for the same number of neurons, same response duration, and larger response amplitude for the long interval. (E) Neural sequences sharing neurons at the beginning and end of each produced interval and a temporally scaled response. (F) Same as in (E), with a temporally scaled activation period and larger number neurons for the long interval. (G) Neural sequences with the same duration and shared neurons at the beginning and end of the two produced intervals but with partially overlapping cells within intervals.
the interval representation of pulse where the neural population generates a cyclic pattern that provide a continuous and relative reading of how much time has elapsed within each produced interval in the rhythmic sequence. This relative time encoding is observed in both neural trajectories and neural sequences. In addition, the line separatrix in the neural trajectories at tapping times is a robust neural correlate for the internal beat. An ideal reader in state space can naturally generate a tick pulse every time the neural trajectories reach the separatrix. In addition, the sudden increase in number of cells just before the tap is a crucial event that could also signal the internal prediction of the metronome’s pulse (Figure 5H). A recent study modeled internal pulse prediction as a probabilistic point process where an observer is continuously inferring the phase of the pulse based on phasic temporal expectation template that was modeled as the sum of Gaussians.51 Crucially, the large recruitment of neurons before the tap can be seen as the temporal expectation signal, where the variability of the predicted times was greater for auditory than visual metronomes and grew as well for longer target intervals. Moreover, the strength of the expectations was greater for longer tempos (Figure 5H). These properties correlated with greater temporal variability for the auditory condition and high accuracy in temporal production for longer intervals (Figure 1).

Many MPC cells were active before every tap across serial orders and modalities, which coincides with the wide generalization of the neural sequences before the tap (Figure 6C). Thus, the line separatrix that is orthogonal to the circular loops in the neural trajectories (Figure 2C) may depend on this last cell type. In addition, this cell population might also be involved in the encoding of the hand movement during tapping, since we found that, around 200 ms before the tap, the neural trajectories sustain changes in speed that represent the changes in hand kinematics. Hence, we suggest that there is a tight coordination in the MPC between the motor commands for the tapping movement and the precise timing of the button press.50,61 The latter is the primary parameter for reward and is probably also a timing signal that directs the rhythmic performance of the monkey.82

We have shown that, independently of the modality, monkeys produce rhythmic intervals in synchrony with a metronome by controlling the dwell time between stereotyped tapping movements.19,61 The temporal control of dwell duration depends on a dynamic combination of temporal scaling and amplitude modulation in the neural trajectories. Temporal scaling as a timing mechanism has been reported in the MPC,26,27,42,63 prefrontal cortex,46,64,65 parietal cortex,38,66 and basal ganglia.79,80 The main concept behind scaling is that the same population of cells represents time to an event (normally a movement) by shrinking or expanding the neural response profile for short or long intervals, respectively.52,49,71 Therefore, the speed of the neural population clock is set at a constant level in order to predict and produce a specific duration in single-interval tasks.72 In contrast, the speed of the neural trajectories during rhythmic tapping is not constant, reaching a large peak at the tapping time across tempos but showing a decrease for long durations during the dwell (Figures 3E and 3F). Furthermore, the temporal scaling of the neural trajectories during the dwell was accompanied by an amplitude modulation in their circular changes in state space.83 Importantly, both the amplitude and the speed change are robustly correlated with the intervals produced by the monkeys. Hence, the mixed coding strategy where neural populations combine temporal scaling and amplitude modulation seems to be a specific signature of rhythmic timing. At the neural sequence level, the changes in the duration of the cells’ activation periods were related to the temporal scaling, while the number of cells in the sequences and the increase in their recruitment lapse were linked with the amplitude modulation of the neural trajectories. Furthermore, the changing numbers of cells recruited in the moving bumps is an indication of interval tuning. Interval tuning during single-interval and beat-based timing has been reported in the MPC,73,74 prefrontal cortex,46 the putamen,75,76 the caudate,77,78 and the cerebellum.19,85 In addition, a chronomap in the MPC has been described in humans using functional imaging, where interval-specific circuits show a topography with short preferred intervals in the anterior and long preferred intervals in the posterior portions of the SMA/pre-SMA.18 Hence, timing depends not only on one population of cells that contracts or expands their activity patterns depending on a constant speed knob but also on interval-specific neurons that generate distinct timing circuits. Tuning and modularity are mechanisms for the division of labor that are widely used in cortical and subcortical circuits to represent sensory, cognitive, and motor information.82–84 Interval tuning can provide large flexibility to encode the passage of time and predict events across behaviors that require the integration of timing with other task parameters that have a different mapping framework in MPC.87,88 Since the width of cell timing is wide, interval-tuned neurons can also show temporal scaling,28,38,46 which can be the substrate of the observed mixed timing encoding that combines amplitude modulation and temporal scaling.

It has been shown in humans that performance to an auditory metronome is more precise and accurate than synchronization to a flashing visual metronome.14,85–91 Conversely, macaques showed a bias toward flashing visual metronomes, with rhythmic timing that was more precise and accurate and with a larger error correction than with auditory isochronous stimuli (Figure 1). This interspecies difference may depend on the anatomofunctional properties of their audiomotor system, especially in the parietal cortex, which is an intermediate processing node.20,62,63 The posterior parietal cortex processes multimodal information73,96 and is deeply involved in sensorimotor control.96 In humans, the posterior auditory cortex is amply connected with the parietal cortex, forming the dorsal auditory stream for the localization and timing of sound.97–100 Conversely, in monkeys, the homologous posterior medial and lateral belt areas of the auditory cortex only send restricted connections to area 7a and VIP of the parietal lobe. This limited auditory input contrasts with the massive reciprocal link between parietal and visual areas in macaques, constituting the dorsal visual stream for spatial, temporal, and motion processing.102,103 For example, area 7a is strongly connected with visual areas that map the foveal and especially the peripheral visual field (V2, V3, PO), as well as with areas involved in visual motion in the middle temporal cortex (MT, MST).104–106 Hence, we suggest that the bias toward visual metronomes in monkeys is rooted in their large visual parietal region, while the remarkable
human abilities for auditory beat perception and entrainment depend on their vast audioparietal link. A fundamental result of the present study is that the modality of the metronome produced a large displacement of the neural trajectories in state space without considerably altering their internal pulse representation or the rhythmic time-keeping mechanism. These findings suggest the existence of a modality-dependent tonic external input that produces a divergence in the cyclic neural trajectories to different subspaces during ST. Jazayeri’s group recently reported a similar mechanism to encode two timing gain contexts for interval reproduction, based on different levels of a static input while preserving the computational timing mechanism in the MPC. Posterior parietal areas, especially area 7a, could provide this differential input. Indeed, the preSMA receives parietal inputs mainly from area 7a. This ample region of the posterior parietal cortex receives differential auditory and visual inputs (see above) and is implicated in sensorimotor coordination during reaching, as well as in timing of single and rhythmic intervals. Consequently, we postulate that area 7a sends partially overlapping auditory and visual inputs to the MPC and that the uneven connectivity between the two modalities imposes the tonic divergence of the cyclical neural trajectories in the different auditory and visual subspaces. We emphasize the notion of partially overlapping audiovisual inputs to the MPC due to both the observed non-orthogonal subspaces for visual and auditory metronomes and the single-cell mixed selectivity across modalities (see simulations in Figure 7G).

Last, error correction is a critical component of tapping synchronization and is constituted by a phase correction mechanism involved in subtle corrections of the relative phase between the metronome and the taps and by a period correction mechanism that adjusts deviations of the internal neural clock period. Consistent with the human auditory bias and monkey visual bias for beat-based timing, the period correction in humans is larger for auditory than visual metronomes, whereas, in monkeys, we observed the opposite (Figures 1E and 1F). EEG experiments in humans have shown that the MPC is involved in error correction during natural tap synchronization and during tapping perturbations. Our individual session results showed that the lag 1 autocorrelation of the amplitude and speed of MPC neural trajectories was negative during the dwell, especially for the condition of 850 ms with a visual metronome. Crucially, the correlations between the lag 1 autocorrelations of the behavior and the speed/amplitude in the trajectories occurred earlier in the dwell, suggesting that the error correction signal occurred before the encoding of the dwell duration. Therefore, we suggest that the error-correction signal containing the amount of over-underestimation of the previous dwell is kept in the circuit to be compared with the elapsed time representation to define the actual produced interval.

Limitations of the study
The small number of simultaneously recorded neurons in monkey 1 and the lack of video recordings during task performance precluded the individual session analysis in this animal. Another caveat of the present study was that we only used two durations, which limited the use of more complete models to determine encoding of dwell duration. Finally, the use of high-density silicon electrodes allows for the stable recording of neurons using a semi-chronic approach. However, it is important to record in different anterior-posterior locations on the MPC to be able to determine the differential role of SMA proper and preSMA in the described representation of ST parameters.

STAR METHODS
Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
- **METHOD DETAILS**
  - Synchronization task
  - Recording
  - Spike detection and discrimination
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Neural activation periods
  - Neural trajectories
  - Generating neural trajectories
  - Coding subspaces and mixed variance
  - Kinematic analysis of neural trajectories
  - Movement kinematics
  - Temporal scaling in the synchronization task
  - Amplitude-modulation-time-scaling index (AMSI)
  - Distance and diagonal asymmetry index
  - Moving bumps simulations

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at [https://doi.org/10.1016/j.celrep.2023.113234](https://doi.org/10.1016/j.celrep.2023.113234).

ACKNOWLEDGMENTS
We thank Ranulfo Romo, Pavel Rueda, Jonathan Cannon, and Roman Rossi for their fruitful comments on the manuscript. We also thank Jessica Gonzalez Norris for proofreading the manuscript and Raul Paulin and Luis Prado for their technical assistance. A.B. is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and has received Consejo Nacional de Ciencia y Tecnología (CONACYT) fellowship 403678. H.M. is supported by CONACYT grant A1-S-8430, UNAM-DGAPA-PAPIIT IN201721.

AUTHOR CONTRIBUTIONS
H.M., A.B., and J.G. conceived the study. A.B. and J.G. collected the data. A.B., O.P., and H.M. performed data analyses. O.P. and G.M. contributed analytical tools and guidance. H.M. supervised the project. H.M. wrote the manuscript. A.B., J.G., G.M., O.P., and H.M. edited the manuscript.
REFERENCES


**STAR METHODS**

**KEY RESOURCES TABLE**

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental models: Organisms/strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhesus macaque (Macaca mulatta)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Deposited data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preprocessed study data</td>
<td>this paper</td>
<td><a href="https://zenodo.org/record/8352554">https://zenodo.org/record/8352554</a></td>
</tr>
<tr>
<td>Software and algorithms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis code</td>
<td>this paper</td>
<td><a href="https://zenodo.org/record/8361030">https://zenodo.org/record/8361030</a></td>
</tr>
<tr>
<td>Python</td>
<td>Python Software Foundation</td>
<td><a href="https://www.python.org/">https://www.python.org/</a></td>
</tr>
<tr>
<td>KiloSort</td>
<td>Pachitariu et al.\textsuperscript{116}; Steinmetz et al.\textsuperscript{117}</td>
<td><a href="https://github.com/cortex-lab/KiloSort">https://github.com/cortex-lab/KiloSort</a></td>
</tr>
<tr>
<td>ABVA</td>
<td>this paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Synapse</td>
<td>Tucker Davis Technologies</td>
<td><a href="https://www.tdt.com/component/synapse-software/">https://www.tdt.com/component/synapse-software/</a></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RZ2 Bioamp Processor</td>
<td>Tucker Davis Technologies</td>
<td><a href="https://www.tdt.com/component/rz2-bioamp-processor/">https://www.tdt.com/component/rz2-bioamp-processor/</a></td>
</tr>
<tr>
<td>Buzsaki64-Z64 probe</td>
<td>Neuronexus</td>
<td><a href="https://www.neuronexus.com/">https://www.neuronexus.com/</a></td>
</tr>
</tbody>
</table>

**RESOURCE AVAILABILITY**

**Lead contact**
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Hugo Merchant (hugomerchant@unam.mx).

**Materials availability**
This study did not generate any new unique reagents.

**Data and code availability**
- Original data is available at Zenodo and is publicly available as of the date of publication. The DOI is listed in the key resources table.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. The DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

**EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**

All the animal care, housing, and experimental procedures were approved by Ethics in Research Committee of the Universidad Nacional Autónoma de México and conformed to the principles outlined in the Guide for the Care and Use of Laboratory Animals (NIH, publication number 85-23, revised 1985). The two adult monkeys (Macaca mulatta), one female (M01) and one male (M02), between 5 and 7 kg, were trained to perform the Synchronization behavioral task (ST). Monkeys were monitored daily by researchers and the animal care staff to check their health and wellbeing.

Visual stimuli were presented on a 23-inch Dell H236HL monitor at a resolution of 1920x1080 at a refresh rate of 60 Hz, and auditory stimuli were played from the form speakers in the recording setup at 60 cm from the monkeys' head.

**METHOD DETAILS**

**Synchronization task**
The Synchronization task has been described before.\textsuperscript{15} Briefly, monkeys were trained to attend to a sequence of brief stimuli with a constant interstimulus interval and push a button in synchrony with the last six stimuli (Figure 1A). At the beginning of a trial, the
animals held a lever and attended to three stimuli, after which they started to move, the goal being to produce six taps in synchrony with the six remaining metronome pulses. Trials were separated by a variable intertrial interval (1.2–4 s). The inter-onset intervals (450 or 850 ms) for the visual (red square with a side length of 5 cm, presented for 33 ms) or auditory (white noise with a 33 ms duration, 75dB) metronomes were presented in blocks of 25 trials. Monkeys received a reward (fruit juice) when the duration of the produced intervals showed an error below 18% of the instructed interval and all asynchronies between stimuli and taps were less than ±75dB. Metronomes were presented in blocks of 25 trials. Monkeys received a reward (fruit juice) when the duration of the produced interval/modality conditions was random across days.

Recording
We used two semichronic, high-density electrode systems placed bilaterally in the limit between SMA and preSMA, with 64 recording sites (Buzsaki64-Z64, Neuronexus) in the left and 64 in the right hemisphere. We refer to the recorded regions as medial premotor cortex (MPC). The probes were connected to a microdrive that allowed the control of the movement of the two electrode systems independently in the dorsoventral axis. The neural data of 128 channels was acquired, amplified, and digitized using a PZ2 preamplifier (Tucker-Davis Technologies, FL, USA, http://www.tdt.com) at 24,414Hz. The signal was transmitted to an RZ2 base station through fiber optics for on-line processing. Data was stored using RS4 Data Streamer. Preliminary spike sorting was performed offline using the ABVA and Kilosort algorithms (see below).

Spike detection and discrimination
We developed a spike sorting pipeline (ABVA), which is based on previous algorithms for massive spike discrimination. Briefly, it considers that high-density silicon probes can record the action potential of the same neuron in different recording sites (up to the 8 recording sites for each silicon probe), producing simultaneous events with different amplitudes in many recording channels. Hence, ABVA used the data from each shank to perform the spike detection and discrimination steps.

In the first stage of the algorithm, spikes are detected using a double threshold methodology with a maximum time window tw and a minimum threshold voltage Vth. Specifically, spikes are detected as spatiotemporally connected events coming from one cell when the snippets for different recording sites exceed the Vth and showed peak times below tw. The next step is the feature extraction, which obtains the d-spike properties that allows larger spike discrimination. We used PCA to extract the most representative waveforms within the eight recording sites of a shank. So, for a spike consisting of n sample points (62 samples, corresponding to 2.5 ms) for each recording site, the feature extraction method produces m variables (m < n), where m is the number of features in PCA space (the number of features is a user-settable parameter, with a default value of 3).

The next step is spike clustering. In this stage, spikes with similar feature vectors across the eight recording sites are grouped into clusters in the low-dimensional space, assuming that each cluster represents the spikes of a single cell. We use K-means clustering to classify the spatiotemporally connected events. The optimal number of clusters was evaluated by the MATLAB Statistics Toolbox function “evalclusters” based on the Calinski-Harabasz clustering criterion. Crucially, we built a mask that corresponds to the voltage shape templates of temporally overlapping events in the group of channels that were clustered as an individual cell (Figures S12A and S12B). This spike-shaped template mask was used to discriminate during task performance the response events of a cell across the eight recording sites of a shank.

Finally, the results are curated to adjust the potential errors made by the clustering algorithm. In our case, the algorithm automatically does the curation, polishing the clustering results using a dual-threshold on the spike mask. This threshold is applied to each recording site template of the mask and includes a minimum number of occurrences as well as a minimum amplitude of the peak-to-valley waveform. This strategy strongly avoids spurious detection of small noise events on the masks.

As a result, in one session, up to 142 isolated single neurons were detected (average of 102, cells per recording, range: 42–142), in eight analyzed recording sessions for Monkey 1 and fourteen sessions for Monkey 2. The total number of analyzed neurons after discrimination was 1189: 225 for Monkey 1 and 964 for Monkey 2. These neurons were recorded for twenty-five trials for each of the four task conditions and were included in the analysis, unless noted otherwise, irrespective of their activity modulation in the task.

Our goal was to produce a practical system that can be used in our laboratory for processing information using high-count silicon probes offline and in real-time. We validated our algorithm by comparing ABVA with the commonly used KiloSort. KiloSort 2.5 was used for spike discrimination, using all default parameters with the exception of ops.Th = [10,4]. We compared the results of both algorithms for different recording sessions using the Pearson correlation over the activation profiles of the resulting cell responses during the task performance of our four conditions. We identified the cells with high correlations (r2 > 0.3, p < 0.001) in their activity pattern between the two algorithms. Figure S12C shows activation profiles of the 72 cells in session 1 of Monkey 2 with high correlations between algorithms: KiloSort identified 138 total cells and ABVA 99. Similar robust correlations in activity patterns were obtained between the two methods across all sessions.

Neural activation periods
We used Poisson-train analysis to identify the cell activation periods within each interval defined by two subsequent taps. This analysis determines how improbable it is that the number of action potentials within a specific condition (i.e., target interval and ordinal sequence) was a chance occurrence. For this purpose, the actual number of spikes within a time window was compared with the
number of spikes predicted by the Poisson distribution derived from the mean discharge rate during the entire recording of the cell. The measure of improbability was the surprise index (SI) defined as:

\[ SI = - \log P \]

where \( P \) was defined by the Poisson equation:

\[ P = e^{-r} \sum_{i=0}^{n} \frac{(rT)^i}{i!} \]

where \( P \) is the probability that, given the average discharge rate \( r \), the spike train for a produced interval \( T \) contains \( n \) or more spikes in a trial. Thus, a large \( SI \) indicates a low probability that a specific elevation in activity was a chance occurrence. This analysis assumes that an activation period is statistically different from the average discharge rate \( r \), considering that the firing of the cell follows a non-homogenous Poisson process. The detection of activation periods above randomness has been described previously. Importantly, the Poisson-train analysis provided the response-onset latency and the activation period for each cell and for each combination of target interval/serial order.

Neural trajectories

**Trial binarization**

For each trial, we computed the produced interval (time between two taps) for the four intervals in the ST sequence. Then, we calculated the \( Bin_{time} \), as the time that is equal to the produced interval divided by \( Bin_{size} \), where \( Bin_{size} \) was 22 and 42 for the target intervals of 450 and 850 ms, respectively. The discharge rate of each cell was computed in time windows as the number of spikes that occurred in the \( Bin_{time} \), divided by \( Bin_{time} \). This time series was smoothed using Gaussian filter with a kernel of 20ms and was normalized for each neuron. The resulting discharge rate was called target interval normalized data (TIND).

**Principal component coefficients matrix**

Given a linear transformation of a matrix \( X \) into a matrix \( Y \), such that each dimension of \( Y \) explains the variance of the original data \( X \) in descending order, PCA can be described as the search for matrix \( P \) that transforms \( X \) into \( Y \) as follows.

\[ Y = PX \]

Hence, we first calculated the matrix \( P \) using a matrix \( X \) that includes all trials of the four target duration x modality ST conditions. Using this \( P \) on across all the task conditions guarantees that the same transformation is applied to different neural activity sets. Therefore, using the TIND framework, we avoided over- or under-representation of the information for different target intervals.

**Generating neural trajectories**

Principal component coefficient matrix \( P \) was multiplied by the \( X \) matrix to transform the neural data into the space of the original \( Y \). Using the same transformation matrix for each trial allowed the comparison of trajectories for different trials and tasks. A locally weighted scatterplot smoothing function was applied to the columns of the \( Y \) matrix. We used PCA to visualize and analyze recorded activity patterns (Y). For both animals, neural trajectories (Figure S4) and the properties of the neural sequences (Figures S8) were qualitatively similar. We thus performed the main analyses on the combined population (visual and auditory condition) of neurons across animals. The first 10 PCs explained the 25% of the variance in the data. In contrast, when we performed PCA over short audiovisual population depicted in Figure S9A the first 10 PCs captured 55% of the variance, while the first 10 PCs of the long audio-motor population (Figure S9B) captured 52% of the variance. For the AS, AL, VS., and VL populations shown in Figure 4F, the first 10 PCs captured 44%, 40%, 48% and 43% of the variance in the data, respectively. Therefore, the low total variance explained using the data from all conditions is due to the heterogeneity of neural responses between durations and modalities.

**Coding subspaces and mixed variance**

The subspaces for duration and modality were computed by projecting the trajectories of the variable of interest (e.g., auditory modality) from the original 3D neural trajectories into the second-order PC for these data. The line separatix was computed by performing Orthogonal Distance Regression (ODR) on the original 3D neural trajectories. For the serial order subspace, we used the 3rd PC the second-order PCs of duration to obtain the change in the trajectories orthogonal to the duration plane that is related to the four consecutively produced intervals. The data shown in Figure 2F corresponds to the overall mean from the original 3D neural trajectories for duration, modality, and elapsed time. Thus, the trajectories in white for elapsed time correspond to the bin-by-bin mean of the trajectories of the four conditions.

To determine the partial and mixed variance, we first computed the total variance of the neural trajectories from the matrix \( R^{(3 \times 16800)} \) corresponding to 3 PCs x 168 bins x 25 trials x 4 conditions (PC’s, elapsed time, trials, modality, duration). Then, for the duration parameter, we defined a matrix \( R^{(3 \times 50)} \) which corresponds to the overall mean from the neural trajectories for duration and elapsed time and calculated the variance divided by the total variance. For the modality parameter, we calculated the variance of matrix \( R^{(3 \times 50)} \), which corresponds to the overall mean from the neural trajectories for modality and elapsed time divided by the total variance. For the time encoding parameter, we computed the variance of a matrix \( R^{(3 \times 4096)} \) corresponding to the overall mean from...
the neural trajectories for the duration and modality divided by the total variance. Finally, the mixed variance corresponds to the difference between the total variance and the sum of the duration, modality and elapsed time variance divided by the total variance.

**Kinematic analysis of neural trajectories**

The first three PCs explained the 7.1, 4.1, and 4 percent of the total variance. PC1 showed a steep change between the oscillatory dynamics in the auditory and visual. In contrast, the PC2 and PC3 showed a strong oscillatory structure with a phase difference of $\pi/2$ radians during ST. For these three PCs, we calculated the amplitude of the trajectory as the Euclidean distance between the anchor point (Figure 2C) and each point in the trajectory segment across the four serial order elements for each target interval. The angle of the trajectory was calculated as the angle from the dot product between the anchor point (Figure 2C) and each point in the trajectory segment across the four serial-order elements for each target interval. The position was calculated as the signed difference of the point (Figure 2C) and each point in the trajectory segment across the four serial order elements for each target interval. The angle of

**Gaussian Process Factor Analysis**

Gaussian Process Factor Analysis (GPFA) extracts low-dimension latent trajectories from noisy, high-dimension time series data. It combines linear dimensionality reduction (factor analysis) with Gaussian-process temporal smoothing in a unified probabilistic framework.

The input consists of a set of trials ($Y$) that includes all trials and target interval combinations for the visual and auditory condition for ST in the TIND cell population, each containing a list of spike trains ($N$ neurons). The output is the projection ($X$) of the data in a space of pre-chosen dimensionality $X_{\text{dim}} < N$.

Under the assumption of a linear relation (transform matrix $C$) between the latent variable $X$ following a Gaussian process and the spike train data $Y$ with a bias $d$ and a noise term of zero mean and covariance $R$ (i.e., $Y = CX + d + N(0,R)$), the projection corresponds to the conditional probability $E[X|Y]$. The parameters $(C, d, R)$, as well as the time scales and variances of the Gaussian process, are estimated from the data using an expectation maximization (EM) algorithm.

**Movement kinematics**

We applied the Lucas-Kanade optic flow method to measure the monkey’s arm speed during the ST. This method calculates a flow field from the intensity changes between two consecutive video frames. The analyzed video was recorded with a high-speed camera (Basler acA750 AG) positioned orthogonally to the hand’s plane of motion with a 640x480 resolution at 250 frames per second. The optic flow method was applied to a smaller area of 140x140 pixels from the original video that contained the monkey’s arm during the whole trial and no other moving objects. The arm’s movement velocity vector was calculated across all frames as the magnitude of the sum of all the individual flow field vectors whose magnitude was larger than a predefined threshold. The velocity vector was calculated from the first to the last tap on each correct trial. We reported the speed as the magnitude of the velocity vector. Subsequently, the kinematic state of the arm was tagged as movement when the velocity vector was larger than a threshold or dwell. The tagging algorithm considered a change in the kinematic state when the new state lasted longer than three consecutive frames.

**Temporal scaling in the synchronization task**

To quantify temporal scaling, we defined the scaling index (SI) of a subspace $S$ as the portion of variance of the projections of trajectories into $S$ that can be explained by temporal scaling (Wang et al. 2018). We computed the $k$th scaling component $u_{SC,k}$ as:

$$U_{SC,k} = \arg \min_{u} \sum r(t,T) \left( u^T (r(t,T) u) - \text{Mean}_{r(t,T)} (r(t,T) u) \right)^2 \sum r(t,T) \left( u^T (r(t,T) u) - \text{Mean}_{r(t,T)} (r(t,T) u) \right)^2$$

Equation 1

where $r(t,T)$ is population activity at the scaled time when the duration of the production epoch is $T$, the denominator is the total variance of the trajectories, and the numerator is the variance that cannot be explained by temporal scaling. As in (71) Biand Zhou (2021), to compute the first scaling component $U_{SC,1}$ we set $r(t,T) = r^{PC}(t,T_p)$ with $0 \leq t \leq 1$, where $r^{PC}$ is the projection of the population activity in the subspace spanned by the first three PCs, and $T_p$ is the interval produced by the monkey in the production epoch; then we minimized $u$ in Equation 1. To calculate the second scaling component $U_{SC,2}$, we set $r(t,T) = r(t,T) - r(t,T)U_{SC,1}$, and then minimized $u$ in Equation 1 in the subspace orthogonal to $U_{SC,1}$. In this way, we computed the three scaling components one at a time. Finally, the scaling index (SI) of a subspace was defined as:

$$SI = \frac{\sum r(t,T) U - \text{Mean}_{r(t,T)} (r(t,T) U)^2}{\sum r(t,T) U - \text{Mean}_{r(t,T)} (r(t,T) U)^2}$$

where $r(t,T)U$ is the projection of the scaled trajectory to the subspace $U$. For implementation we used https://github.com/zedongbi/IntervalTiming.71
Amplitude-modulation-time-scaling index (AMSI)

The AMSI was computed as follows:

\[ \text{AMSI} = \frac{k_a - k_v}{k_a + k_v} \]

where \( k_a \) corresponds to the amplitude index, defined as:

\[ k_a = \frac{a_2 - a_1}{\max(a_1, a_2) \left( 1 - \frac{t_1}{t_2} \right)} \]

And \( k_v \) corresponds to the velocity index, defined as:

\[ k_v = \frac{v_1 - v_2}{\max(v_1, v_2) \left( 1 - \frac{t_1}{t_2} \right)} \]

and \( a, v \) and \( t \) correspond to the amplitude, velocity, and time (target duration) of the neural trajectories for short (1) and long (2) durations (see Figure S4). A logarithmic sigmoid function \( \sigma \) was used to adjust \( k_a = \sigma(k_a) \) and \( k_v = \sigma(k_v) \) to get values within a 0 to 1 range. Thus, AMSI reached a value of \(-1\) when the neural trajectory was fully temporal-scaled, whereas it reached a value of \(1\) when the neural trajectories were fully amplitude-modulated.

Distance and diagonal asymmetry index

To quantitatively determine the change in the dynamics of neural sequences across conditions we adopted a geometric approach that has been used before. First, we constructed the heatmaps of Figure 6A with the cell response profiles segmented into 20 bins and averaged across the four serial order elements of the ST, for each of the four task conditions. Second, we computed the Euclidean distance matrix between all possible pairs of task conditions using the sorted activity in Figure 6A. Third, we identified the minimum distance value for each bin within each distance matrix and extracted two vectors: one vector keeps the distance values at the minimum (called the minimum distance vector), and the other keeps the time bin difference between the diagonal and the minimum values (called the diagonal asymmetry vector). We transformed the diagonal asymmetry vector into angular differences, where a difference of 20 bins between the diagonal and the minimum value corresponds to 360°. For the self-sorted neural sequences, the minimum distance and diagonal asymmetry vectors were constituted by twenty zeros, one for each time bin, since we were comparing a condition with itself (Figure 6B, white dots in task comparison diagonal). In contrast, for cross-sorted moving bumps, the minimum values of the distance matrix can fall into different values between the rows and columns and both vectors were different from zero (Figure 6B, white dots in task comparison off-diagonal). Fourth, we defined the distance index as the mean of the minimum distance vector, and the diagonal asymmetry index as the sum of the diagonal asymmetry vector. The former is a measure of the overall distance between the neural sequences of a pair of task conditions, whereas the latter is a cumulative measure of how congruent over time the two neural sequences within each produced interval of the ST are.

Moving bumps simulations

To investigate how the properties of neural sequences were associated with the geometry and kinematics of population neuronal trajectories, we performed different simulations. Each simulation comprises a set of \( N_d \) neurons with a firing rate:

\[ r_i(t, t_d) = m_d \eta \left( \frac{\mu_i(t_d)}{\sigma_i} \right)^2 \]

where \( t \) is the elapsed time, and \( m_d \) and \( \sigma_d \) are the maximum discharge rate and duration of the activation period, respectively, which depend on the target duration \( t_d \). \( \mu_i \) and \( \sigma_i \) correspond to the time of peak activation of neuron \( i \). The peaks of activity at \( \mu_i \) cover the entire interval \( t_d \). To create neural sequences that reset at the next produced interval in the sequence, we used the distance function \( d(t, \mu_i) = \min(|t - \mu_i|, t_d - |t - \mu_i|) \), allowing for the generation of cyclic kinematics.

Regarding the changes in duration of cell activation of Figures 7C, 7E and 7F, we used \( \sigma_{850} > \sigma_{450} \). For the remaining simulations we used \( \sigma_{850} = \sigma_{450} \). For the changes in response magnitude of Figure 7D we used \( m_{850} > m_{450} \) and \( m_{850} = m_{450} \) for all the other cases. The dynamics of the activation peaks depicted in Figures 7A and 7B followed an absolute time representation \( \mu_i = \mu_{450} \). The same representation is used for the shared cells at the beginning and end on the interval of Figures 7E–7G. In contrast, for the rest of the simulations the dynamics of the activation peaks followed a relative time scaling with \( \mu_{450} = \frac{850}{850-450} \). We used \( N_{450} < N_{850} \) for Figures 7B and 7F. Finally, for Figures 7F and 7G we randomly inserted neurons within the intermediate portion of the produced interval.
Supplemental information

Amodal population clock in the primate medial premotor system for rhythmic tapping

Abraham Betancourt, Oswaldo Pérez, Jorge Gámez, Germán Mendoza, and Hugo Merchant
**Supporting information**

Supplementary Table 1. Individual session analysis for the Dwell epoch, where we computed the Pearson correlation coefficient between the behavioral (top labels) and kinematic parameters of the neural trajectories (bottom labels) in the fourteen sessions of Monkey 2 with videos on ST behavior. Orange shading is associated with the sessions that had the following three significant correlations: dwell time vs dwell amplitude, temporal variability vs variability of position, and Lag-1 autocorrelation of produced interval vs Lag-1 autocorrelation of dwell amplitude. Sessions in blue shading showed at least one or two significant correlations.

<table>
<thead>
<tr>
<th>#Session</th>
<th>DWELL TIME</th>
<th>TEMPORAL VARIABILITY</th>
<th>AUTOCORRELATION LAG-1 PI</th>
<th>#Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dwell amplitude</td>
<td>Dwell speed</td>
<td>Variability of position</td>
<td>Autocorrelation Lag-1 Dwell amplitude</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>p</td>
<td>r</td>
<td>p</td>
</tr>
<tr>
<td>2</td>
<td>0.59</td>
<td>0.0001</td>
<td>-0.76</td>
<td>0.0001</td>
</tr>
<tr>
<td>3</td>
<td>0.9</td>
<td>0.0001</td>
<td>-0.54</td>
<td>0.0001</td>
</tr>
<tr>
<td>6</td>
<td>0.55</td>
<td>0.0001</td>
<td>-0.62</td>
<td>0.0001</td>
</tr>
<tr>
<td>8</td>
<td>0.7</td>
<td>0.0001</td>
<td>-0.41</td>
<td>0.0001</td>
</tr>
<tr>
<td>9</td>
<td>0.86</td>
<td>0.0001</td>
<td>-0.24</td>
<td>0.03</td>
</tr>
<tr>
<td>11</td>
<td>0.85</td>
<td>0.0001</td>
<td>-0.17</td>
<td>0.19</td>
</tr>
<tr>
<td>14</td>
<td>0.92</td>
<td>0.0001</td>
<td>-0.35</td>
<td>0.001</td>
</tr>
<tr>
<td>1</td>
<td>0.95</td>
<td>0.0001</td>
<td>-0.95</td>
<td>0.0001</td>
</tr>
<tr>
<td>4</td>
<td>0.95</td>
<td>0.0001</td>
<td>-0.81</td>
<td>0.0001</td>
</tr>
<tr>
<td>5</td>
<td>0.93</td>
<td>0.0001</td>
<td>-0.79</td>
<td>0.0001</td>
</tr>
<tr>
<td>7</td>
<td>0.94</td>
<td>0.0001</td>
<td>-0.7</td>
<td>0.0001</td>
</tr>
<tr>
<td>10</td>
<td>0.98</td>
<td>0.0001</td>
<td>-0.38</td>
<td>0.01</td>
</tr>
<tr>
<td>12</td>
<td>0.99</td>
<td>0.0001</td>
<td>-0.28</td>
<td>0.05</td>
</tr>
<tr>
<td>13</td>
<td>0.98</td>
<td>0.0001</td>
<td>-0.61</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

**Significant sessions**

| 14 | 13 | 13 | 7 | 2 |

Orange shading is associated with three significant correlations: dwell time vs dwell amplitude, temporal variability vs variability of position, and Lag-1 autocorrelation of produced interval vs Lag-1 autocorrelation of dwell amplitude. Sessions in blue shading showed at least one or two significant correlations.
<table>
<thead>
<tr>
<th>#Session</th>
<th>MOVEMENT TIME</th>
<th>AUTOCORRELATION LAG-1 PI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Movement amplitude</td>
<td>Movement speed</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>p</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
<td>0.21</td>
</tr>
<tr>
<td>3</td>
<td>-0.03</td>
<td>0.81</td>
</tr>
<tr>
<td>6</td>
<td>-0.45</td>
<td>0.0001</td>
</tr>
<tr>
<td>8</td>
<td>-0.04</td>
<td>0.79</td>
</tr>
<tr>
<td>9</td>
<td>-0.61</td>
<td>0.0001</td>
</tr>
<tr>
<td>11</td>
<td>-0.4</td>
<td>0.0001</td>
</tr>
<tr>
<td>14</td>
<td>-0.16</td>
<td>0.28</td>
</tr>
<tr>
<td>1</td>
<td>0.23</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>-0.27</td>
<td>0.04</td>
</tr>
<tr>
<td>5</td>
<td>-0.4</td>
<td>0.0001</td>
</tr>
<tr>
<td>7</td>
<td>-0.17</td>
<td>0.22</td>
</tr>
<tr>
<td>10</td>
<td>0.06</td>
<td>0.67</td>
</tr>
<tr>
<td>12</td>
<td>-0.06</td>
<td>0.0001</td>
</tr>
<tr>
<td>13</td>
<td>0.2</td>
<td>0.15</td>
</tr>
<tr>
<td>#Cells</td>
<td>42</td>
<td>28</td>
</tr>
</tbody>
</table>

**Supplementary Table 2.** Individual session analysis where we computed the Pearson correlation coefficient between the behavioral (top labels) and kinematic parameters of the neural trajectories (bottom labels) in the fourteen sessions of Monkey 2 with collected videos on behavior. Orange and blue shading follows the significant effects of Supplementary Table 1.
Figure S1. Neural population traces.

(A) Neural population geometry of the first three principal components of the auditory and visual modality used to calculate the scaling index and scaling components. Color code format as in Figure 2B.

(B) Scaling components for the auditory (Top panels) and visual (Bottom panels) modality calculated in the subspace spanned by the first three principal components.

(C) Portion of the 3 PCs during dwell time for the auditory (top) and visual (bottom) modalities.

(D) Same as in C but for the movement time epoch of the ST.
Figure S2. Neural population trajectories during ST and their oscillatory dynamic properties using alternative reduction method (GPFA). Format as in Figure 2.
Figure S3. Correlation between different behavioral and neural trajectories parameters of Session 2 of Monkey2.
(A) Neural trajectories in the first three PCs as in Figure 2A. The tested arbitrary points for computing the amplitude and angle are depicted in state space with two colors. Dark orange for the positions where the properties of behavior were replicated in the neural population kinematics, as in Figure 3A-C and pink for the positions where the kinematics did not follow these properties. Note that similar kinematic properties were obtained when the arbitrary point was located within a contiguous region (shaded) in state space.

(B) The same as in A but only for the projected neural activity of Session 2 of Monkey 2.

(C) Dwell and Movement time amplitude (± 2xSEM), computed as area under curve from neural trajectories of Session 2 of Monkey 2, as a function of target interval. The ANOVA showed significant main effects of modality, $F(1,196) = 1580.9$, $p < .0001$; epoch, $F(1,196) = 3716.2$, $p < .0001$; but not statistical significance on duration, $F(1,196) = .45$, $p = .5$; as well as significant effects on duration x epoch interaction, $F(1,192) = 4194.82$, $p < .0001$ and modality x epoch interaction, $F(1,196) = 38.7$, $p < .0001$.

(D) Lag 1 autocorrelation of the amplitude of the neural trajectories during the Dwell time as a function of target duration for Session 2 of Monkey 2. The ANOVA showed significant main effects of duration, $F(1,196) = 4.17$, $p < .04$, but no statistical significant effect on modality, $F(1,196) = 3.21$, $p = .07$ and on duration x modality interaction, $F(1,196) = .14$, $p = .70$.

(E) Variability of the position (SD within and across trials) as a function of target interval (± 5xSEM) for the recording session in D. The ANOVA showed significant main effects of duration, $F(1,196) = 26.1$, $p < .0001$, modality, $F(1,196) = 31.63$, $p < .0001$; as well as significant effects on duration x modality interaction, $F(1,196) = 20.3$, $p < .0001$.

(F) Negative significant correlation between the produced interval and movement amplitude ($r = -0.79$, $p < 0.0001$) for the recording session in B.

(G) Significant correlation between the produced interval and dwell amplitude ($r = .86$, $p < .0001$) for recording session 2 of Monkey 2.

(H) Not significant correlation between the produced interval and movement time speed of the neural trajectories ($r =-0.15$, $p = 0.14$) for the recording session in B.

(I) Negative significant correlation between the produced interval and dwell time speed of the neural trajectories ($r =-0.73$, $p < 0.0001$) for the recording session in B.

(J) Significant correlation between the temporal variability of the produced intervals and the variability of the trajectory position ($r = .72$, $p < .0001$) for the recording session in G.

(K) Significant correlation between the autocorrelation Lag-1 of the produced interval vs autocorrelation Lag-1 of the movement amplitude ($r = 0.26$, $p < .007$) for the recording session in B.

(L) Significant correlation between the autocorrelation Lag-1 of the produced interval vs autocorrelation Lag-1 of the dwell amplitude ($r = .2$, $p < .04$) for the recording session in G.

(M) Significant correlation between the autocorrelation Lag-1 of the produced interval vs autocorrelation Lag-1 of the movement speed ($r = 0.25$, $p < .009$) for the recording session in B.

(N) Significant correlation between the autocorrelation Lag-1 of the produced interval vs autocorrelation Lag-1 of the dwell speed ($r = 0.26$, $p < .007$) for the recording session in B.
Figure S4. Amplitude-modulation time-scaling index (AMSI).

Top. Geometric description of the changes in amplitude (a) and velocity (v) for trajectories that show a full temporal scaling (Velocity Index = 1; AMSI = -1) or a full amplitude modulation (Amplitude index = 1; AMSI = 1). t1 = 450, t2 = 850 ms.

Bottom. Equations for all velocity and amplitude ratios and indexes.

<table>
<thead>
<tr>
<th></th>
<th>Time scaling</th>
<th>Amplitude modulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplitude Ratio</td>
<td>$a_1 = a_2$</td>
<td>$a_1 = \frac{t_1}{t_2} a_2$</td>
</tr>
<tr>
<td>Velocity Ratio</td>
<td>$v_2 = \frac{t_1}{t_2} v_1$</td>
<td>$v_2 = v_1$</td>
</tr>
<tr>
<td>Amplitude Index</td>
<td>$k_a \rightarrow 0$</td>
<td>$k_a \rightarrow 1$</td>
</tr>
<tr>
<td>Velocity Index</td>
<td>$k_v \rightarrow 1$</td>
<td>$k_v \rightarrow 0$</td>
</tr>
<tr>
<td>Amplitude Modulation</td>
<td>$AMSI = \frac{k_a - k_v}{k_a + k_v}$</td>
<td>$AMSI \rightarrow -1$</td>
</tr>
</tbody>
</table>
Figure S5. Neural population trajectories and AMSI during ST and their oscillatory dynamic properties.

(A) Projection of the neural activity in MPC for monkey 1 (173 neurons, left panel) and monkey 2 (846 neurons, right panel) during ST onto the first three PCs. The first three PCs explained the 7.9, 4.4, and 3.8% of the total variance for monkey 1 and the first three PCs explained the 8.4, 6.9, and 4.4% of the total variance for monkey 2. We show the average trajectories across the five consecutively produced intervals included in the analysis. Each point in the trajectory represents the neural network state at a particular moment, where the trajectory completes an oscillatory cycle on every produced interval across conditions. Cyan: 450 ms auditory (AS); Blue: 850 ms auditory (AL); Orange: 450 ms visual (VS); Red: 850 ms visual (VL). The green spheres indicate the tapping times across the trial sequence.

(B) General codification of the relative passage of time across the four conditions (white trajectories), as well as the overall state space values for target duration (cyan and cinnamon diamonds), modality (red/blue asterisks) and tapping times (green spheres). The data shown in the left panels correspond to monkey 1 and the right panels to monkey 2.
(C) AMSI as a function of trial time for the auditory (blue) and visual (orange) conditions, the dwell and movement periods are depicted at the bottom for the four produced intervals using conventions in Figure 4B.
Figure S6. Neural correlates of error correction.
(A) Significant correlation ($r = 0.416$, $p < 0.001$) between the amplitude of the neural trajectories and the produced interval for Session 14 of Monkey2. The 25 trials of the 4 produced intervals in the sequences are depicted. We subtracted the average to center both parameters.

(B) Significant correlation ($R = 0.38$, $p = 0.01$) between the trial-by-trial lag 1 autocorrelation of the produced intervals and the lag 1 autocorrelation of the amplitude of the neural trajectories for the same trials for the 850 ms auditory condition. The 25 trial autocorrelations were computed from the four serially produced intervals. Note the large negative values of the lag 1 autocorrelation for the behavior and the amplitude of the neural trajectories.

(C) Top. Box plots of the correlation values between the trajectories’ amplitude and the produced interval for all the significant sessions depicted in D across the four ST conditions. Bottom. Box plots of the correlation values between the lag 1 autocorrelation of the amplitude and the lag 1 autocorrelation of the produced intervals for the same sessions and conditions of the top.

(D) Plots of the time bins with significant correlations ($p < 0.05$) between the lag 1 autocorrelation of the amplitude and the lag 1 autocorrelation of the produced intervals (light green), and the time bins with significant correlations between amplitude of the neural trajectory and the duration of the produced interval (dark green). These significant correlations are shown for the 14 sessions of Monkey 2 as a function of interval duration (phase in $2\pi$) for the four experimental conditions individually. Note the large tendency of the correlation in the lag1 autocorrelations of behavior-trajectory’s amplitude to occur before the correlations between amplitude-produced duration.

(E) Significant correlation ($r = 0.49$, $p < 0.001$) between the speed of the neural trajectories and the produced interval for Session 11 of Monkey2. Format as in A. Note the large decrease in speed as a function of produced duration, evident of temporal scaling in the neural trajectories.

(F) Significant correlation ($R = 0.39$, $p = 0.05$) between the trial-by-trial lag 1 autocorrelation of the produced intervals and the lag 1 autocorrelation of the speed of the neural trajectories for the same trials. Format as in B.

(G) Top. Box plots of the correlation values between the speed of the neural trajectories and the produced interval for all the significant sessions depicted in H and the four ST conditions. Bottom. Box plots of the correlation values between the lag 1 autocorrelation of the speed and the lag 1 autocorrelation of the produced intervals for the same sessions and conditions of the top.

(H) Plots of the time bins with significant correlations ($p < 0.05$) between the lag 1 autocorrelation of the speed of the trajectories and the lag 1 autocorrelation of the produced intervals (light green), and the time bins with significant correlations between amplitude
Figure S7. Dynamics of the neuronal sequences for Monkey 1 and Monkey 2.

(A) Venn Diagram for number of cells with significant effects for duration (green), modality (magenta), and/or serial order duration (purple). The data shown in the left panels correspond to monkey 1 and the right panels to monkey 2.

(B) The neural recruitment lapse showed a larger increase in the first quarter that plateaued until the third quarter and acquired lower values on the last quarter (ANOVA main effect duration, $F(1,51) = 15.23$, $p < 0.001$; main effect modality, $F(1,51) = 31.64$, $p < 0.001$; main effect quarter, $F(3,51) = 5.48$, $p < 0.002$; modality x duration interaction, $F(1,51) = 15.85$, $p < 0.001$; modality x quarter interaction, $F(3,51) = 4.24$, $p < 0.009$; duration x quarter interaction, $F(3,51) = 5.38$, $p < 0.001$).

ANOVA nonsignificant main effect duration, $F(1,51) = 0.20$, $p = 0.65$; main effect modality, $F(1,51) = 27.12$, $p < 0.001$; main effect quarter, $F(3,51) = 52.25$, $p < 0.0001$; nonsignificant modality x duration interaction, $F(1,51) = 1.41$, $p = 0.23$; nonsignificant duration x quarter interaction, $F(3,51) = 0.20$, $p = 0.88$; modality x quarter interaction, $F(3,51) = 3.83$, $p < 0.01$.

(C) The number of cells showed an initial decrease followed by rebound that was steeper for the visual condition, and a peak in the fourth quarter followed by a sharp decrease at the end of the produced interval (ANOVA nonsignificant main effect duration, $F(1,51) = 1.11$, $p = 0.29$; main effect modality, $F(1,51) = 14.50$, $p < 0.001$; main effect quarter, $F(3,51) = 66.34$, $p < 0.0001$; nonsignificant modality x duration interaction, $F(1,51) = 1.98$, $p = 0.16$; nonsignificant duration x quarter interaction, $F(3,51) = 2.35$, $p = 0.08$; modality x quarter interaction, $F(3,51) = 5.81$, $p < 0.01$).

ANOVA nonsignificant main effect duration, $F(1,51) = 1.29$, $p = 0.25$; nonsignificant main effect modality, $F(1,51) = 3.36$, $p = 0.07$; main effect quarter, $F(3,51) = 60.59$, $p < 0.0001$; nonsignificant modality x duration interaction, $F(1,51) = 1.21$, $p = 0.27$; nonsignificant duration x quarter interaction, $F(3,51) = 2.39$, $p = 0.07$; modality x quarter interaction, $F(3,51) = 4.73$, $p < 0.005$. 
Figure S8. Neural sequences for audiovisual neurons.

(A) Average normalized firing rate of cells (y-axis) with significant MI on at least one of the ST parameters for both the auditory and visual conditions of the 450 ms interval. The four vertical black lines represent the tapping times. The cells were aligned to the bin of peak activity.

(B) The same as in A but for cells with significant MI for both the auditory and visual conditions of the 850 ms interval.

(C) Average normalized firing rate of cells (y-axis) with a nonsignificant MI on the ST parameters displayed as a function of trial time for the fours task conditions (AS, AL, VS, VL). Format as in Figure 4F.
Figure S9. Population SDF (mean ± SEM) of the neurons with activation periods during each of the four quarters of a produced interval.
The neurons of moving bumps showed instantaneous activity changes that correspond to different types of ramping patterns reported previously.

(A) Neurons of the first quarter showed ramping profile of the swinging ramps.

(B-C) The neurons of the second and third quarters showed an up-and-down profile of activation that correspond to ramps that encode elapsed time in the width or height of the ramp.

(D) In the last quarter, the neurons encode the time-remaining for an action, reaching a peak at a particular time before the next tap.
Figure S10. Activation profiles across the four conditions and the four serial order elements of the ST.

(A) Correlation matrix of self-sorted cells for auditory short condition (significant r-squared correlations between pairs of cells are illustrated as red points). The significant correlations were distributed not only around the diagonal but also for pairs of cells that were active at the beginning and end of the neural sequence.

(B) Probability that pairs of cells showed large correlation values ($r > 0.5$) in their response profile dividing the complete neural sequence into quarters. The color code goes from zero for blue quarter to one for red quarters. Large probability values are mainly observed for the last two quarters of the neural sequences across all pairwise task comparisons.
(C) Activation periods for the 124 neurons that showed similar response profiles across the four conditions and the four serial order elements of the ST. Figure conventions as in Figure 4A.
(D) Distance index for initial intermediate and final bins using the Euclidean distance across task conditions (left, from Figure 6B) and serial order elements of the auditory large condition (right, from Figure 6C).
Figure S11.
Normalized distance between the short and long interval trajectories at the tapping times for the simulations in Figure 7 panels A-G.
Figure S12. Semiautomatic spike discrimination of the cells with ABVA.

(A) Waveforms (mean ± SD) of the 4-clustered cells (Spk1-Spk3) across the 8 neighboring recording sites of 1 silicon probe with eight recording sites.

(B) Spike clusters of the three cells displayed along the first principal components of spike waveforms extracted from the 1st, 3rd, 4th and 6th channel. Same color code as in A. C: calibration: vertical 150 µV; horizontal 2 ms.

(C). Cells showed large numbers of neurons with significant correlations in their response profiles between spike sorting frameworks for one recorded session. We show the discharge rate of neural activity sorted with ABVA (left panel) and KiloSort (right panel). We found 72 cells with a high correlation (r > .3) in their response between sorting methods (p < .0001).