scientific reports

Check for updates

OPEN

Dynamical mechanisms of how an RNN keeps a beat, uncovered with a low-dimensional reduced model

Klavdia Zemlianova[1], Amitabha Bose[2] & John Rinzel[3]✉

Despite music's omnipresence, the specific neural mechanisms responsible for perceiving and anticipating temporal patterns in music are unknown. To study potential mechanisms for keeping time in rhythmic contexts, we train a biologically constrained RNN, with excitatory (E) and inhibitory (I) units, on seven different stimulus tempos (2–8 Hz) on a synchronization and continuation task, a standard experimental paradigm. Our trained RNN generates a network oscillator that uses an input current (context parameter) to control oscillation frequency and replicates key features of neural dynamics observed in neural recordings of monkeys performing the same task. We develop a reduced three-variable rate model of the RNN and analyze its dynamic properties. By treating our understanding of the mathematical structure for oscillations in the reduced model as predictive, we confirm that the dynamical mechanisms are found also in the RNN. Our neurally plausible reduced model reveals an E-I circuit with two distinct inhibitory sub-populations, of which one is tightly synchronized with the excitatory units.

The ability to estimate time is important for many activities like dancing and playing or listening to a musical instrument. Experimental and computational studies of the neural mechanisms underlying the ability to estimate time have primarily focused on the temporal estimation of isolated intervals[1–3]. However, how the brain keeps a musical beat, called rhythmic timing, likely involves different neural mechanisms that are yet to be explained[4–7].

Unlike timing isolated intervals, rhythmic timing, also called beat-based timing, necessitates computationally distinct requirements. In particular, rhythmic timing requires aligning to a beat in both phase and tempo and requires the ability to internally maintain the learned rhythm in the absence of stimulus input (imagine tapping along to a song after it stops playing). Furthermore, the two modes of timing may be at least partially separated in the brain: beat-based timing is thought to rely on the premotor and supplementary motor areas[8–11], basal ganglia[6,12] and the parietal cortex[13,14] unlike interval timing, which is more dependent on the cerebellum[7,15–17]. Alternative proposals, however, posit that the cerebellum may not be necessary for timing[18,19] at all or that the parietal cortex may play a role in both[20].

Tapping along to an isochronous beat is the simplest form of beat-based timing and has been extensively studied in human psychoacoustic studies[21,22]. Recent work has shown that macaques, too, can be trained to synchronize their motor behavior to a beat[23,24] although, unlike humans, they don't spontaneously do so[25]. Prior studies on electrophysiological recordings from macaques have characterized a number of features of neural data hypothesized to be related to the underlying rhythmic timing mechanism[23,24]. One such prominent feature is sequential firing, a type of dynamic coding[26]: neurons fire at varying phases between tap times thereby forming activity sequences that repeat between taps. Further, when neural activity is projected to a lower dimensional space using principal component analysis, the neural firing rates trace out circular trajectories whose circumference increases with decreasing stimulus frequency. Although previously proposed models for neural timing may show synchronization to a rhythmic stimulus and maintenance of the learned rhythm[27], neuromechanistic explanations of the observed dynamical features of existing neural data are lacking.

Motivated by recent successes of recurrent neural networks (RNNs) to reproduce dynamical features of neural data[28–32], we train an RNN on a standard rhythmic timing task to replicate observed neural features from macaque experiments. To mimic previous parametrizations of the RNN in related tasks, we assume the existence of an input drive as well as an adaptive neuronal process (e.g. error correction process[33]) that dynamically adjusts the magnitude of the input drive to the RNN. Based on this neuro-meaningful context parametrization that mimics the driving current in a biophysical model, training leads to a controllable network that can produce oscillations over a wide range of frequencies and matches key dynamical features of experimental data. Using the

[1]Center for Neural Science, New York University, New York, NY 10003, USA. [2]Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102, USA. [3]Center for Neural Science and Courant Institute of Mathematical Sciences, New York University, New York, NY 10003, USA. ✉email: rinzeljm@gmail.com

RNN, we then develop a neurally plausible reduced model (E-I-I). Using this mathematically tractable model, we uncover dynamical mechanisms that explain the RNN's oscillatory capabilities and propose distinct functional roles for two inhibitory subpopulations. Our novel dissection of the RNN's oscillatory dynamics provides insights into how the brain potentially represents a musical beat.

Results

To examine potential mechanisms for rhythmic timing, we trained a recurrent neural network (RNN) with 500 units (80% excitatory, 20% inhibitory) on a synchronization and continuation task (Fig. 1a)—a standard task in rhythmic timing literature. In this task, stimulus pulses are delivered equally spaced in time and a subject is asked to synchronize a motor action, such as a finger tap, with the stimulus sequence (synchronization phase) and then to continue producing that motor output at the same rate and phase after the stimulus is stopped (continuation phase). To model this task, the inputs to the RNN consisted of stimuli onset times (I_{stim}), modeled as short pulsatile inputs delivered at the frequency of the stimulus, as well as a context cue (I_{cc}), modeled as a constant input drive whose amplitude was inversely related to the interval between stimulus pulses (Fig. 1b).

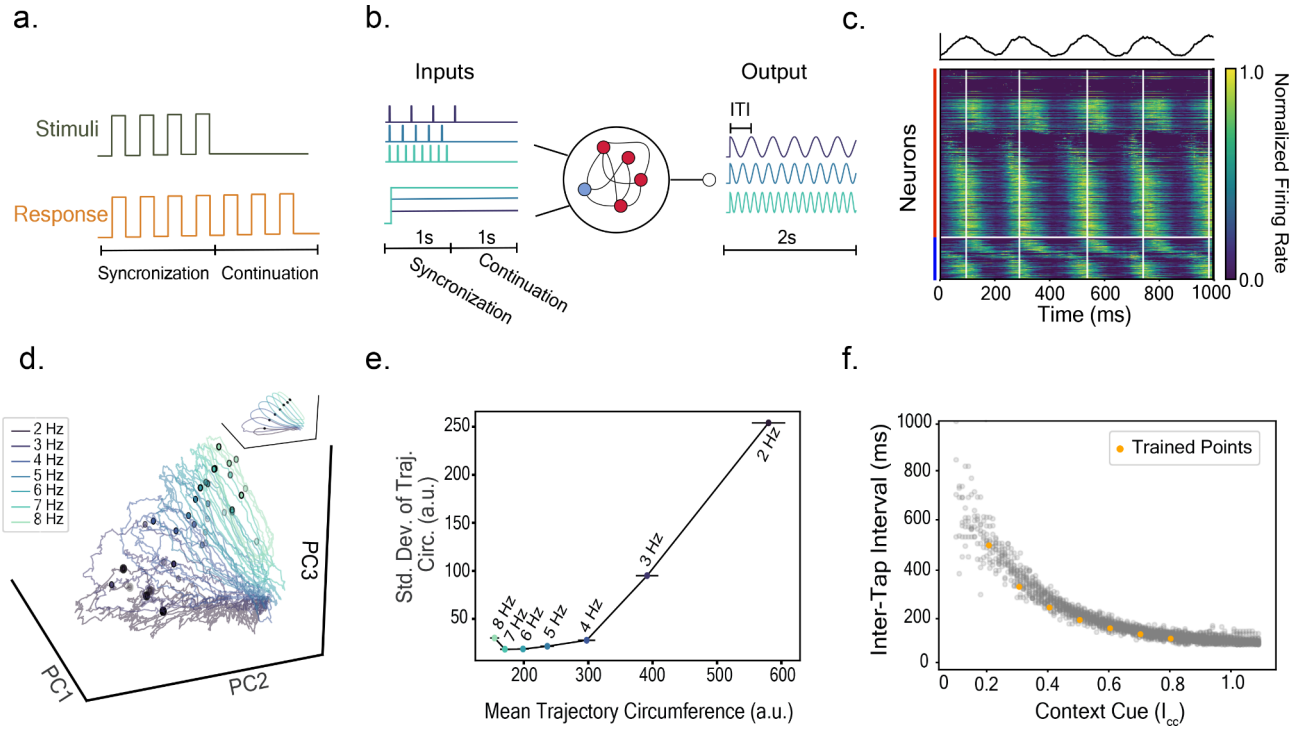

Fig. 1. RNN trained on the synchronization and continuation task replicates key features of neural dynamics. (**a**) Synchronization and continuation task. Stimulus pulses are presented equally spaced in time and the subject responds by first synchronizing their motor output with the stimulus (synchronization phase) and then continuing to produce the motor output at the same rate and phase in the absence of the stimulus (continuation phase). (**b**) Schematic of the RNN training. The RNN, composed of 500 hidden units with 80% excitatory (red) and 20% inhibitory (blue), receives two types of inputs: a context cue and stimulus pulses. The amplitude of the context cue is proportional to the frequency of the stimulus (purple for low frequency, teal for high frequency). The model output is a sinusoid of the same frequency as the stimulus input with the peaks of the sinusoids ("taps") aligned to the stimulus tone times. Black horizontal bar on the output indicates one Inter-Tap Interval (ITI). (**c**) Firing rates of RNN units are normalized so that each unit's activity is between 0 and 1 (color gradient) and then sorted by the time at which each unit hits its maximal firing rate during the first ITI. Excitatory (red) and Inhibitory (blue) units are sorted separately. Time course of the output unit is shown above the heatmap in black. White vertical lines indicate model "tap" times. (**d**) Firing time courses of RNN units trace out circles when projected to the space spanned by the top three principal components (explain 61% of the variance). Color gradient indicates frequency of stimulus (purple = low frequency to teal = high frequency). Black circles indicate model output "tap" times. Inset at top right shows the noise-free time courses projected to the same space for the same trained frequencies with model "tap" times indicated with filled circles. (**e**) Standard deviation vs. mean of the circumference of trajectories from panel d. Trajectory circumference is computed as the arclength of the discretized trajectory in $\mathbb{R}^N$ taken between model output 'tap' times. Mean and standard deviation is computed across individual cycles for each I_{cc} (see "Methods"). (**f**) Simulated ITIs of the RNN output for I_{cc} (grey) at every 0.01 step from 0.1 to 1.1. Yellow circles indicate the points for which the RNN was trained. All dynamics for panels **c**.-**f**. are taken from the RNN simulated in the continuation phase, after transients have settled and the time course is nearly periodic, with noise $\sigma_{in} = \sigma_{rec} = 0.01$ (see "Methods").

Essentially, the pulsatile inputs provided the network with information about phase and the context cue provided information about oscillation period. Finally, the stimulus tones were presented for the first one second of network input only (synchronization phase only) as in the experiments[23,34], while the context cue was provided for the entire two seconds of network input (synchronization and continuation phases). We implicitly assume controllability through an adaptive process (not explicitly modeled) that dynamically adjusts the context cue parameter (e.g. by error correction[33]) to match and maintain the target beat. In short, the network is adaptive and not based on entrainment.

The output of the network, a scalar time course o, was required to produce a sinusoidal response lasting for at least two seconds with peaks of the output cycles aligned with the stimulus tones. We use the peak times of the sinusoidal output as a proxy for the command signal that would be sent to execute a motor action that would be synchronized with the stimulus pulses in this task and we will refer to peak times of the output cycles as model 'tap' times. The network was trained on seven different frequencies: 2, 3, 4, 5, 6, 7 and 8 Hz—a natural range of frequencies for musical and speech rhythms[22,35].

RNN reproduces key features of experimental data

The task-trained RNN replicates several features of the neural dynamics previously observed in neural data of macaques performing the same task[23] such as sequential firing rates and firing rate trajectories that increases in circumference with increasing stimulus period[20,33]. To see if our model replicates the first finding, we sort RNN hidden units according to their peak firing times with units that fire early appearing in the top of the sequence (Fig. 1c). Since the peaks of the sinusoidal output of our model are a proxy for the tap events, we define the inter-tap-interval (ITI) for our model to be the time between consecutive peaks in the output. This analysis of the RNN hidden units reveals weak sequential activity that repeats each ITI (Fig. S1). This sorting also reveals that most excitatory units are synchronized and fire maximally around the "tap" times (white vertical lines in Fig. 1c) with some units having weaker firing peaks that lead or trail the tap-focused units. In contrast, inhibitory units show stronger sequential structure in their firing patterns with some units firing at maximum rates at different ITI phases—28% of the inhibitory units fire within a phase range of 0.2 to 0.8 and are characterized as the Int-I group. In comparison, only 14% of the excitatory units are characterized as the Int-E group. While the sequential activity is weaker than observed experimentally, the RNN replicates the over-representation of units firing at the "tap" just as observed in the experimental data[23].

Examination of the experimental data also showed that when the neural firing rate time courses are projected onto the space spanned by the top three principal components, the trajectories trace out circles such that the radius and variability increase with stimulus period[23,34]. Projecting the firing rates of the RNN units into their corresponding space similarly reveal circular-like trajectories (Fig. 1d) that also increase in radius and variability (Fig. 1e) with stimulus period. The model 'tap' times (indicated by colored dots in Fig. 1d) align themselves along a line in state space—also in agreement with experimental findings.

When visualized in state space, the trained RNN trajectories are organized as a cone-like shape with slow trajectories (Fig. 1d dark colors, large radius) on one end and fast trajectories (Fig. 1d light colors, small radius) on the other end. Although we only trained the RNN on seven frequencies, the trained RNN could interpolate between the learned context cues (driving the RNN with intermediate values of I_{cc} produces oscillations of intermediate periods, see Fig. 1f) as well as extrapolate beyond the trained regime (Fig. 1f—notice ITIs, shorter and longer, for contexts cues above 0.8 and below 0.2, respectively). The finding that the RNN is able to generalize is likely related to a low-dimensional structure in the network whose output varies continuously with a tonic context cue, consistent with previous studies[36–38]. However, unlike previous work on single-interval timing, our model's generalization capabilities extend this finding to oscillatory systems. A context-parameterized structure is intuitively appealing as speeding up (slowing down) tempo can be accomplished by transitioning up (down) along the cone-manifold's axis by increasing (decreasing) the neural drive (I_{cc}).

Developing a three-variable reduced model of the RNN

Having shown that the trained RNN reproduces several qualitative features of the neural data, we sought to understand the underlying dynamic mechanism of the RNN oscillatory solution by developing a reduced variable description. Close inspection of the RNN firing rate time courses (Fig. 1c) reveals that many of the units have highly similar activity patterns allowing us to identify four subpopulations: excitatory and inhibitory units that fire with a phase between (0.8,1) and (0, 0.2) of the "tap" time—which we will refer to as the tap excitatory (Tap-E) and tap inhibitory groups (Tap-I), respectively, as well as units that fire with a phase between 0.2 and 0.8 of the model "tap" times—the inter-tap excitatory (Int-E) and inter-tap inhibitory groups (Int-I), respectively (Fig. 2a). The Tap-E, Int-E, Tap-I and Int-I account for 344/400 (86%), 56/400 (14%), 72/100 (72%), and 28/100 (28%) of the units, respectively. Simulations show that silencing the entire Int-E group in the trained RNN does not destroy the ability of the network to form oscillations, though it affects the frequency range of the network output (Fig. S2). Silencing any one of the remaining groups prevents the network from oscillating or from producing oscillations at a wide enough range of frequencies suggesting that the remaining three groups are necessary to the mechanism (Fig. S2).

Motivated by this observation, we define a three-variable reduced model based on the three necessary RNN subpopulations: Tap-E (x), Tap-I (y), and Int-I (z) (Fig. 2b). The variables x, y, and z correspond to the subgroup-averaged activities of the hidden units. The evolution of these variables is defined by rate equations modeled after the formulation of the RNN:

$$\tau_x x' = -x + W_{xx}F_x(x) + W_{xy}F_y(y) + W_{xz}F_z(z) + W_x^{in}I_{cc} + I_{bx} \quad (1)$$

$$\tau_y y' = -y + W_{yx}F_x(x) + W_{yy}F_y(y) + W_{yz}F_z(z) + W_y^{in}I_{cc} + I_{by} \quad (2)$$

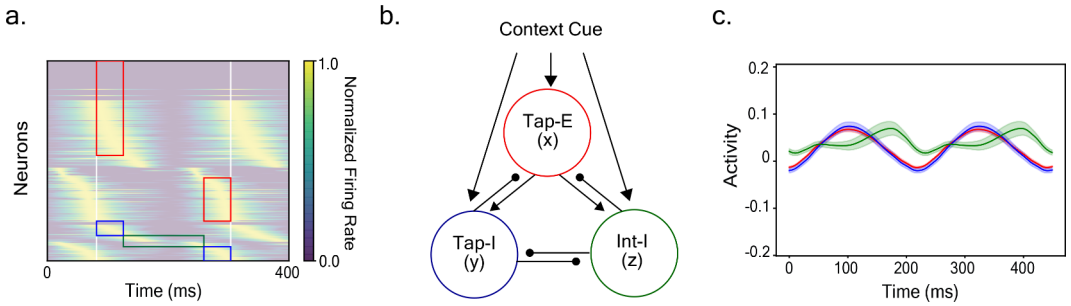

Fig. 2. Distinct subpopulations in the RNN allow for the formulation of a three-variable reduced model. (**a**) Four distinct neuronal populations in the RNN are identified by first sorting the normalized firing activity (noise-free simulation) by peak firing time and grouping according to phase of peak firing rate. Example sorting during the continuation phase is shown for a stimulus of 5 Hz ($I_{cc} = 0.5$). Excitatory and inhibitory units are sorted separately. Excitatory and inhibitory units that peak within a phase of 0.2 to the tap times (white vertical lines) form the tap excitatory and tap inhibitory groups (red and blue boxes, respectively). Inhibitory units that have peak firing rates with phases between 0.2 and 0.8 form the inter-tap inhibitory (green box) group. (**b**) Schematic of reduced model with three populations of units defined in panel a. Connectivity weights between groups are taken to be group-averaged RNN connection weights except for a few weights (see "Methods"). (**c**) Time course of the pre-rectified activity (x-variables in Eq. (1) of "Methods") of RNN hidden units (averaged over units) according to group membership: red, blue and green curves correspond to the average Tap-E, Tap-I and INT-I time courses, respectively. Time course shown during the continuation (no stimulus input) phase ($I_{cc} = 0.5$) and shading indicates $\pm$ standard error of the mean computed across units from the RNN simulated without noise during the continuation phase and after initial transients have settled, so that the individual units show their periodic response.

Parameter	Value	Parameter	Value
W_{xx}	8.949	I_{bz}	−0.1
W_{xy}	−7.40	τ_x, τ_y	10 *ms*
W_{xz}	−2.952	τ_z	50 *ms*
W_{yx}	9.123	a_x	150.134
W_{yy}	−7.386	b_x	−0.476
W_{yz}	−3.072	c_x	0.007
W_{zx}	8.935	a_y	62.873
W_{zy}	−4	b_y	0.481
W_{zz}	−1	c_y	0.016
W_x^{in}	0.048	a_z	87.0
W_y^{in}	0.055	b_z	−0.781
W_z^{in}	0.068	c_z	0.012
I_{bx}, I_{by}	0		

Table 1. Reduced model parameters.

$$\tau_z z' = -z + W_{zx} F_x(x) + W_{zy} F_y(y) + W_{zz} F_z(z) + W_z^{in} I_{cc} + I_{bz} \tag{3}$$

$$F_k(x) = c_k \log\left(1 + \exp\left(a_k x - b_k\right)\right), k = x, y, z \tag{4}$$

where connectivity weights between units and the input weight on the context cue (I_{cc}) are given by W and W^{in} , respectively. The bias and time constant for each population k are given by I_{b_k} and τ_k, respectively. Since we are interested in the structure of the model that maintains oscillations during the continuation phase, we drop the stimulus tone input so that the only external input to the reduced model is the context cue (I_{cc}). To approximate the effects of the RNN units' heterogeneous properties, we replace the non-linearity in the reduced form with a soft-plus function, $F(\cdot)$, that is fit to the averaged ReLU functions for each subpopulation (Eq. (4)). Finally, we take the connectivity weights to be the group average weights from the RNN except for τ_z, W_{zy}, W_{zz} which we have hand-tuned (see "Methods"). Values for the model parameters are found in Table 1. The reduced model is simulated with zero noise, throughout.

As a first step, we look at the subgroup-specific activity of the RNN hidden units. Averaging the hidden unit activity in the RNN according to subgroup (Fig. 2c) revealed low amplitude activity for all three groups.

Furthermore, the Tap-E and Tap-I groups appeared to be highly synchronized (firing rate balanced) and nearly identical while the Int-I group showed a phase offset in its time of peak firing compared to the Tap groups.

Analysis of the three-variable model

Qualitatively, we see agreement between the dynamics of the reduced model and the RNN. For small and large values of context cue (I_{cc}), the system converges to a steady state (Fig. 3a and c, respectively). For intermediate values of I_{cc}, the reduced model displays oscillations (Fig. 3b) with x (red) and y (blue) variables highly synchronized and z (green) showing a phase delay, in agreement with the RNN behavior of the averaged Tap-E, Tap-I and Int-I time courses shown in Fig. 2c. The RNN dynamics also converge to a steady state for I_{cc} values beyond the RNN's extrapolation regime, $I_{cc} < 0$ and $I_{cc} > 1.1$ (not shown). The reduced model produces a wide range of oscillation frequencies spanning 1 to 17 Hz (Fig. 3d) which overlaps with the produced frequency range of the RNN (Fig. 1f). The rise of ITI with decreasing I_{cc} is steep and localized rather than gradual as in the RNN; the difference is perhaps due to a lack of mechanisms for sequential activity or explicit account of heterogeneity in the reduced model.

The parameter regimes that lead to oscillations can be identified in a compact description, the bifurcation diagram, of the response dynamics (Fig. 3e) using I_{cc} as the control parameter. For I_{cc} negative there can be multiple steady states, as explained in the Supplementary Information Sect. 3 and Fig. S3, but in this parameter regime, only one is stable (stable in red, unstable in black). Since $x, y,$ and z are negative for this stable state, this corresponds to a state of no firing consistent with the time course shown in Fig. 3a. For large enough I_{cc} (Fig. 3e red line for $I_{cc} > 1.57$), there is a single stable state. Oscillations exist for intermediate I_{cc} with trough to peak amplitude (Fig. 3e, green for max and min of oscillation during a cycle) decreasing and frequency increasing as I_{cc} increases. The change in amplitude and period of the oscillations (as a function of context cue) can also be seen when projecting of the periodic solution onto the x vs. z plane (Fig. 3f). Note that points along each trajectory are plotted at each 1 *ms* increment, implying that longer period trajectories slow down near the southwest portion of the trajectory.

To understand how oscillations arise and disappear in this system, we take a closer look at the how the fixed points change stability. Starting at small values of the context cue ($I_{cc} < 0.19$), there are three fixed points whose existence can be seen in Fig. 3e (black and red lines). As the context cue increases, the lower two branches of the bifurcation curve meet at a saddle node point and disappear. This is the parameter value at which the system changes between a regime with a stable steady state attractor to an oscillatory regime. The solution that appears for $I_{cc} = I_{cc}^{HC}$ has large amplitude and zero frequency (infinite period); it's called a homoclinic orbit, a unique trajectory that approaches the saddle-node as t tends to plus or minus infinity (illustrated in Fig. S3a with a simple canonical model). The oscillation period is long for I_{cc} near, but just greater than I_{cc}^{HC} and decreases with

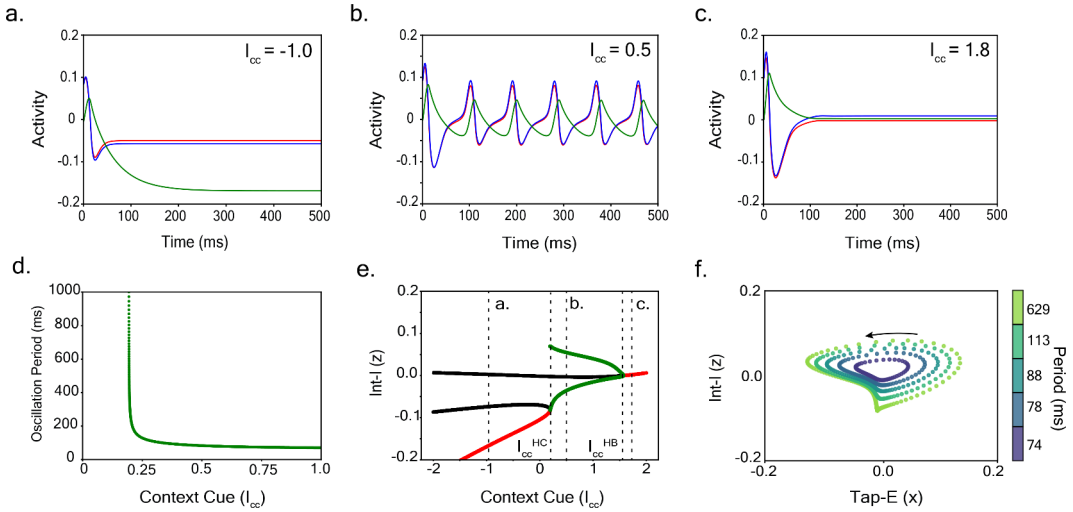

Fig. 3. In the reduced model, long-period oscillations arise from a homoclinic bifurcation and disappear as short-period oscillations via a Hopf bifurcation. (**a**) Example time courses of x (red), y (blue), and z (green) variables reduced model (Eqs. (1)–(4)) simulated with context cue $I_{cc} = -1.0$ and $\tau_z = 50ms$. (**b**, **c**) Time courses of the reduced model for context cues $I_{cc} = 0.5$ and $I_{cc} = 1.8$. Colors are the same as in **a**. Time courses in panels **a**., **b**., and **c**. are simulated without noise. (**d**) Oscillation period vs. context cue for the reduced model shown for $\tau_z = 50\ ms$. (**e**) One dimensional bifurcation diagram of the reduced model showing the effect of context cue (I_{cc}) on model dynamics. The homoclinic and Hopf bifurcation points are indicated by I_{cc}^{HC} (0.19) and I_{cc}^{HB} (1.57), respectively. Red and black lines indicate stable and unstable fixed points, respectively. Green lines indicate the max and min z-values of the stable periodic orbit. Dashed lines at context cue levels of $I_{cc} = -1$, $I_{cc} = 0.5$ and $I_{cc} = 1.8$ correspond to panels a, b, and c, respectively. (**f**) Example oscillatory trajectories for context cue values $I_{cc} = 0.197, 0.3, 0.5, 0.8$ and 1.2 (gradients from yellow to purple; dots plotted for every 1 *ms* timestep) projected into the two-dimensional phase plane, z vs. x. Direction of flow is counterclockwise (black arrow).

increasing I_{cc}. At large values of context cue ($I_{cc} > 1.57$), there is only one fixed point and it is a stable attractor. As the context cue is decreased from the right side of the Fig. 3e, oscillations emerge with small amplitude by destabilizing the steady state via a Hopf bifurcation (I_{cc}^{HB}) (Fig. S3b). These dynamical features reveal that for context cues less than I_{cc}^{HC}, the reduced model converges to a stable fixed point (e.g., Fig. 3a); for context cues between I_{cc}^{HC} and I_{cc}^{HB}, the system shows stable oscillations (e.g., Fig. 3b); and for context cue values above I_{cc}^{HB}, the system again converges to a stable fixed point via a spiraling attractor (e.g., Fig. 3c). Furthermore, given that the reduced model is meant to inform the RNN, these observations allow us to make a prediction for the RNN: as the context cue increases, oscillations arise as long period, large amplitude oscillations from a homoclinic bifurcation and disappear as short period, small amplitude oscillations via a Hopf bifurcation. Further details about the oscillation dynamics are provided in the Supplementary Information Sect. 3 and Figs. S4, S5.

Assuming E-I firing rate balance allows further reduction to a two-variable model

The tight balance observed between the Tap-E (x) and Tap-I (y) variables suggests a further reduction of our model to a two-variable system by assuming that x and y are related by a scalar multiple taken to be the slope of the overlapping region of the nullclines (Fig. 3e solid line $x, y \geq 0$). Introducing this assumption ($y = 1.1x$), the resulting two-variable system (Fig. 4a) is amenable to phase plane analysis without the need for projecting from a higher dimensional space. The oscillation trajectory strikingly resembles the projected limit cycle of the $x - y - z$ model (compare Fig. 4a to 3f). An example trajectory is plotted at every 1 *ms* timestep (grey) and shows a slowing down near the southern end of the trajectory as observed in the three-variable reduced model. Furthermore, the qualitative dynamical features are preserved: oscillations emerge with long period via a homoclinic bifurcation that then disappear via a Hopf bifurcation as I_{cc} changes from medium to high values (compare Fig. 4b to 3e). Finally, we see that the system retains small amplitude oscillations and a phase delay in the peak activity of the z variable (compare Fig. 4c to 3b).

Analysis of RNN dynamics confirms predictions from reduced model

Having identified the reduced model's dynamic structure and attractor transitions, we searched for the corresponding features in the RNN. We identified a two-dimensional plane that captures the key fixed points affecting the RNN dynamics and investigated the changes in the RNN dynamics as a function of changes in context cue, (I_{cc}) in this plane. To identify the relevant plane, we performed Principal Component Analysis on RNN time courses from trajectories initialized on the balance manifold in the non-oscillatory regime (see "Methods") and chose the plane spanned by the first two principal components (this captures 80% of the variance for all noise-free RNN trajectories). Note that the projection plane was kept constant for the following analysis.

First, consider the RNN behavior for context cue values for which the system converges to a unique stable steady state. The top row of Fig. 5a shows the projection of three example trajectories (red curves) for three small values of the context cues ($I_{cc} = -0.5, -0.2$ and 0.1). While it may appear that the example trajectories cross against the direction of the vector field (white arrows), it is important to remember that the vector field changes over the course of the trajectory; the shown vector field is computed at the final time point in the trajectory, after the system reaches its steady state (the complete dynamics over the course of the trajectory can be seen in Supp. Movie 2). From the shown vector field, we see evidence for three fixed points of the system (Fig. 5a top row, pink circle and triangles). As the context cue increases from −0.5 to 0.0, the stable node (Fig. 5a top row pink circle) and saddle (Fig. 5a top row left pink triangle) come together and eventually coalesce at the onset of oscillations at $I_{cc} = 0.18$ (not shown). Recall, we had predicted this behavior from our bifurcation analysis of the reduced model: a long-period oscillation appears as a (large amplitude) homoclinic orbit as the saddle-node bifurcation point is approached from negative to low values of context cue (Figs. 3g and 4b, red and black branches come together at the start of the green branch). The RNN dynamics transition from an excitable to oscillatory mode.

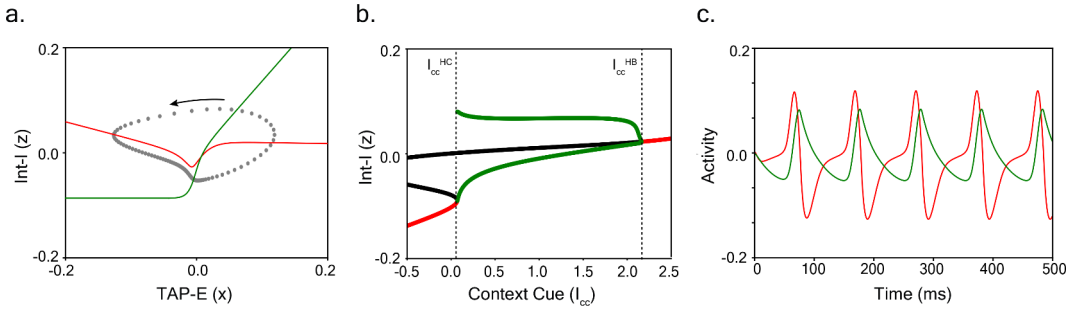

Fig. 4. Two variable reduced model. (**a**) Phase plane representation of the two-variable reduced model (obtained by assuming strong E-I balance: $y = 1.1x$) for Tap-E (x) and Int-I (z) variables. The limit cycle trajectory (grey; dots plotted for every 1 *ms* of trajectory) is a global attractor, $I_{cc} = 0.2$; it crosses the x -nullcline (red) vertically and the y-nullcline horizontally (green). The flow is counterclockwise (black arrow); peak of the Int-I inhibition occurs on the downstroke of the Tap-E excitation. (**b**) Bifurcation diagram of the two-variable reduction with context cue (I_{cc}) as control parameter resembles that for three-variable model (Fig. 3e). The limit cycle attractor (green) transitions to a stable steady state (red) via homoclinic and Hopf bifurcations for $I_{cc}^{HC} = 0.07$, $I_{cc}^{HB} = 2.16$, respectively. Black lines indicate unstable fixed points. (**c**) Example time course of x and z variables for the same trajectory shown in (**a**)

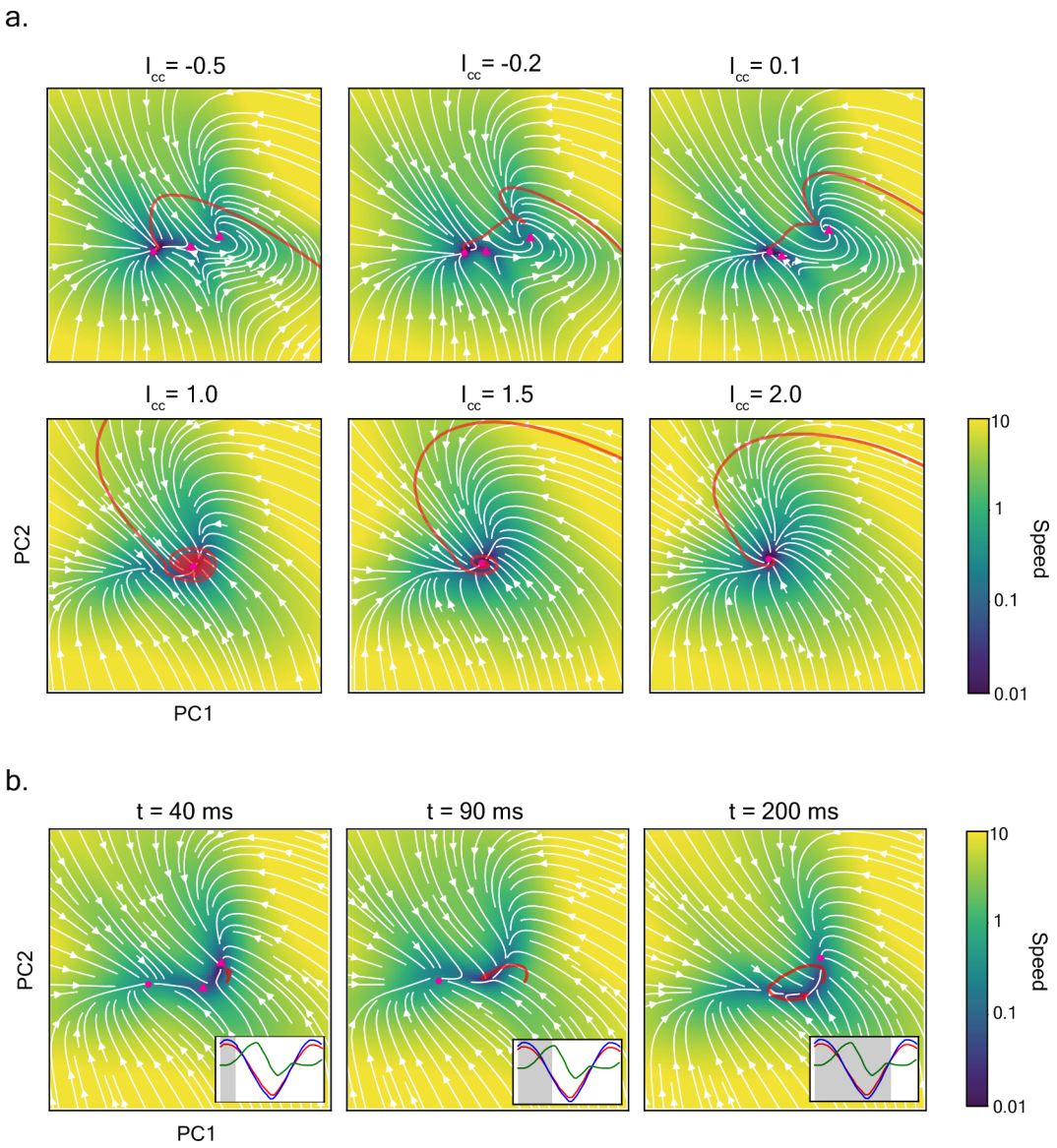

Fig. 5. RNN dynamics agree with reduced model predictions. (**a**) Two-dimensional projection of RNN dynamics for different values of context cue. In the top row, context cues $I_{cc} = -0.5, -0.2$ and 0.1, there is a global stable steady state attractor on the left-hand side (dark purple region) and all trajectories end up there. In the bottom row, context cues $I_{cc} = 1.0, 1.5$ and 2.0, the dynamics show a stable attractor again but now the trajectories spiral towards this fixed point. White arrows indicate the vector field in the projection plane; projected directions computed after the dynamics have evolved to be near steady state. Red curves indicate example trajectories. Color indicates speed with dark purple corresponding to slow speeds and bright yellow corresponding to fast speeds. Pink markers indicate fixed points deduced from flow maps with circles indicating stable fixed points and triangles indicating unstable fixed points. (**b**) Two-dimensional projection of RNN dynamics for $I_{cc} = 0.5$ at three different time points during one cycle (period = 222 *ms*) of the oscillating trajectory. Colors are the same as in **a**. Pink circles and triangle indicate fixed points that are transient. Since the trajectory is oscillating, the vector field changes over time and the transient fixed points appear and disappear over the course of the oscillation. Inset shows averaged time courses for Tap-E (red), Tap-I (blue) and INT-I (green) sub-populations from the RNN with grey shading to indicate depicted time point relative to the time course.

The second row of Fig. 5a shows example trajectories for larger values of context cue, $I_{cc} = 1.0, 1.5$ and 2.0, projected into the same plane. The vector field for these projections is also computed once the system has reached steady state. An example trajectory (red) spirals towards the fixed point and as the context cue is increased, the fixed point becomes a stronger attractor and less spiraling is observed (Fig. 5a bottom row, compare panels from left to right). This behavior is indicative of a Hopf bifurcation in the RNN (at a context cue around $I_{cc} = 0.95$) and corroborates its prediction from the reduced model (Figs. 3e and 4b, for $I_{cc} > I_{cc}^{HB}$).

The behavior of the RNN in the oscillatory regime can be seen at three different snapshots in time in Fig. 5b. Note that the pink markers indicate "pseudo fixed points" of the system at the depicted point in time. These points are not fixed points of the system since they will disappear over the course of the oscillation. Pseudo fixed points are transient and not uncommon in dynamical systems where flows along manifolds are examined with some (slower) variables held constant. We show that the pseudo fixed points in these RNN projections and their ensuing effects on the dynamics may be associated with similar objects found in the three-variable reduced model (see Supplementary Information Sects. 3 and 4).

At the level of individual units, one signature of a homoclinic bifurcation is an elongation of near constant activity in the firing time course as the oscillation period increases. This happens because, after the saddle and attractor coalesce, the flow near to where the fixed points had been is slow. In other words, this "ghost" of the fixed points slows the oscillation. As the context cue increases, the ghost has less of an effect thereby reducing the window of slow speeds and thus decreasing the oscillation period. We see this elongation of the period of near constant activity with decreasing I_{cc} prominently in the reduced model as well as the RNN (Fig. S6). Finally, another feature of the reduced model is that the oscillation amplitude (the distance between the green curves in Fig. 3e) decreases as the Hopf bifurcation is approached—this is confirmed in the RNN (Fig. S6).

Discussion

We developed a neurally plausible mechanism for the synchronization and continuation task that is controllable via the context cue to oscillate over a frequency range that matches the range of beat frequencies perceived by macaques and humans. By developing a reduced model, we were able to predict and confirm the dynamical features that gave rise to oscillations in the RNN. Based on our identification of three distinct neural populations in the RNN network that are necessary for oscillations, our reduced model is comprised of an excitatory and one group of inhibitory units that form a balanced sub-network with a second group of inhibitory units that create an elongated window of inhibition that allows the system to oscillate.

Our choice to include the context cue in the model formulation is meant to suggest that an animal performing the same task would develop a map from the stimulus inter-tone times to the produced rhythm. We do not describe the learning process for this map. Furthermore, our model is idealized to a functional brain region for the representation of the oscillation, such as the supplementary motor area or premotor cortex. We assume that the context cue is coming from another brain region that associates a tonic drive to the oscillator rhythm that matches the inter-stimuli tone interval. In other models[37,39–41], a tonic input drive is a common controllability parameter. Finally, we found that introducing the context cue is necessary for the model to generalize between the trained stimulus frequencies.

The oscillation mechanism can be differentiated from previous models for rhythmic timing. Unlike entrained oscillator models[5,42–45] where an endogenous brain rhythm entrains to the stimulus frequency, our model does not depend on sustained entrainment. Rather, the internally generated oscillation is tuned to the appropriate frequency through adjustment of the context cue. Unlike our model, entrainment alone is not be able to maintain the oscillation in the absence of a rhythmic stimulus (the continuation phase). The mechanism that we discovered through the RNN is more closely aligned to that of tunable oscillator models such as the Beat Generator Model[33,46] or the SAM-MPM model[39] where the oscillation frequency adapts to match that of the stimulus through learning rules. However, unlike the Beat Generator model which achieves slow oscillation by the incorporation of a slow variable, our model shows long period oscillations arising through a homoclinic bifurcation. Moreover, our model can be viewed as a novel network-based circuit model in contrast to a cellular oscillator model in the Beat Generator framework with the context cue (I_{cc}) as the control parameter that sets the oscillation period. In the context of predictive coding theory for rhythmic timing[47], our model can be viewed as the mechanism underlying the dynamic neural population trajectories in the medial premotor cortex that encode the distribution of timing of events in a rhythmic sequence. Finally, other models have addressed more complex rhythms[48,49] but have done so at the cost of losing the ability to explain the neural dynamics observed on the short timescale as we have done here. Future work could apply the current framework to more complex tasks. Our RNN training paradigm utilized a sinusoidal output to align our model with dynamical features observed in the data such as circular trajectories, sequential firing patterns, and the ability of the model to generalize to untrained frequencies. Our investigation into other choices for the output found that non-sinusoidal, such as pulsatile-like, time courses replicated fewer of the observed features from the neural data.

With insights garnered from the RNN and reduced models, we suggest some experimental predictions that could help test our model. First, our models have two distinct sub-populations of inhibitory units: the Tap-I group that is highly synchronized to the Tap-E units and the Int-I group that shows sequential structure in the firing pattern. This architecture is consistent with previous experimental findings in other brain areas such as the visual cortex[50] and auditory cortex[51], where parvalbumin-expressing inhibitory interneurons pair up with co-tuned excitatory cells as well as with somatostatin interneurons. Oscillations are among the growing evidence for functional importance of such canonical circuitry[52,53]. A model of gamma oscillations in visual cortex is based on strong E-I balance based on parvalbumin-expressing inhibitory interneurons with somatostatin interneurons playing a critical role in producing the oscillations[54]—albeit this is for higher frequency oscillations than considered here. Taken together, these results suggest that the Tap-I units in our RNN might be mapped onto parvalbumin-expressing interneurons while Int-I units might be mapped onto somatostatin interneurons. We therefore suggest that somatostatin-like functional inhibition provides a mechanism for the elongated inhibition. Second, subject to neuronal subgroups being identifiable in experimental data, the firing rate time courses for this E-I pairing are expected to be highly synchronized. Third, the presence of a homoclinic bifurcation in our model suggests that as the oscillation period slows, there should be a sub-threshold plateau of near constant activity that immediately precedes firing during the "Tap" phase of activity. Another expected signature of homoclinic behavior is the lack of resonance in the autocorrelation of ITIs. In contrast, resonance

would be expected for the shorter period oscillations associated with a Hopf bifurcation mechanism[55]. Fourth, the formulation of the reduced model allowed us to disregard the inter-tap excitatory group suggesting that these cells could be silenced without destroying the network oscillation (albeit the overall frequency may change; Fig. S2). Interestingly, from the reduced model, we also found that we can silence the Int-I group and get oscillations along the firing rate balance manifold (Fig. S5).

In the three-variable reduced model, the variables x and y are in strong balance as exemplified by the highly synchronized time courses of x and y (Fig. 2c). This balance is reminiscent of an inhibition-stabilized-network balance operating in a regime where slow modulation of input leads to x and y both increasing or both decreasing in super-threshold activity ranges[56,57]. This behavior is unlike the noise-driven activity found near a subthreshold E-I balance state in the fluctuation-dominated regime described in spiking networks[58].

The RNN shows sequential firing thereby achieving ITI timescales of hundreds of milliseconds despite the utilization of 10 *ms* time constants. In an effort to retain simplicity in our three-variable model, we chose the time constant, τ_z, of the Int-I group to be longer than that of Tap-E and Tap-I. We propose the interpretation that, functionally, sequential firing in the RNN (and in the neural data) serve to provide an extended duration of inhibition, acting effectively as if Int-I decays more slowly. Sequential activity has been found in many parts of the brain for a variety of different tasks (not restricted to timing) although the underlying mechanism for it has been shown to vary between regions. For instance, we found little evidence of strong chain-like connectivity suggesting that our sequential structure is unlike that seen in songbirds[59] and is more related to sequences formed from parametric gradients of excitability as seen in the hippocampus[60] and in the posterior parietal cortex[61]. During interval timing tasks, sequential firing activity has also been documented in the premotor cortex and the striatum[62]. Models have been used to replicate this finding[38] where, unlike our model, the sequential activity arises from a structured difference between incoming excitation and inhibition into a unit. Interestingly, although more units are involved in slower frequency oscillations, we saw no increase in dimensionality (as measured by number of principal components need to explain the same level of variance) for longer sequences. In future work, it would be worthwhile to explore the effect of explicitly adding chain-like[63] structures into the reduced model or to train a spiking network with spike-time-dependent plasticity with heterosynaptic competition on the synchronization and continuation task and to test if sequential activity emerges as suggested by previous computational studies[64].

The ease of training RNNs and their success at replicating features of neural dynamics seen in experimental data has fueled interest in uncovering the mechanisms underlying the RNN solutions to tasks[65–72]. While we do not provide a one-size-fits-all approach to reducing a high dimensional network to an interpretable circuit, we do show how close inspection of the RNN unit activity can lead to the formulation of a circuit model with distinct neuron subpopulations. In doing so, we are better able to understand the functional mechanisms that underlie activity of the hidden layers of the RNN beyond those layers simply being a black box.

By training and then dissecting an E-I RNN, we have generated a novel, neurally plausible network model of rhythmic timing. Although previous work has described the occurrence of homoclinic bifurcations in rate models for neural networks[73–76], finding such a dynamical feature in a timing task is new. Furthermore, our model and analysis not only offers interpretations of neural dynamics previously observed in experimental data but also posits functional significance for different interneuron subtypes.

Methods

RNN Formulation

The RNN consists of 500 firing rate units with distinct excitatory (80%) and inhibitory (20%) populations.

The dynamics of the RNN are described by the following equations:

$$\tau x' = -x + W^{rec} r + W^{in}_{cc} I_{cc} + W^{in}_{stim} I_{stim} + b^{rec} + \sqrt{2\tau\sigma^2_{rec}}\zeta \tag{5}$$

$$o = W^{out} r + b^{out} \tag{6}$$

$$r(x) = \max(0, x) \tag{7}$$

where $x \in \mathbb{R}^N$ is activity (analogous to a mean voltage[77]) of the N (=500) RNN hidden units which are transformed into firing rates r via the ReLU non-linearity in Eq. (7). The recurrent weight matrix is given by W^{rec}, the vector b^{rec}denotes the biases for each unit and ζ denotes Gaussian white noise drawn independently for each unit. The input weight scalars are given by $W^{in}_{cc}, W^{in}_{stim} \in \mathbb{R}^N$. The two types of input: context cue and stimulus pulses are indicated by I_{cc} and I_{stim}, respectively. The time constant of $\tau = 10\ ms$ was used for all units. The output (o) of the network is a linear combination of the firing rates with the weight matrix $W^{out} \in \mathbb{R}^N$ and a bias term b^{out}. The RNN was trained with additive white noise for the stimulus input ($\sigma_{in} = 0.01$) and for each hidden unit ($\sigma_{rec} = 0.01$) . Self-connections were not permitted in the network to encourage sequential firing[78]. The quantities $W^{in}_{cc}, W^{in}_{stim}, W^{rec}, W^{out}, b^{rec},$ and b^{out} are 'learned' during training of the RNN.

RNN training

The RNN was trained using continuous-time dynamics—Eqs. (5)–(7). Integration in time was by Euler's method with a step size, dt, equal to 1 *ms*. The RNN was trained on the NYU computing cluster for 72 h and reached an $RMSE = 0.13$.

For each of the inputs (I_{cc}, I_{stim}) to the network, the Euler step increment at time t is modeled using:

$$\Delta_I = \max\left(0, dt\ I_k + \sqrt{2\alpha\sigma^2_{in}}\boldsymbol{N}(0,1)\right)$$

where $I_k = I_{cc}$ or $I_k = I_{Stim}$, $\alpha = dt/\tau$ and $\sigma_{in} = 0.01$ corresponds to the standard deviation of the input noise.

The network was trained using the Pycog library[79] using stochastic gradient descent algorithm with a learning rate of 0.01. We used the mean squared error cost function between the network output (o) and target function ($\hat{o}$):

$$RMSE = \sqrt{\sum \left(o\left(t\right) - \hat{o}\left(t\right)\right)^2}$$

Weights $W^{out}, W^{in}, W^{brec}, b^{rec}, b^{out}$ were all trained. Unless otherwise specified, all other training parameter are the defaults as given in Song et al.[79].

Synchronization and continuation task

The RNN was trained with seven different pairs of inputs for $T = 2s$ (frequencies = 2, 3, 4, 5, 6, 7, 8 Hz, equivalently periods = 500, 333.3, 250, 200, 166.6, 142.8, 125 ms). The inputs are of two types: I_{stim} which was provided for $1\ s$, and I_{cc} which was provided for $2\ s$ and defined as follows:

$$I_{Stim} = \left\{ \begin{array}{ll} 1 & for\ T_{onset} + T_{tones} \leq t \leq T_{onset} + T_{tones} + 10\,ms\ and\ t \leq T_{onset} + 1s \\ 0 & otherwise \end{array} \right\}$$

where T_{tones} = [0, period, 2*period, …]

$$I_{cc} = \left\{ \begin{array}{ll} \frac{100}{period} & for\ t \geq T_{onset} \\ 0 & otherwise \end{array} \right\}$$

The output target was defined as

$$\hat{o}\left(t\right) = \left\{ \begin{array}{ll} \frac{\cos\left(\frac{2\pi t}{period}\right)+1}{2} & for\ t \geq T_{onset} \\ 0 & otherwise \end{array} \right\}$$

T_{onset} was random on each trial and drawn from Uniform(0,100 ms).

Generalization performance

To test how the trained RNN generalizes to previously unseen frequency inputs, we gave the trained RNN context cue inputs in the range of 0.1 to 1.1 with a step size of 0.01. RNN inter-tap intervals were computed from 50,000 *ms* of the continuation phase RNN output and each ITI was plotted in Fig. 1f.

Computation of RNN unit dynamics

Figure 1 panels **c**-**e**: Statistics of RNN trajectories are computed by first simulating the RNN with noise ($\sigma_{in} = 0.01, \sigma_{rec} = 0.01$) only during the continuation phase (i.e. the RNN was not provided with a stimulus pulse sequence) for 52,000 timesteps (dt = 1 *ms*) at each I_{cc}. The first 2,000 timesteps are deleted to make sure that there are no transients and units are in their near-periodic activity patterns. Each 50,000 timestep time course is then divided into its inter-tap-intervals (ITIs) and statistics are computed over the corresponding set of ITI's for each I_{cc}.

Figure 2 panels **a**, **c**: RNN trajectories are simulated without noise ($\sigma_{in} = 0.0, \sigma_{rec} = 0.0$) only during the continuation phase. The first 2000 *ms* of a time course are dropped to make sure transients have passed and units are in their effectively periodic activity states. Each unit is normalized so that the firing activity falls between 0 and 1. The normalized units are sorted by time of peak firing (when firing activity is at 1) during their first ITI after the 2000 *ms* of dropped time course (excitatory and inhibitory are sorted separately). Units that hit their max firing rate with a phase of within 0.2 of the model "tap" times are labeled 'tap' units and the remaining units are labeled inter-tap units.

Analysis of RNN dynamics

The projection plane for the analysis of RNN dynamics was kept constant for all panels in Fig. 5. This plane was selected by first simulating one hundred T = 500 *ms* noise-free trajectories of RNN units with context cues from the range of −1 to 0.2. Units were assigned to one of Tap-E, Tap-I or Int-I groups based on their firing phase during a noise-free simulation of a 5 Hz stimulus ($I_{cc} = 0.5$) and group membership was fixed thereafter. On each simulated trajectory, the initial conditions for the hidden variables were selected so that the units that were grouped into Tap excitatory and Tap inhibitory units had the same value which was randomly chosen on each trial from the Uniform (0,1) distribution. All other units were individually randomly initialized in Uniform (−1,1). These trajectories were then concatenated together into a matrix of size T * 100 x N where T is the number of timesteps computed for each trajectory and N is the number of units in the RNN. Principal component analysis was performed on this matrix after it was normalized for each unit. The plane spanned by the first two principal components was selected for the projection. For noise-free trajectories, the top 2 principal components (PCs) accounted for 80% of the variance and the top 3 PCs accounted for 85% of the variance.

This choice of initial conditions for these trajectories was motivated by the observation that in the reduced model, several fixed points lie on the balance manifold (in the region where x and y nullclines are almost overlapping for $x, y \geq 0$). Initializing trajectories at different points along this manifold meant that the resulting

trajectories should have their flows affected by these fixed points or the resulting strong vector field near to the balance manifold if the reduced model is a good proxy for the dynamics in the full RNN. PCA would then represent the directions in state-space that capture these dynamics as well. Indeed, we find that initializing points in the RNN near the predicted balance-manifold allowed PCA to find a plane capturing the predicted unstable fixed points. Initializing trajectories randomly in the full 500-dimensional space, in contrast, will capture the stable fixed point since all trajectories head toward the attractor but is unlikely to create trajectories that pass close enough to the balance manifold since the region of space that it influences is relatively small. In agreement with this, we found that initializing trajectories randomly (instead of around the balance manifold) and then performing PCA consistently only captured the stable fixed point but not the predicted unstable fixed points. Although we don't know with certainty where the balance manifold lies in the full RNN, initializing all units that were grouped into either tap excitatory or tap inhibitory units to the same (random) value for each simulated trajectory worked well. It was important that we capture the plane with the fixed points on the balance manifold because, according to the reduced model, these fixed points are involved in the homoclinic bifurcation and we wanted to investigate if this was the case in the RNN as well.

Having defined the projection plane, the vector field was computed by first defining a vector, I_t^{proj}, for each point in the projection plane to take the values in the plane for the first two components and filling the remaining components with their corresponding value at time t for a given example trajectory. This vector was then projected back to the full RNN state space, I_t^{orig}, and simulated one step forward using the RNN to get, I_{t+1}^{orig}. I_{t+1}^{orig} was then projected back into the space defined by the principal components to get I_{t+1}^{proj} and the vector flow at point I_t^{proj} was given by the direction of $I_{t+1}^{proj} - I_t^{proj}$. Speed is the L^2norm of the RNN hidden unit activity, dx_i/dt, computed similarly with dt=1 *ms*:

$$speed = \sqrt{\sum_{i=0}^{500} (dx_i/dt)^2}\ .$$

Reduced model

The three- (Eqs. (1)–(4)) and two-variable reduced models were discretized for simulations using Euler's method with a timestep of $dt = 1\ ms$. Reducing the timestep to $dt = 0.1\ ms$ did not noticeably change the results.

Weights for the reduced model were taken to be the between sub-population averages from the RNN except for τ_z, W_{zy}, W_{zz}. As discussed in the Supplementary Information, the τ_z was increased to have oscillation periods in the reduced models more closely match up with the RNN. The exact value of τ_z does not affect the key findings as long as it is set to be in the oscillatory regime (Fig. S4b). The RNN averaged connectivity weights were adjusted for W_{zy} (from -7.817 to -4.0) and W_{zz} (from -2.599 to -1.0) so that the model shows oscillations that have a frequence dependence on context cue. More specifically, the population averaged weights put the reduced model in the regime where the intersection of the nullclines occurred in the linear region of all of the nullclines. When this happens, there is no frequency dependence or change of steady state stability on the context cue because a small shift of the nullclines (due to changes in context cue) does not result in a change in slope of the nullclines at the point of intersection. However, if the nullclines intersect in the non-linear section, then small shifts of the nullclines (from changes in the context cue) do result in slope changes at the point of intersection of the nullclines and correspondingly, this results in changes in the oscillation frequency or change of steady state stability of the system. The two hand-tuned weights (W_{zy} and W_{zz}) were adjusted based on this principle to make the corresponding nullclines intersect in the non-linear region of the nullclines.

Bifurcation analysis

All bifurcation analysis was done using XPPAUT[80] and the resulting data was imported into python to create the figures.

Data availability

The datasets used and/or analyzed, as well as computer codes used for simulations, during the current study are available from the corresponding author on reasonable request.

Received: 13 June 2024; Accepted: 25 October 2024
Published online: 02 November 2024

References

1. Matell, M. S. & Meck, W. H. Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Brain Res. Cogn. Brain Res.* **21**, 139–170 (2004).
2. Buhusi, C. V. & Meck, W. H. What makes us tick? Functional and neural mechanisms of interval timing. *Nat. Rev. Neurosci.* **6**, 755–765 (2005).
3. Gibbon, J. & Malapani, C. Time perception and timing, neural basis of encyclopedia. *Cogn. Sci.* https://doi.org/10.1002/0470018860.s00426 (2006).
4. Cannon, J. J. & Patel, A. D. How beat perception co-opts motor neurophysiology. *Trends Cogn. Sci.* **25**, 137–150 (2021).
5. Large, E. W., Almonte, F. V. & Velasco, M. J. A canonical model for gradient frequency neural networks. *Phys. D* **239**, 905–911 (2010).
6. Grahn, J. A. Neural mechanisms of rhythm perception: current findings and future perspectives. *Top. Cogn. Sci.* **4**, 585–606 (2012).
7. Breska, A. & Ivry, R. B. Double dissociation of single-interval and rhythmic temporal prediction in cerebellar degeneration and Parkinson's disease. *Proc. Natl. Acad. Sci. USA* **115**, 12283–12288 (2018).
8. Chen, J. L., Penhune, V. B. & Zatorre, R. J. Listening to musical rhythms recruits motor regions of the brain. *Cereb. Cortex* **18**, 2844–2854 (2008).

9. Merchant, H., Harrington, D. L. & Meck, W. H. Neural basis of the perception and estimation of time. *Annu. Rev. Neurosci.* **36**, 313–336 (2013).
10. Grahn, J. A. & Rowe, J. B. Finding and Feeling the Musical Beat: Striatal Dissociations between Detection and Prediction of Regularity. *Cereb. Cortex* **23**, 913–921 (2013).
11. Bengtsson, S. L. et al. Listening to rhythms activates motor and premotor cortices. *Cortex* **45**, 62–71 (2009).
12. Grahn, J. A. & Brett, M. Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex* **45**, 54–61 (2009).
13. Ross, J. M., Iversen, J. R. & Balasubramaniam, R. The role of posterior parietal cortex in beat-based timing perception: a continuous theta burst stimulation study. *J. Cogn. Neurosci.* **30**, 634–643 (2018).
14. Kung, S. J., Chen, J. L., Zatorre, R. J. & Penhune, V. B. Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *J. Cogn. Neurosci.* **25**, 401–420 (2013).
15. Ivry, R. B. & Keele, S. W. Timing functions of the cerebellum. *J. Cogn. Neurosci.* **1**, 136–152 (1989).
16. Grube, M., Lee, K. H., Griffiths, T. D., Barker, A. T. & Woodruff, P. W. Transcranial magnetic theta-burst stimulation of the human cerebellum distinguishes absolute, duration-based from relative, beat-based perception of subsecond time intervals. *Front. Psychol.* **1**, 1946 (2010).
17. Teki, S., Grube, M., Kumar, S. & Griffiths, T. D. Distinct neural substrates of duration-based and beat-based auditory timing. *J. Neurosci.* **31**, 3805–3812 (2011).
18. Rao, S. M., Mayer, A. R. & Harrington, D. L. The evolution of brain activation during temporal processing. *Nat. Neurosci.* **4**, 317–323 (2001).
19. Karmarkar, U. R. & Buonomano, D. V. Timing in the absence of clocks: encoding time in neural network states. *Neuron* **53**, 427–438 (2007).
20. Walsh, V. A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* **7**, 483–488 (2003).
21. Repp, B. H. & Su, Y. H. Sensorimotor synchronization: A review of recent research (2006–2012). *Psychon Bull. Rev.* **20**, 403–452 (2013).
22. Repp, B. H. Sensorimotor synchronization: A review of the tapping literature. *Psychon Bull. Rev.* **12**, 969–992 (2005).
23. Gámez, J., Mendoza, G., Prado, L., Betancourt, A. & Merchant, H. The amplitude in periodic neural state trajectories underlies the tempo of rhythmic tapping. *PLoS Biol.* **17**, e3000054 (2019).
24. de Lafuente, V. et al. Keeping time and rhythm by internal simulation of sensory stimuli and behavioral actions. *Sci. Adv.* **10**, 20220103474812 (2024).
25. Patel, A. D. Vocal learning as a preadaptation for the evolution of human beat perception and synchronization. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20200326 (2021).
26. Stroud, J. P., Duncan, J. & Lengyel, M. The computational foundations of dynamic coding in working memory. *Trends Cogn. Sci.* **28**, 614–627 (2024).
27. Large, E. W. et al. Dynamic models for musical rhythm perception and coordination. *Front. Comput. Neurosci.* **17** (2023).
28. Russo, A. A. et al. Motor cortex embeds muscle-like commands in an article motor cortex embeds muscle-like commands in an untangled population response. *Neuron* **97**, 953–966e8 (2018).
29. Goudar, V. & Buonomano, D. V. Encoding sensory and motor patterns as time-invariant trajectories in recurrent neural networks. *Elife* **7** (2018).
30. Sohn, H., Narain, D., Meirhaeghe, N. & Jazayeri, M. Bayesian computation through cortical latent dynamics. *Neuron* **103**, 934–947e5 (2019).
31. Sussillo, D., Churchland, M. M., Kaufman, M. T. & Shenoy, K. V. A neural network that finds a naturalistic solution for the production of muscle activity. *Nat. Neurosci.* **18** (2015).
32. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
33. Bose, A., Byrne, Á. & Rinzel, J. A neuromechanistic model for rhythmic beat generation. *PLoS Comput. Biol.* **15**, e1006450 (2019).
34. Betancourt, A., Pérez, O., Gámez, J., Mendoza, G. & Merchant, H. Amodal population clock in the primate medial premotor system for rhythmic tapping. *Cell Rep.* **42**, 113234 (2023).
35. Patel, A. D. & Iversen, J. R. The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Front. Syst. Neurosci.* **8** (2014).
36. Beiran, M., Meirhaeghe, N., Sohn, H., Jazayeri, M. & Ostojic, S. Parametric control of flexible timing through low-dimensional neural manifolds. *SSRN Electron. J.* **111**, 739–753e8 (2021).
37. Zhou, S., Masmanidis, S. C. & Buonomano, D. V. Encoding time in neural dynamic regimes with distinct computational tradeoffs. *PLoS Comput. Biol.* **18**, e1009271 (2022).
38. Hardy, N. F. & Buonomano, D. V. Encoding time in feedforward trajectories of a recurrent neural network model. *Neural Comput.* **30**, 378–396 (2018).
39. Egger, S. W., Le, N. M. & Jazayeri, M. A neural circuit model for human sensorimotor timing. *Nat. Commun.* **11**, 3933 (2020).
40. Ausborn, J., Snyder, A. C., Shevtsova, N. A., Rybak, I. A. & Rubin, J. E. State-dependent rhythmogenesis and frequency control in a half-center locomotor CPG. *J. Neurophysiol.* **119**, 96–117 (2018).
41. Butera, R. J., Rinzel, J. & Smith, J. C. Models of respiratory rhythm generation in the pre-Botzinger complex. I. Bursting pacemaker neurons. *J. Neurophysiol.* **82**, 382–397 (1999).
42. Kim, J. C. & Large, E. W. Signal processing in periodically forced gradient frequency neural networks. *Front. Comput. Neurosci.* **9**, 152 (2015).
43. Kim, J. C. & Large, E. W. Multifrequency Hebbian plasticity in coupled neural oscillators. *Biol. Cybern.* **115**, 43–57 (2021).
44. Tichko, P., Kim, J. C. & Large, E. W. Bouncing the network: A dynamical systems model of auditory–vestibular interactions underlying infants' perception of musical rhythm. *Dev. Sci.* **24**, e13103 (2021).
45. Large, E. W., Herrera, J. A. & Velasco, M. J. Neural networks for beat perception in musical rhythm. *Front. Syst. Neurosci.* **9**, 159 (2015).
46. Byrne, Á., Rinzel, J. & Bose, A. Order-indeterminant event-based maps for learning a beat. *Chaos* **30**, 083138 (2020).
47. Snyder, J. S., Gordon, R. L. & Hannon, E. E. Theoretical and empirical advances in understanding musical rhythm, beat and metre. *Nat. Rev. Psychol.* **3**, 449–462 (2024).
48. Calderon, C. B., Verguts, T., Frank, M. J. & Thunderstruck. The ACDC model of flexible sequences and rhythms in recurrent neural circuits. *PLoS Comput. Biol.* **18**, e1009854 (2022).
49. Cannon, J. Expectancy-based rhythmic entrainment as continuous Bayesian inference. *PLoS Comput. Biol.* **17**, e1009025 (2021).
50. Znamenskiy, P. et al. Functional specificity of recurrent inhibition in visual cortex. *Neuron* **112**, 991–1000e8 (2024).
51. Li, L. Y. et al. A feedforward inhibitory circuit mediates lateral refinement of sensory representation in upper layer 2/3 of mouse primary auditory cortex. *J. Neurosci.* **34**, 13670–13683 (2014).
52. Bos, H., Oswald, A. M. & Doiron, B. Untangling stability and gain modulation in cortical circuits with multiple interneuron classes. *bioRxiv* 1–30. https://doi.org/10.1101/2020.06.15.148114 (2024).
53. Kumar, M. et al. Cell-type-specific plasticity of inhibitory interneurons in the rehabilitation of auditory cortex after peripheral damage. *Nat. Commun.* **14**, 1–23 (2023).
54. Veit, J., Hakim, R., Jadi, M. P., Sejnowski, T. J. & Adesnik, H. Cortical gamma band synchronization through somatostatin interneurons. *Nat. Neurosci.* **20**, 951–959 (2017).

55. Tateno, T., Harsch, A. & Robinson, H. P. C. Threshold firing frequency-current relationships of neurons in rat somatosensory cortex: Type 1 and type 2 dynamics. *J. Neurophysiol.* **92**, 2283–2294 (2004).
56. Murphy, B. K. & Miller, K. D. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron* **61**, 635–648 (2009).
57. Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J. & McNaughton, B. L. Paradoxical effects of external modulation of inhibitory interneurons. *J. Neurosci.* **17**, 4382–4388 (1997).
58. Van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science (1979)* **274**, 1724–1726 (1996).
59. Long, M. A., Jin, D. Z. & Fee, M. S. Support for a synaptic chain model of neuronal sequence generation. *Nature* **468**, 394–399 (2010).
60. Itskov, V., Curto, C., Pastalkova, E. & Buzsáki, G. Cell assembly sequences arising from spike threshold adaptation keep track of time in the hippocampus. *J. Neurosci.* **31**, 2828–2834 (2011).
61. Rajan, K., Harvey, C. D. D. & Tank, D. W. W. Recurrent network models of sequence generation and memory. *Neuron* **90**, 128–142 (2016).
62. Zhou, S., Masmanidis, S. C. & Buonomano, D. V. Neural sequences as an optimal dynamical regime for the readout of time. *Neuron* **108**, 651–658e5 (2020).
63. Zemlianova, K., Bose, A. & Rinzel, J. A biophysical counting mechanism for keeping time. *Biol. Cybern*. **116**, 205–218 (2022).
64. Fiete, I. R., Senn, W., Wang, C. Z. H. & Hahnloser, R. H. R. Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron* **65**, 563–576 (2010).
65. Langdon, C. & Engel, T. A. Latent circuit inference from heterogeneous neural responses during cognitive tasks. bioRxiv 2022.01.23.477431 (2022).
66. Mastrogiuseppe, F. & Ostojic, S. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron* **99**, 609–623e29 (2018).
67. Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F. & Ostojic, S. The role of population structure in computations through neural dynamics. *Nat. Neurosci.* **25**, 783–794 (2022).
68. Langdon, C., Genkin, M. & Engel, T. A. A unifying perspective on neural manifolds and circuits for cognition. *Nat. Rev. Neurosci.* **24**, 363–377 (2023).
69. Schaeffer, R., Khona, M., Meshulam, L. & Fiete, I. R. Reverse-engineering recurrent neural network solutions to a hierarchical inference task for mice. *Adv. Neural Inf. Process. Syst.*, 4584–4596 (2020).
70. Barak, O., Sussillo, D., Romo, R., Tsodyks, M. & Abbott, L. F. From fixed points to chaos: Three models of delayed discrimination. *Prog Neurobiol.* **103**, 214–222 (2013).
71. Maheswaranathan, N., Williams, A. H., Golub, M. D., Ganguli, S. & Sussillo, D. Reverse engineering recurrent networks for sentiment classification reveals line attractor dynamics. *Adv. Neural Inf. Process. Syst.* **32**, 15696 (2019).
72. Sussillo, D. & Barak, O. Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Comput.* **25**, 626–649 (2013).
73. Cowan, J. D., Neuman, J. & van Drongelen, W. Wilson–Cowan equations for neocortical dynamics. *J. Math. Neurosci.* **6**, 1–24 (2016).
74. Borisyuk, R. M. & Kirillov, A. B. Bifurcation analysis of a neural network model. *Biol. Cybern*. **66**, 319–325 (1992).
75. Ermentrout, G. B. & Cowan, J. D. Temporal oscillations in neuronal nets. *J. Math. Biol.* **7**, 265–280 (1979).
76. Hoppensteadt, F. C. & Izhikevich, E. M. *Weakly Connected Neural Networks* **126** (1997).
77. Ermentrout, G. B. & Terman, D. H. *Mathematical Foundations of Neuroscience. Interdisciplinary Applied Mathematics*, vol. 35 (Springer, 2010).
78. Orhan, A. E. & Ma, W. J. A diverse range of factors affect the nature of neural representations underlying short-term memory. *Nat. Neurosci.* **22**, 275–283 (2019).
79. Song, H. F., Yang, G. R. & Wang, X. J. Training excitatory-inhibitory recurrent neural networks for cognitive tasks: a simple and flexible framework. *PLoS Comput. Biol.* **12**, e1004792 (2016).
80. Ermentrout, B. Simulating Analyzing, and Animating Dynamical Systems. A guide to XPPAUT for researchers and students (SIAM, Society for Industrial and Applied Mathematics, 2002). https://doi.org/10.1137/1.9780898718195

Acknowledgements

K.Z: This work is supported by the Google Ph.D. Research Fellowship. A.B.: This material is based upon work supported by the National Science Foundation under Grant No. DMS-1929284 while the author was in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI, during the Math + Neuroscience: Strengthening the Interplay Between Theory and Mathematics program.

Author contributions

K.Z. proposed the project and carried out all RNN simulations and analysis. K.Z., J.R., A.B. formulated and simulated the reduced model and interpreted the results. K.Z. drafted the manuscript and edited it with J.R. and A.B.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-77849-x.

Correspondence and requests for materials should be addressed to J.R.