Title: Transformation of temporal sequences in the zebra finch auditory system

Running Title: Zebra Finch auditory transformations

Authors: Yoonseob Lim[1,5], Ryan Lagoy[2], Barbara Shinn-Cunningham[3], Timothy J. Gardner[4]

Affiliations:

[1]Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215, USA

[2]Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215, USA

[3]Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

[4]Department of Biology and Biomedical Engineering, Boston University, Boston, MA 02215, USA

[5]Convergence Research Center for Diagnosis, Treatment, and Care System for Dementia, Korea Institute of Science and Technology, Seoul 02792, Korea

Contact: Dr. Timothy J. Gardner, timothyg@bu.edu

Abstract:

This study examines how temporally patterned stimuli are transformed as they propagate from primary to secondary zones in the thalamorecipient auditory pallium in zebra finches. Using a new class of synthetic click stimuli, we find a robust mapping from temporal sequences in the primary zone to distinct population vectors in secondary auditory areas. We tested whether songbirds could discriminate synthetic click

sequences in an operant setup and found that a robust behavioral discrimination is present for click sequences composed of intervals ranging from 11-40ms, but breaks down for stimuli composed of longer inter-click intervals. This work suggests that the analog of the songbird auditory cortex transforms temporal patterns to sequence-selective population responses or “spatial codes,” and that these distinct population responses contribute to behavioral discrimination of temporally complex sounds.

Introduction

A highly developed auditory cortical network supports auditory-vocal behavior in songbirds. The core of the auditory processing system consists of anatomical areas named Field L, NCM (Caudomedial Nidopallium), and CM (Caudomedial Mesopalium) (Vates et al. 1996) (Figure 1c). These areas and other associated auditory areas are directly and indirectly connected with the song motor pathway (Vates et al. 1996; Mandelblat-Cerf et al. 2014). Field L, the primary thalamorecipient area, is composed of four different sub-regions (L2a, L2b, L1, and L3) that are reciprocally connected (Vates et al. 1996). Among these sub-regions, L2a receives the strongest input from the core of Ov (Nucleus Ovoidalis), the primary auditory thalamus (Müller & Leppelsack 1985; Rübsamen & Dörrscheidt 1986; Hose et al. 1987). Secondary auditory areas, L2b, L3, and L1 receive feedforward input from L2a and thalamus, but also receive feedback from higher cortical areas such as CM. These hierarchically and reciprocally connected auditory areas are thought to be analogous to the early stages of mammalian auditory cortex, but the details of the homologies remain a subject of debate (Jarvis et al. 2005; Wang et al. 2010; Calabrese & Woolley 2015).

For zebra finches and other songbirds, temporal cues in song provide reliable information about the identity of the singer and are used for perceptual discrimination of songs (Gentner & Margoliash 2003; Gentner et al. 2006; Grace et al. 2003; Shaevitz & Theunissen 2007). The songbird auditory processing stream is well adapted to this information processing task and reliably relays temporal information in conspecific song. In the zebra finch auditory system, there are neurons from midbrain to the highest levels of auditory association areas that respond with precise spike times to playback of conspecific song. This is true for both dense spiking neurons and the highly selective, sparse firing neurons recently described in the high level auditory area, NCM (Schneider & Woolley 2013), as well as the auditory-motor association area, HVC (High Vocal Center) (Prather et al. 2008). Using a Spectrotemporal Receptive Field (STRF) analysis, the effective temporal integration window of neurons in L2a, the first thalamorecipient zone, was observed to be very brief compared with responses one step further from the periphery in areas L1 and L3 (Kim & Doupe 2011). The secondary areas, including L2b, L1, and L3 but not the primary zone, L2a, are recipients of significant feedback from high order auditory areas (Vates et al. 1996). Combined, these few studies suggest that an interesting transformation of temporal sequences could take place between primary and secondary zones in Field L.

Here we developed new experimental paradigms to examine how temporally patterned auditory stimuli are transformed in the transition from the primary thalamorecipient zone, L2a, to secondary auditory processing areas, L2b and L3. We first demonstrate that from primary area, L2a, to secondary areas, L2b and L3, neurons

responding to song become less synchronous in their relative response times yet more informative about the identity of specific syllables.

To zero in more closely on the nature of the transformation, we examined responses to a set of simplified auditory stimuli consisting of click sequences. The chosen stimuli were akin to “Morse code” - the sounds differed only in the temporal ordering of intervals between clicks. These intervals were drawn from a distribution similar to the intervals between sub-syllabic acoustic transitions in zebra finch song (Gardner et al. 2001; Amador et al. 2013). For click sequence listening, a distinctive transformation of auditory responses was found between primary and secondary auditory zones. In the primary zone, each click elicited a similar low latency response in all recorded neurons, and the structure of this response was largely insensitive to the temporal context of the click. One synapse further from the periphery, in secondary auditory areas, L2b and L3, neurons responded asynchronously and selectively, depending on the temporal context of the click. In effect, temporal sequences are transformed to distinct population vectors in the transition from primary to secondary auditory areas. In this process, temporal patterns come to be represented in a format that could directly form the basis of perceptual discriminations based on simple thresholds.

We next tested whether songbirds could discriminate different temporal click sequence patterns in an operant training paradigm. A novel “restart-go” operant paradigm was developed for this purpose, a paradigm that we found effective for particularly challenging discrimination tests in zebra finches. Using this training procedure, zebra finches rapidly learned to discriminate click sequences composed of

song-like intervals. When the stimulus set was slowed by a factor of two, the strength of the temporal to spatial transformation in the secondary auditory was reduced, and there was a corresponding degradation of behavioral discrimination.

Taken together, these results indicate that the ascending auditory pathway in zebra finches transforms temporal sequences into distinct population vectors. This transformation applied to click sequences consisting of intervals that overlap with sub-syllabic acoustic structure in song, and may provide an important substrate for song perception and discrimination in sub-syllabic time-scales.

Results

General note: the electrophysiological recordings reported here were gathered using four-shank, 32 channel silicon electrodes. From each bird, we recorded activity simultaneously from the primary thalamorecipient zone in auditory area, Field L2a, and neighboring auditory areas in L2b and L3 (Figure 1a,1b). All stimuli were presented in an interleaved fashion, and each animal was recorded acutely, with all data gathered in a single session. All data presented in figures and quantified below were gathered from well-sorted single-unit responses – a minority of recordings (Figure 3 – figure supplement 3, The only exception to this rule is Figure 3 – figure supplement 2, which includes a few channels of high SNR multi-unit traces that did not satisfy our criterion for single unit isolation. These traces are marked with an asterisk). For additional details, see Methods.

Transformation of song responses in the auditory hierarchy

We first compared the temporal coding of song in primary (L2a) and secondary auditory areas (L2b and L3) of unanesthetized songbirds. Our intent was not to thoroughly catalog song responses, but rather to calibrate responses in order to design a set of synthetic stimuli that could be used for the remainder of the study. Primary and secondary recording sites were distinguished based on the distinct response profiles found in the two areas (Figure 1a,3a). This classification was confirmed by spatial mapping of the recording sites (Figure 1 - figure supplement 1) showing that the primary cells were spatially segregated from the secondary neurons (Due to small anatomical and surgical variations and the small scale of the primary zone, this area could not be reliably identified by spatial coordinates alone).

Precise spike timing could be found in both primary and secondary areas in response to song. Focusing first on responses in the primary auditory area, L2a, we found a surprising degree of response synchrony across neurons and across birds (Figure 1a). The population peri-stimulus temporal histogram (PSTH) for each song was deeply modulated for neurons in L2a (Figure 1 – figure supplement 3, Figure 2a reveals the histogram of inter-peak intervals in this population PSTH). In contrast, neurons in secondary auditory areas, L2b and L3, showed a broader repertoire of response profiles. This increase in the diversity of response timing leads to a decrease in the magnitude of the cross-correlation between the PSTHs of individual neurons in the secondary auditory areas relative to a similar cross-correlation performed in primary area, L2a (Figure 1d).

Transformation of click-sequence responses in the auditory hierarchy

Our next objective was to examine whether a similar transformation from synchronous to asynchronous coding could be seen for more elementary stimuli consisting of irregularly spaced clicks. This synthetic stimulus would allow us to probe whether the sequence transformation from the primary to secondary auditory areas requires complex spectral content. If secondary auditory neurons have more complex or more selective spectral receptive fields, the emergence of asynchronous coding in the secondary auditory areas could be explained on the basis of this acoustic selectivity alone. However, if the transformation from synchronous primary response to asynchronous secondary responses could be reproduced with click trains, the result would indicate that the auditory processing pathways contain intrinsic temporal dynamics that transform temporal sequences independent of spectral selectivity.

The chosen synthetic stimuli were three seconds long and composed of clicks separated by ten specific inter-click intervals (11,14,16,20,23,26,29,34,36,40ms). We chose these intervals based on the timescale of neural responses to birdsongs in L2a (Figure 2a,2b). The inter-peak intervals of the population PSTH in response to these click sequences was similar to inter-peak intervals in response to natural song. In effect, we chose click patterns that, in primary auditory area, elicited a temporal response loosely overlapping with the natural song response. We note that the selected inter-click intervals are also similar to intervals between sub-syllabic acoustic transitions found in zebra finch song (Amador et al. 2013; Norton & Scharff 2016). For comparison, Figure 2 also shows the L2a PSTH inter-peak interval histogram for click sequences slowed by a factor of two.

160 The duration of all ten click intervals summed together is 249ms. The longer
161 three-second sequences were built from 249ms blocks, where each block contains a
162 permutation of the ten click intervals. In some stimuli the blocks were repeating and in
163 others non-repeating. For all sequences the stimuli differed only in the ordering of click
164 intervals. On timescales longer than the block duration the statistical properties of all
165 stimuli were equivalent. The set of stimuli used in this study can be seen in Figure 3 -
166 figure supplement 1 (Sample audio files are also provided. See supplementary file 1).
167 Raster plots for single units in primary and secondary auditory areas are shown
168 in Figure 3a (Example spike waveforms of single units and corresponding rasters are
169 shown in Figure 3 – figure supplement 3). Raster plots for the full ensemble of single
170 and multi-units are shown in Figure 3 – figure supplement 2, including a breakdown of
171 secondary cells into narrow (red) and broad-spiking (black) neuron waveforms. Only
172 narrow units were found in the primary auditory area (This figure is the only time in the
173 paper that poorly sorted units, or “multi-units” are included). A distinct change in the
174 temporal response to click sequences can be found in the transition from primary to
175 secondary areas. In the primary auditory area, the click responses are fairly insensitive
176 to the local context – to first approximation, each click evoked a synchronous, low
177 latency response across channels, whereas secondary auditory areas were
178 characterized by sparser and less synchronous responses that were more sensitive to
179 the sequence context of the click (Figure 3 – figure supplement 4,5). The click
180 sequence, by definition, contains no significant spectral cues for frequencies above
181 100Hz (the shortest interval in the click set was 11ms, thus below the 100Hz cutoff).
182 Zebra finch hearing thresholds for pure tones are attenuated by about 20dB relative to

183 humans at 100Hz (Okanoya & Dooling 1987; Moore 2007), and the fundamental
184 frequency of conspecific song is typically 500Hz or higher in zebra finches.

185 As for song responses, the transition from primary to secondary thalamorecipient
186 areas reveals a desynchronizing transformation that maps temporal click sequences
187 onto distinct neuronal ensembles. For the click sequences used here, this
188 transformation is even more apparent than for song responses. The diversification of
189 neuronal responses increases the information about the preceding temporal context of a
190 given click that the population vector contains. To demonstrate this, we computed
191 phase space trajectories of the population vector in response to click sequences, and
192 then quantified the Euclidean distance between these phase space trajectories. In this
193 analysis, every neuron recorded defines a direction in a phase space hypercube, and
194 the average firing rate of the cell defines a position along the respective axis.

195 The phase space trajectory for three cells in the primary auditory area and three
196 cells in the secondary auditory areas are shown in Figure 4a during playback of two
197 distinct sequences. In the primary auditory area, L2a, the phase space trajectories of
198 distinct stimuli overlap for all time points, meaning that the pattern of active cells
199 contains little population-vector information that can distinguish the stimuli. In contrast,
200 in secondary auditory areas, specific points in the phase space trajectory diverge from
201 one another in a stimulus-dependent manner. That is, the pattern of cell responses in
202 secondary auditory areas contain information about one or more intervals preceding the
203 click. To summarize simply - there are particular configurations of active cells that occur
204 only during playback of one stimulus or another — a useful feature for a system that is
205 tuned to make fine discriminations about temporal sequence patterns.

206 To quantify the degree to which the click stimuli can be distinguished based on
207 the neural responses, we defined a simple decoding mechanism based on the
208 population vector of the ensemble response (see Methods for details). In this decoding,
209 the discriminability of the sequence at a particular time is given by the distance in phase
210 space to the nearest trajectory belonging to a different stimulus. The power of this
211 “spatial” code for sequence discrimination is quantified through an ROC (Receiver
212 Operating Characteristic) analysis in Figure 4b. We analyzed coding in primary and
213 secondary areas using the ROC analysis, based on a fixed number of single unit
214 recordings (n=10) in both cases. In the secondary auditory areas, but not the primary
215 thalamorecipient area, temporal sequences are mapped onto distinct population
216 patterns, revealing a better sequence decoding in the ROC analysis. We repeated this
217 analysis just for the first 500ms of the stimulus, and still found a high degree of
218 sequence discriminability in the secondary auditory areas (Figure 4 – figure supplement
219 1). This shorter analysis is more directly relevant to the behavioral discriminations
220 reported below, since trained birds performing behavioral discriminations typically
221 respond within this time frame (Figure 6 – figure supplement 2). To further validate
222 this approach, we applied the same analysis to the PSTH of the song syllable
223 responses (n=13 syllables, Figure 1a) and found an increase in syllable discriminability
224 in the secondary auditory area (Figure 1 - figure supplement 2). Given the rich spectral
225 content of song relative to clicks, the primary area, L2a, already shows a high degree of
226 response selectivity, better than in response to the click sequences.
227 We next repeated the click electrophysiology using a stimulus set composed of
228 intervals twice as long as the first (22-80ms, rather than 11-40ms, Figure 2c). This

change in stimulus timescale had a minimal impact on spike rate in secondary auditory cortex (Figure 5 – figure supplement 1), but a significant change in the power of the temporal to spatial transformation. Using the same phase plane ROC analysis, we found that the timescale dilation led to reduced sequence discrimination in secondary auditory areas (Figure 5).

Behavioral recognition of click sequences

The preceding electrophysiology experiments demonstrated a transformation of click responses to distinct population vectors in secondary auditory areas of unanesthetized zebra finches. As a result, areas downstream of secondary auditory areas could, in principle, solve a click-sequence classification problem based on a simple summation and threshold applied to subsets of secondary cell inputs. Given the robust transformation of temporal click sequences in zebra finch auditory areas, we next sought to determine if songbirds could be trained to behaviorally discriminate this class of artificial stimuli, and whether or not properties of the electrophysiological responses correlated with behavioral discriminations.

Songbirds were trained using a new operant training procedure developed for this study (Figure 6 - figure supplement 1). We call this automated training “reset-go.” A detailed description of the training procedure can be found in Methods. In essence, a bird can demonstrate learning through two behaviors - by interrupting the playback of a non-rewarding stimulus to “request” the reset of an unfavorable trial, or by accessing the water port during playback of rewarding stimuli. In all experiments, two sounds were presented - a rewarded stimulus (click sequence 2 from Figure 3 - figure supplement 1)

and non-rewarded stimulus (click sequence 1 or 9 from Figure 3 - figure supplement 1). Zebra finches in this task were mildly water restricted, and worked for 1-5µl drops of water, routinely performing a thousand trials in a five hour training session.

Figure 6a reveals the time-course of discrimination learning for one bird. After ten days of initiation of training this bird would interrupt the playback of the unrewarded stimulus (sequence 1) within three seconds and access the water port while the rewarded stimulus (sequence 2) was presented. Figure 6b shows summary statistics for learning in eight birds trained to discriminate sequence 1 vs. sequence 2. Figure 6 – source data 1 documents the groups of birds trained and Figure 6 - figure supplement 2 shows the time-course of learning for the various groups. The detailed training procedure is described in Methods. Over a population of trained birds (n=53), a majority (n=35 birds) showed high levels of performance (d' > 1) within 14 days of training onset, revealing that songbirds could readily learn to discriminate the fast temporal click sequences used in this study.

Catch trials probe the nature of auditory discrimination

To probe the underlying nature of the auditory discrimination, we examined catch trials for two conditions. For time-reversed click stimuli, behavior fell to chance levels (Figure 7a), indicating that the ordering of the click intervals was critical to the behavioral discrimination. The next test examined cyclic permutations of the training stimuli. Rather than beginning playback at the normal starting interval of each sequence, the cyclic permutation initiated each stimulus at a random click interval in the three second stimulus – effectively a phase shift in the stimulus. For this group of catch

trials, a small decrease in performance was found when the cyclic stimulus was first introduced, but within four days, performance returned to baseline (Figure 7b). The conclusion from this is that the birds are listening for patterned sequences of intervals irrespective of their absolute time of occurrence relative to the onset of the trial.

Breakdown of behavioral recognition

Since the click sequence contains no spectral structure above 100Hz, stretching the click sequence is a manipulation that has no impact on the frequency content of the sound in the spectral range of zebra finch syllables (>500Hz). We found that birds trained to discriminate fast sequences failed to respond above chance when the timescale of the clicks was slowed by a factor of two. The slow sequences were truncated at three seconds to match the original stimulus duration.

We next examined whether naive birds could learn to discriminate the slower click sequences if they were exclusively trained on the slower sequences from the outset. Eleven birds were trained in a single stage training and four birds trained on the first stage of a two-stage training procedure that is also documented in the methods section. In contrast to the high success rate for faster click sequences, no bird developed a discrimination ability for the slower click sequences (Figure 6c, Figure 6 - figure supplement 2g,2h). The ability to discriminate click stimuli was found only for the faster click sequences.

Discussion

297 Songbirds form detailed auditory memories for complex songs, and these
298 memories serve to guide imitative vocal learning (Nottebohm 1972; Brainard & Doupe
299 2002; Bolhuis & Gahr 2006; Gardner et al. 2005). In parallel, a range of songbird
300 species can perform at high levels in operant tasks involving song and synthetic
301 stimulus discrimination (Austen et al. 2011; Sturdy & Weisman 2006; Cynx &
302 Nottebohm 1992; Scharff et al. 1998; Stripling et al. 2003; Abe & Watanabe 2011).
303 While songbird auditory performance has been well documented, the network
304 mechanisms underlying song discriminations have not been studied. In particular, one
305 of the least understood aspects of auditory sequence processing concerns the
306 transformations applied to complex temporal sequences (Mauk & Buonomano 2004).
307 The present study provides insight into the processing of a simple class of
308 temporal sequences composed of irregularly spaced clicks. We find that after the
309 stimulus passes through the primary thalamorecipient zone, L2a, L2b, and L3, these
310 temporal sequences are transformed into distinct population vectors or “spatial codes.”
311 The mapping of temporal patterns to spatial patterns or ensemble codes provides an
312 opportunity for downstream neurons to perform stimulus discrimination based on simple
313 linear classifiers acting on the population vector. For the click stimuli used in this study,
314 reliable discriminations could be made based on the distinct population vectors that
315 arise in L2b and L3, binned in 5ms time bins.
316 Operant training revealed that songbirds readily learn to discriminate the Morse-
317 code like click stimuli. The fast click sequences were behaviorally discriminable with
318 high accuracy in a majority of trained birds. Surprisingly, no animals learned to
319 discriminate click sequences that were slowed by a factor of two, even though

secondary auditory areas responded with similar spike rates to the slower stimulus. The slowed sequences evoked inter-peak intervals in primary auditory area PSTH that were longer than the typical intervals between peaks in the PSTH during natural song exposure. We suggest that the ascending auditory pathway in the transition from L2a to L2b and L3 is tuned to process temporal events on the faster timescale (11-40ms) in a manner that is particularly useful for song memorization and discrimination.

We mention two caveats in the present study. First, the high-pass cutoff frequency of the loudspeakers was 3kHz (High frequency tweeters were used for stimulus delivery limiting the spectral content of each click). We do not know how the spectral content of the click impacts the behavioral discrimination of the slower sequences. In another prior study, zebra finches were able to discriminate sequences of beeps spaced by intervals of up to 300ms – intervals much longer than those used in our study (van der Aa et al. 2015). It is likely that brief clicks and longer tones tap into auditory processing pathways with distinct temporal dynamics, explaining the performance difference (In addition, many details of the temporal discrimination tasks were different in the two studies, and the distinct results may also relate to these task differences). Additional tests will be needed to determine whether or not the spectral content of each click impacts the behavioral performance. Opportunities also exist to further examine the ability of the zebra finch to generalize temporal pattern recognition through time-dilations (Nagel et al. 2010).

The second caveat is that the single unit ensembles studied here were “virtual ensembles” recorded in different animals; noise correlations within animals could further impact discrimination in ways that were not addressed here (Zohary et al. 1994; Abbott

& Dayan 1999). While we did not acquire enough high quality single unit data to perform the ROC analysis within single animals, enough units were recorded simultaneously to qualitatively reveal the transformation from primary to secondary responses in summary raster plots (Figure 3 - figure supplement 2,3). These rasters support the view that the sequence transformation described for virtual ensembles will also hold for ensembles of neurons in individual birds.

Much theoretical interest has focused on the question of how brains composed of neurons with short intrinsic timescales can process long timescale stimuli and generate long timescale behaviors (Lashely 2004). For temporal stimuli composed of identical units such as clicks, intrinsic cellular or circuit mechanisms must bridge intervals of time from one interval to the next in order to create sequence-specific population responses. To encode the history of the stimulus in the present state of the network, synfire chains, avalanches, or more complex transient dynamics in recurrent networks have all been proposed (Abeles 1991; Grossberg 1969; Maass et al. 2002). In other models, persistent currents in single cells bridge intervals of time (Egorov et al. 2002). In each of these models, intrinsic dynamics of cortical cells or circuits are used to transfer information about past events into the network responses at a given time.

One effective way of transferring information about prior events into current responses is through feedback connections. Primary auditory area (L2a) in songbirds reportedly receives no feedback from higher level auditory zones (Vates et al. 1996), and the synchronous, low-latency responses in this region may reflect a feedforward response to thalamic drive. In contrast, all other areas, including the secondary auditory zones examined here (L2b and L3), are more densely interconnected both with each

other and with higher level auditory areas. This anatomical distinction raises the possibility that L2b and L3, but not the primary auditory area, L2a, can sustain reverberant activity that could underlie the temporal sequence transformation observed in L2b and L3. Relevant theoretical constructs for this model include liquid state machine theories (Maass et al. 2002). By way of illustration, Figure 8 reveals the output of a simple reverberant model that recapitulates key features of the observed dynamics. In this case, the model is simply a linear dynamical system driven by click sequences and additive noise, and tick marks represent time points when the vector (v) crosses arbitrary threshold amplitude.

$$dv/dt = \alpha M v - \gamma v - \eta \quad (1)$$

In this example, matrix M is a random anti-symmetric matrix, with all imaginary eigenvalues, and η is a random noise term. By choosing the time-constants α and γ appropriately, the model can produce patterns that resemble spike rasters observed in L2b and L3. Figure 8, generated by this linear system, simply illustrates the point that even the simplest recurrent dynamical systems have the capability of transforming click-sequence information into distinct population vectors when properly tuned. In this example, the anti-symmetric matrix, M, provides a form of "critical tuning" in which multiple oscillatory timescales are equally excitable, providing for richer temporal dynamics than would occur for a typical nonsymmetric random connectivity matrix (Magnasco et al. 2009). While the hypothesis that recurrence explains the auditory sequence transformation is appealing, new experimental studies are needed to examine the role of recurrent dynamics in temporal sequence processing in songbird auditory pallium.

While the reverberant models provide an attractive explanation for the sequence transformation observed in L2b and L3, the behavioral discriminations the birds exhibited here cannot be taken as evidence for the reverberant model, strictly speaking. A purely feed-forward counter-hypothesis is that the neurons in secondary auditory areas could demonstrate biophysical integration timescales that solve the sequence discrimination through single cell properties. To illustrate this hypothesis, we first smoothed the click sequences used for behavior training with three different rectangular windows of timescale T or shorter and built phase-plane traces of these hypothetical "smoothing units" in 3D space. In this case, three different smoothing windows correspond to hypothetical units with different integration timescales. We then analyzed the minimal time-scale T for which the behaviorally trained sequences could be perfectly segregated in the ROC analysis performed earlier for actual neural sequences. From this analysis, we found that phase plane traces used in the behavioral studies can be perfectly separated if the width of the longest rectangular window was 100ms or greater. To state this more simply, while the sequences presented to the birds were three seconds long, click rates measured in time bins as short as 100ms provide a population vector that is adequate for sequence discrimination. This 100ms timescale cannot rule out either feedforward single cell biophysics or recurrent dynamics as contributors to the sequence transformation.

While the circuit mechanisms remain to be established, this study serves to demonstrate both a distinctive transformation of temporal sequences in the transition from L2a to higher order areas, L2b and L3, and a behavioral capacity of zebra finches to discriminate synthetic click sequences. The transformation of temporal sequences to

412 distinct population vectors may underlie the songbird's advanced discrimination abilities
413 for temporally structured conspecific song.

Materials and methods

All procedures were approved by the Institutional Animal Care and Use Committee of Boston University.

In vivo experimental preparation:

Subjects. For the neural recordings, we examined a total of 11 different adult male zebra finches (*Taeniopygia guttata*).

Stimulus. The artificial stimulus set used for electrophysiology consists of nine click sequences with different interval ordering (Figure 3 - figure supplement 1). Sequence 9 was used only for a subset of operant training tests. The natural song stimulus set consisted of three conspecific songs (n=13 syllables). During electrophysiological recordings, a neural data acquisition system (RZ-5, Tucker-Davis Technology) triggered a pulse generator to create rectangular pulses (100μs width) with different intervals, or played the conspecific bird songs. All stimuli were presented in free-field (60~65dB Peak amplitude) over a loudspeaker (bird song) (Companion 2, BOSE Corporation, Framingham, MA, USA) or a tweeter (for clicks) (PLWT4, Pyle Audio, frequency response range: 3kHz – 30kHz).

Neural recording. Prior to the electrophysiological recording, the birds were injected with the anti-inflammatory analgesic Meloxicam, via intramuscular injection, and anesthetized (1-2% isoflurane in 0.6-0.8ℓ/min O_2) for a preparatory surgical procedure to

implant a custom-made head-post. Local scalp anesthetic (bupivicaine) was administered subcutaneously (40µl, 4mg/kg) and a small (0.18g) head plate affixed to the skull through light-bonded dental acrylic. This was attached so that the head could later be held at a fixed 55 degree angle during unanesthetized auditory recordings. After recovery from general anesthetic (two hours), the bird was given a booster dose of bupivicaine along the margins of the scalp, placed in a foam restraint, and transferred to a double-walled sound proof chamber (40A-Series, Industrial Acoustics Company, Bronx, NY, USA), facing a loudspeaker or a tweeter. The sound source was located 25 cm away from the bird beak.

We used a four-shank multichannel silicon probe (Impedance: 1-2 MΩ, A4x8-5mm-50-200-177-A32, Neuronexus, Ann Arbor, MI, USA) to record extracellular spikes. The coordinates for recording were 1.5mm lateral and 0.8mm anterior to the bifurcation point of the mid-sagittal sinus. The probe was advanced slowly at the speed of 1-2µm/sec using a motorized manipulator (MP-285, Sutter Instrument Company, Novato, CA, USA) until the tip of electrode was located 1.0-1.6 mm below the surface of the brain. Recordings lasted for 4-5 hours. At the end of the recording an electrolytic lesion was made at the location of one of the silicon shank tips using a tungsten electrode (10µA for 10sec). Following this, the birds were deeply anesthetized (110µl, Sodium Pentobarbital [250mg/kg]) and perfused. The extracted brains were stored in 4% paraformaldehyde solution for histology. On the next day after perfusion, parasagittal 100µm sections of the brains were prepared (Vibratome® Series 1000, Technical Products International, St. Louis, MO, USA) and stained with cresyl violet. Electrode

placement was verified by comparing electrolytic lesions to histological landmarks that define the boundaries of Field L (Fortune & Margoliash 1992).

Spike sorting. To isolate single units, the extracellular voltage traces were high-pass filtered at 500Hz (3rd order Butterworth filter) and putative spikes were detected if the voltage traces crossed the positive and negative-going threshold (Quiroga et al. 2004). Then, spikes were re-aligned to the negative peak after resampling up to 250kHz using cubic spline interpolation method. Features of the aligned spikes were composed of the first three principal components and wavelet coefficients of spike waveforms (Quiroga et al. 2004). A mixture of Gaussians model was fit to the spike features using an Expectation Maximization (EM) algorithm to build distinctive clusters of spikes with similar spike waveforms (Pham et al. 2005). Unit quality was then assessed by signal-to-noise ratio (SNR) and refractory period violations to select well isolated single units (Ludwig et al. 2009). All analyses for spike sorting were performed using custom software written in MATLAB (The Mathworks Inc. Natick, MA, USA).

Spike pattern classification. We classified spike patterns into primary and secondary responses based on cross-correlations between spike trains and click sequence stimuli. The similarity score was defined as the maximum cross-correlation value of normalized PSTH (bin size: 5ms) with the normalized click stimulus. A unit was classified as primary If the similarity score was above 0.5 and secondary if the score was below 0.5. Physiological classification was validated by histology (Figure 1b and Figure 1- figure supplement 1), which revealed that although exact coordinates differed in different

animals, primary neurons formed a contiguous island within the surrounding zone of secondary-like responses. The continuity and scale of these islands of primary responses were consistent with the known anatomy and location of primary thalamoricipient zone L2a.

Timescales of neural responses. The timescales of ensemble responses to songs and click sequences in the primary auditory area L2a were characterized by the distribution of intervals between neighboring peaks of the smoothed PSTH (5ms bin). To smooth the PSTH, we filtered the PSTH with an FIR band-pass filter (Kaiser window, passband: 5-110 Hz, number of coefficient: 2233, sampling rate: 1kHz, passband ripple is 5% and stopband attenuation is 40dB). The filtered PSTH was then normalized so that the values were distributed between 0 and 1. Local peaks of normalized PSTH are selected based on three conditions: distance between peak and valley > 0.01, peak value > 0.3, and peak height relative to the neighboring valley > 0.05.

Phase space trajectory. After dividing responses into two groups (primary or secondary), we built a population vector array that contained all PSTHs of units for different stimuli (bin size: 5ms). Each vector had n dimensions of data, where n is the number of neurons. To visualize the behavior of multiple neurons (Figure 4a), we applied principal component analysis (PCA) on the population vector arrays using functions from MATLAB's Statistical Toolbox.

Stimulus discriminability. We defined discriminability of neural responses as the minimum Euclidean distance between two different population vector arrays in response to distinct sounds. Before calculating distances, each spike rate trace in a population vector was smoothed with a 30ms rectangular window. Then, we divided the recording session into two groups (odd vs. even numbered trials) and obtained the distribution of distances in population vector space built from either the same stimulus or across different stimuli. To build the ROC curve, we calculated the true and false positive ratio for discriminating two different stimuli while changing the decision boundary position.

Auditory operant training preparation:

Here we describe a method for auditory operant training that is useful for training zebra finches on challenging discriminations with little shaping procedure. The proposed method uses water reward rather than seed (Picardo et al. 2015). Zebra finches are adapted to arid conditions and can survive for months in a laboratory setting without access to water (Cade et al. 1965), yet they remain highly motivated to work for water. The quantity of water provided in each reward can be as low as 1-5µl. With this reward quantity, birds work for hundreds or thousands of trials per day.

Subjects. In the operant task, we trained 53 adult (>90 days post-hatch on the first day of training) male zebra finches (*Taeniopygia guttata*). All birds were housed in the same aviary room and were experimentally naive at the start of training. Once a bird entered

526 the training cage, he remained in the training cage 24 hours a day until the end of
527 training period.
528
529 ***Food and water.*** Dehydrated seed (100-110F^{o} for 12 hours, D-5 Dehydrator, TSM
530 Products) was supplied every two days (seed is dehydrated the day before it is provided
531 in the cage). Soft food (ABBA 97 Ultimate nestling food, ABBASEED) was available
532 once per week. Birds had unlimited access to water on the weekends and every day
533 access to grit. Birds were not exposed to water deprivation conditions prior to training.
534 On a single day of training, birds normally initiated around 800-1300 trials (with a
535 maximum of 4,000 trials in one individual). This corresponded to 300-1000µl of water
536 consumption during training. We provided additional water (0.5-1ml) after the training if
537 the total volume of water consumption for two days was less than 1ml. The birds usually
538 drank 0.5-1ml of water over night when this supplement was provided. In total, through
539 reward and supplement, the experimental birds received an average of 1-1.5ml of water
540 every day, a number that corresponds to 50% of normal water consumption for zebra
541 finch under certain environmental conditions (Cade et al. 1965).
542
543 ***Operant Chamber.*** In this experiment, 12 identical operant training cages were used.
544 The training cages (11 inch wide x 8 inch high) were kept inside sound attenuation
545 chambers (22 inch wide x 14.5 inch high x 16 inch deep, Figure 6 - figure supplement
546 1). All inside surfaces of the chambers were lined with embossed acoustic foam
547 (PROSPEC® Composite, Pinta acoustics inc). Inside each training cage, there were two
548 infrared switches (OPB815WZ, OPTEK Technology): one for trial initiation (called the

trial port) and one for water reward (called the water port). The water reservoir was located 24 inches above the cage floor and the water valve (EW-01540-02, Cole-Parmer) was placed between the reservoir and spout. The water spout was located in the middle of the infrared switch assembly (water port) so that whenever the bird accessed the water spout, he broke the infrared beam automatically. We used two different sizes of incompressible plastic tubes to make water flow slow enough for a proper drop size (1-5µl). An illustration of the tubing is shown in Figure 6 - figure supplement 1. The sound stimulus was presented through the same tweeter used for the electrophysiology study (PLWT4, Pyle Audio). A microprocessor dedicated to each cage (Arduino Mega 2560, Arduino) controlled stimulus presentation, water delivery, and infrared switches. Individual clicks generated by the Arduino microprocessor were 100µs long rectangular pulses. Using this microprocessor, the mean jitter in the click interval was 93µs (data is not shown). Every time the bird tried a new trial, data from the previous trial was transmitted by ethernet to a central data processing computer in the lab and analyzed in real time by a custom made Matlab program (Mathworks, Natick MA, USA). Training ran for five hours per day from Tuesday to Friday each week. The behavior of all birds was monitored through USB webcams in each chamber (Webcam Pro 9000, Logitech).

Auditory operant training Procedure:

In this procedure, a bird can demonstrate learning through two behaviors – by interrupting ongoing playback of a non-rewarding stimulus to reset the trial, or by accessing the water port selectively for rewarding stimuli. We trained birds with two

572 different methods: a two- stage method and a single-stage method. In all experiments,
573 two sounds were presented – a rewarded stimulus (click sequence 2) and a non-
574 rewarded stimulus (click sequence 1 or 9).
575
576 ***Training procedure during stage 1 of two-stage training.*** This training starts with
577 only one infrared switch (for trial initiation, on the left side of the cage, Figure 6 - figure
578 supplement 1). The bird can start a new trial or interrupt playback of the stimulus by
579 breaking the infrared beam any time 200ms after the start of the stimulus playback. The
580 water spout is on the right side of the cage and water reward is passively given at the
581 end of the rewarded stimuli, which constitutes 20% of total trials. In this setup, the bird
582 learns to be “impatient” and interrupt stimuli that are not followed by reward. In a sense,
583 the bird is “foraging” for a low-probability rewarded sound. On each day of training, we
584 monitored the latency of trial initiations to two different sequences. During the first one
585 or two days, birds simply explored the training environment and explored the trial port
586 randomly. Gradually, birds realized the existence of passive water reward and started to
587 reinitiate trials earlier on non-rewarding trials than rewarding trials (right middle panel of
588 Figure 6 - figure supplement 2a, and 2d. Note the bump of red curve around 5-6sec). In
589 1-2 weeks of training, birds could re-initiate trials only for non-rewarded trials, and wait
590 for water reward on the rewarded trials.
591
592 ***Training procedure during stage 2 of two-stage training.*** Once birds showed
593 significant learning in stage 1, another infrared switch was activated on the water
594 delivery port. Water was no longer delivered passively, but only if the water port was

accessed during or just after the playback of the rewarding stimulus. This period during which reward port access leads to water is called the “response time-window.” This window was 7 seconds long from the end of a sequence. If the water port was accessed at any time during non-rewarding trials, or outside of the 7 second response window, a 10 second time-out period ensued, during which the green LED was turned off. Introducing another infrared switch in this stage did not alter the trial reset behavior that was acquired in the first stage of training (Figure 6 – figure supplement 2c,2f).

Training procedure for single-stage training. In single-stage training, the bird starts training with both infrared switch-contingencies active from the beginning. However, to jumpstart the process, water was also delivered passively at the end of the rewarded stimulus if the bird did not access the water port during playback of the rewarded stimulus. Once the bird learned to initiate trials and encounter water at the water port location, the passive water delivery was shut off. Other than this brief passive delivery period this method involved no shaping or staging. Birds learned strategies for the use of both infrared switches through exploration (re-initiating trials within 3sec when the non-rewarded pattern was presented and accessing the water port during playback of the rewarding stimulus, Figure 6a).

Operant task behavior evaluation. We used a d-prime measure to estimate the progress of learning:

$$d' = z(H) - z(F) \quad (2)$$

where H is the proportion of correct responses (hit rate) and F is the proportion of incorrect responses (false alarm rate) (Green & Swets 1966).

Author Contributions

YL, BSC and TJG designed the study. YL and RL collected data. YL analyzed the data. YL and TJG wrote the paper.

Acknowledgment

This work was supported by CELEST, a National Science Foundation Science of Learning Center (NSF OMA-0835976), by the National Institute of Health (NIH R01NS089679) and by the National Research Council of Science and Technology (NIST) grant by the Korea government (MSIP)(No.CRC-15-04-KIST). We thank Frederic Theunissen, Luke Remage-Healey, Kathy Nordeen, Ofer Tchernichovski, Sarah Bottjer, Elizabeth Regan, and Richard Hanhloser for providing zebra finch song samples from their colonies. We would also like to thank Aniruddh Patel for comments that improved the analysis and manuscript.

Competing interests

None of the authors have any conflicts of interest.

References

Abbott LF, Dayan P. 1999. The effect of correlated variability on the accuracy of a population code. *Neural Computation*, 11(1), pp.91–101. doi: 10.1162/089976699300016827.

Abe K, Watanabe D. 2011. Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience*, 14(8), pp.1067–1074. doi: 10.1038/nn.2869.

Abeles M. 1991. *Corticonics*, Cambridge University Press.

Amador A, Perl YS, Mindlin GB, Margoliash D. 2013. Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature*, 495(7439), pp.59–64. doi: 10.1038/nature11967.

Austen G, Schneider DM, Akshat V, Woolley SMN. 2011. Automated auditory recognition training and testing. *Animal Behaviour*, 82(2), pp.285–293. doi: 10.1016/j.anbehav.2011.05.003.

Bolhuis JJ, Gahr M. 2006. Neural mechanisms of birdsong memory. *Nature Reviews Neuroscience*, 7(5), pp.347–357. doi: 10.1038/nrn1904.

Brainard MS, Doupe AJ. 2002. What songbirds teach us about learning. *Nature*, 417(6886), pp.351–358. doi: 10.1038/417351a.

Cade TJ, Tobin CA, Gold A. 1965. Water Economy and Metabolism of Two Estrildine Finches. *Physiological zoology*, 38(1), pp.9–33.

Calabrese A, Woolley SMN. 2015. Coding principles of the canonical cortical microcircuit in the avian brain. *Proceedings of the National Academy of Sciences of the United States of America*, 112(11), pp.3517–3522. doi: 10.1073/pnas.1408545112.

Cynx J, Nottebohm F. 1992. Testosterone facilitates some conspecific song discriminations in castrated zebra finches (Taeniopygia guttata). *Proceedings of the National Academy of Sciences of the United States of America*, 89(4), pp.1376–1378. doi: 10.1073/pnas.89.4.1376.

Egorov AV, Hamam BN, Fransén E, Hasselmo ME, Alonso AA. 2002. Graded persistent activity in entorhinal cortex neurons. *Nature*, 420(6912), pp.173–178. doi: 10.1038/nature01171.

Fortune ES, Margoliash D. 1992. Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (taenopygia guttata). *The Journal of Comparative Neurology*, 325(3), pp.388–404. doi: 10.1002/cne.903250306.

Gardner TJ, Cecchi GA, Magnasco M, Laje R, Mindlin GB. 2001. Simple motor gestures for birdsongs. *Physical Review Letters*, 87(20), p.208101. doi: 10.1103/PhysRevLett.87.208101.

673 Gardner TJ, Naef F, Nottebohm F. 2005. Freedom and rules: the acquisition and
674 reprogramming of a bird's learned song. *Science*, 308(5724), pp.1046–1049. doi:
675 10.1126/science.1108214.

676 Gentner TQ, Margoliash D. 2003. Neuronal populations and single cells representing
677 learned auditory objects. *Nature*, 424(6949), pp.669–674. doi: 10.1038/nature01731.

678 Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. 2006. Recursive syntactic pattern
679 learning by songbirds. *Nature*, 440(7088), pp.1204–1207. doi: 10.1038/nature04675.

680 Grace JA, Amin N, Singh NC, Theunissen FE. 2003. Selectivity for conspecific song in
681 the zebra finch auditory forebrain. *Journal of Neurophysiology*, 89(1), pp.472–487. doi:
682 10.1152/jn.00088.2002.

683 Green DM, Swets JA. 1966. *Signal detection theory and psychophysics*, New York:
684 Wiley.

685 Grossberg S. 1969. On learning of spatiotemporal patterns by networks with ordered
686 sensory and motor components. 1. Excitatory components of the cerebellum. *Studies in*
687 *Applied Mathematics*, 48, pp.105–132. doi: 10.1002/sapm1969482105.

688 Hose B, Langner G, Scheich H. 1987. Topographic representation of periodicities in the
689 forebrain of the mynah bird: one map for pitch and rhythm? *Brain Research*, 422(2),
690 pp.367–373. doi: 10.1016/0006-8993(87)90946-2.

691 Jarvis ED et al., 2005. Avian brains and a new understanding of vertebrate brain
692 evolution. *Nature Reviews Neuroscience*, 6(2), pp.151–159. doi: 10.1038/nrn1606.

693 Kim G, Doupe AJ. 2011. Organized representation of spectrotemporal features in
694 songbird auditory forebrain. *The Journal of Neuroscience*, 31(47), pp.16977–16990. doi:
695 10.1523/JNEUROSCI.2003-11.2011.

696 Lashely K. 2004. The probelm of serial order in behavior. *In First language acquisition:*
697 *the essential readings*, pp.316–334.

698 Ludwig KA, Miriani RM, Langhals NB, Joseph MD, Anderson DJ, Kipke DR. 2009.
699 Using a common average reference to improve cortical neuron recordings from
700 microelectrode arrays. Journal of Neurophysiology, 101(3), pp.1679–1689. doi:
701 10.1152/jn.90989.2008

702 Maass W, Natschläger T, Markram H. 2002. Real-time computing without stable states:
703 A new framework for neural computation based on perturbations. *Neural Computation*,
704 14(11), pp.2531–2560. doi: 10.1162/089976602760407955.

705 Magnasco MO, Piro O, Cecchi GA. 2009. Self-Tuned Critical Anti-Hebbian Networks.
706 *Physical Review Letters*, 102(25), p.258102. doi: 10.1103/PhysRevLett.102.258102.

707 Mandelblat-Cerf Y, Las L, Desisenko N, Fee MS. 2014. A role for descending auditory

cortical projections in songbird vocal learning. *eLife*, 3, p.e02152. doi: 10.7554/eLife.02152.

Mauk M, Buonomano DV. 2004. The neural basis of temporal processing. *Annual Reviews*, 27, pp.307–340. doi: 10.1146/annurev.neuro.27.070203.144247.

Moore BCJ. 2007. *An Introduction to the Psychology of Hearing* 5 ed., Academic Press.

Müller CM, Leppelsack HJ. 1985. Feature extraction and tonotopic organization in the avian auditory forebrain. *Experimental Brain Research*, 59(3), pp.587–599. doi: 10.1007/BF00261351.

Nagel KI, McLendon HM, Doupe AJ. 2010. Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *Journal of Neurophysiology*, 104(3), pp.1426–1437. doi: 10.1152/jn.00028.2010.

Norton P, Scharff C. 2016. Bird Song Metronomics": Isochronous Organization of Zebra Finch Song Rhythm. Frontiers in neuroscience, 10.309. doi: 10.3389/fnins.2016.00309

Nottebohm F. 1972. The Origins of Vocal Learning. *American Naturalist*, 106(947), pp.116–140.

Okanoya K, Dooling R. 1987. Hearing in passerine and psittacine birds: a comparative study of absolute and masked auditory thresholds. *Journal of comparative psychology*, 101(1), pp.7–15. doi: 10.1037/0735-7036.101.1.7.

Pham DT, Dimov SS, Nguyen CD. 2005. Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1), pp.103–119. doi: 10.1243/095440605X8298.

Picardo MA, Katlowitz KA, Okobi DE, Benezra SE, Clary RC, Merel J, Paninski L, Long MA. 2015. Analyzing the population dynamics underlying a complex motor act. Program No. 181.01. 2015 Neuroscience Meeting Planner. Chicago, IL: Society for Neuroscience, Online

Prather JF, Peters S, Nowicki S, Mooney R. 2008. Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature*, 451(7176), pp.305–310. doi: 10.1038/nature06492.

Quiroga RQ, Nadasdy Z, Ben-Shaul Y. 2004. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, 16(8), pp.1661–1687. doi: 10.1162/089976604774201631.

Rübsamen R, Dörrscheidt GJ. 1986. Tonotopic organization of the auditory forebrain in a songbird, the European starling. *Journal of Comparative Physiology A*, 158(5), pp.639–646. doi: 10.1007/BF00603820.

Scharff C, Nottebohm F, Cynx J. 1998. Conspecific and heterospecific song

743 discrimination in male zebra finches with lesions in the anterior forebrain pathway.
744 *Journal of Neurobiology*, 36(1), pp.81–90. doi: 10.1002/(SICI)1097-
745 4695(199807)36:1<81::AID-NEU7>3.0.CO;2-6.

746 Schneider DM, Woolley SMN. 2013. Sparse and background-invariant coding of
747 vocalizations in auditory scenes. *Neuron*, 79(1), pp.141–152. doi:
748 10.1016/j.neuron.2013.04.038

749 Shaevitz SS, Theunissen FE. 2007. Functional connectivity between auditory areas field
750 L and CLM and song system nucleus HVC in anesthetized zebra finches. *Journal of*
751 *Neurophysiology*, 98(5), pp.2747–2764. doi: 10.1152/jn.00294.2007.

752 Stripling R, Milewski L, Kruse A, Clayton DF. 2003. Rapidly learned song-discrimination
753 without behavioral reinforcement in adult male zebra finches (Taeniopygia guttata).
754 *Neurobiology of Learning and Memory*, 79(1), pp.41–50. doi: 10.1016/S1074-
755 7427(02)00005-9.

756 Sturdy CB, Weisman RG. 2006. Rationale and methodology for testing auditory
757 cognition in songbirds. *Behavioural processes*, 72(3), pp.265–272. doi:
758 10.1016/j.beproc.2006.03.007.

759 van der Aa J, Honing H, Cate ten C. 2015. The perception of regularity in an
760 isochronous stimulus in zebra finches (Taeniopygia guttata) and humans. *Behavioural*
761 *processes*, 115, pp.37–45. doi: 10.1016/j.beproc.2015.02.018.

762 Vates GE, Broome BM, Mello CV, Nottebohm F. 1996. Auditory pathways of caudal
763 telencephalon and their relation to the song system of adult male zebra finches. *The*
764 *Journal of Comparative Neurology*, 366(4), pp.613–642. doi: 10.1002/(SICI)1096-
765 9861(19960318)366:4<613::AID-CNE5>3.0.CO;2-7.

766 Wang Y, Brzozowska-Prechtl A, Karten HJ. 2010. Laminar and columnar auditory
767 cortex in avian brain. *Proceedings of the National Academy of Sciences of the United*
768 *States of America*, 107(28), pp.12676–12681. doi: 10.1073/pnas.1006645107

769 Zohary E, Shadlen MN, Newsome WT. 1994. Correlated neuronal discharge rate and its
770 implications for psychophysical performance. *Nature*, 370(6485), pp.140–143. doi:
771 10.1038/370140a0.

Figures

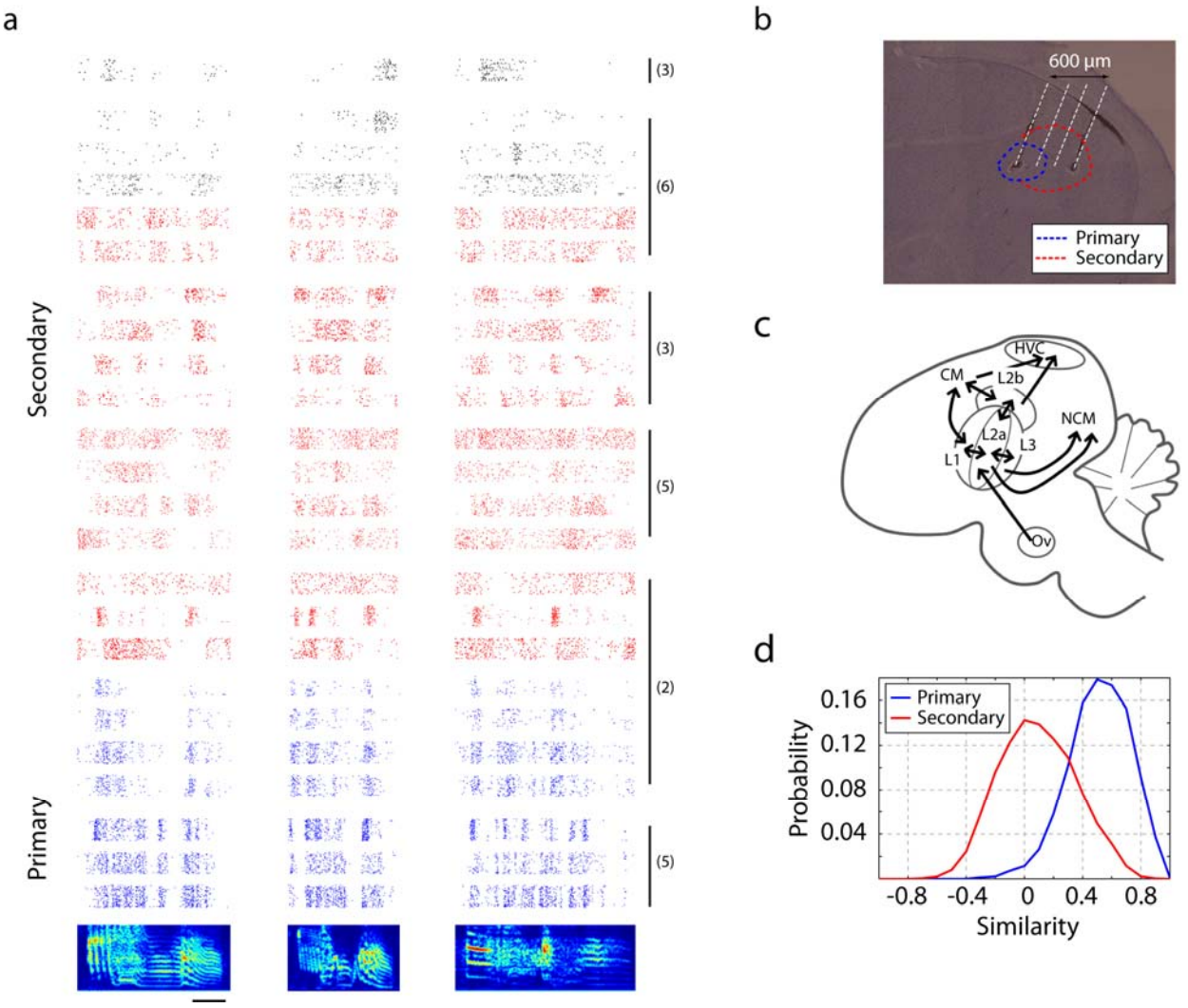


Figure 1. Neural responses in primary and secondary auditory areas to birdsongs.

(a) Example of neural responses in primary (blue) and secondary auditory areas (red and black) to birdsongs. Syllable responses were extracted from playback of whole songs. Individual cells in this figure were recorded in different birds. Numbers on the right correspond to the bird indices shown in Figure 1 – figure supplement 1. Cells in the primary auditory area, L2a, respond more synchronously than cells in the secondary area. Red and black colors in the raster denote two classes of cells in secondary

auditory areas defined by spike-width (For red, spike width is less than 250 μs and for black, greater than 250 μs). The scale bar is 50ms.

(b) Sagittal section located at 1.5mm lateral of the midline with estimated electrode shank positions (dotted white line). Physiological locations are confirmed by the anatomy (Figure 1 - figure supplement 1).

(c) Schematic of a sagittal section of male zebra finch brain.

(d) Response similarity scores between all pairs of cells in the secondary auditory area are lower than similarity scores in the primary auditory area (Secondary auditory responses to song are more diverse across neurons).

The following figure supplements are available for Figure 1:

Figure 1 - figure supplement 1. Estimated recording location of units in primary and secondary auditory areas.

Figure 1 - figure supplement 2. Song syllable discriminability analysis.

Figure 1 - figure supplement 3. PSTH of song responses

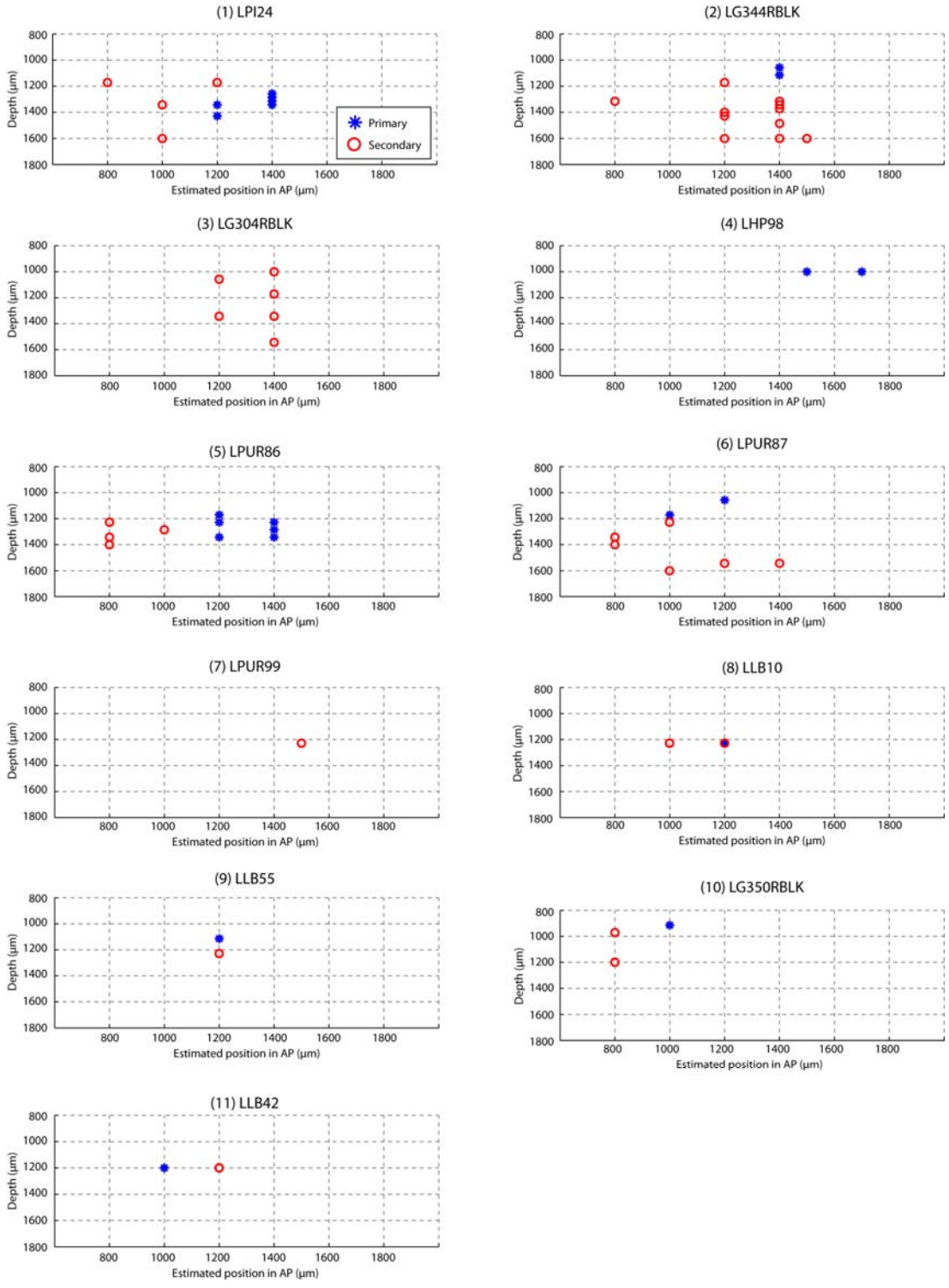
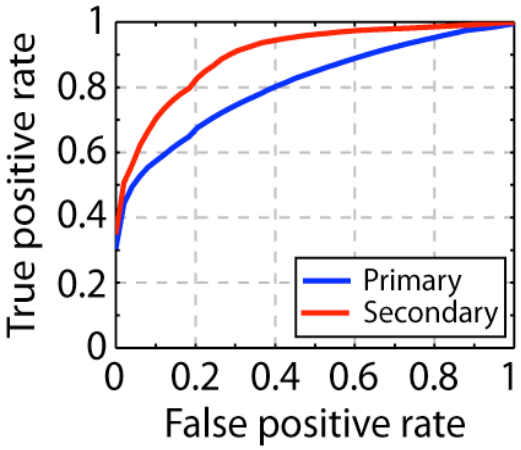


Figure 1 - figure supplement 1. Estimated recording location of units.

Cells were colored by their classification as primary or secondary cells based on response latency and similarity scores (Figure 1d, 3b). This figure shows that for each

801 bird the primary and secondary cells were spatially separable, providing independent
802 confirmation that the classification as primary and secondary cortical neurons was
803 accurate. On each graph, estimated spatial positions of primary (blue star) and
804 secondary (red circles) units are shown. Positions were approximated based on the
805 configuration of electrode and recording coordinates.



806 **Figure 1 - figure supplement 2. Song syllable discriminability analysis.**

807 ROC analysis shows increased discriminability of song syllables in secondary auditory

808 areas, L2b and L3, relative to primary auditory area, L2a (n=13 syllables).

809

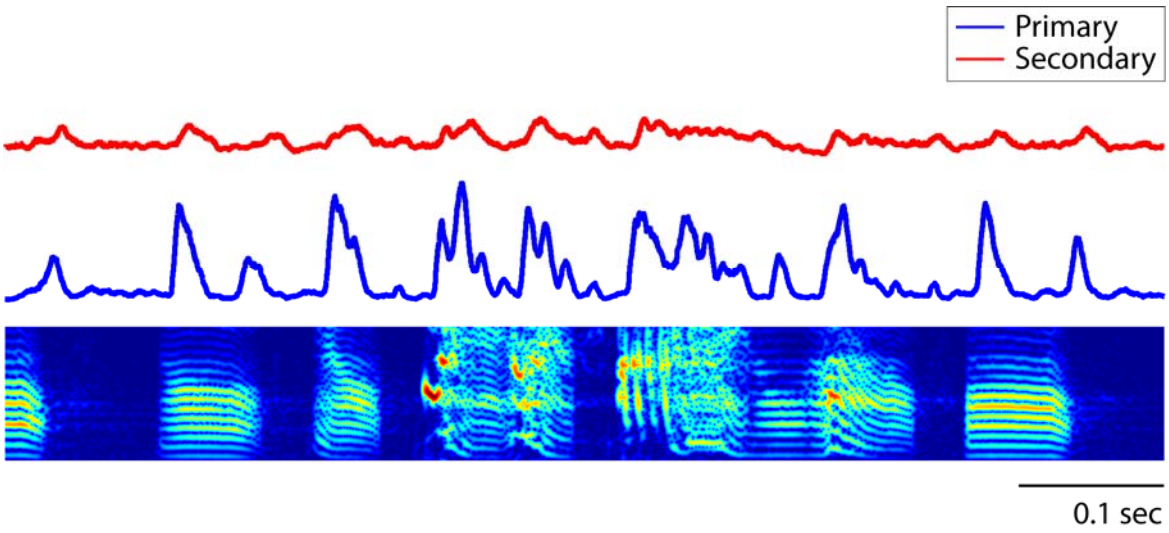


Figure 1 - figure supplement 3. Peri-Stimulus Time Histogram (PSTH) of song responses.

The PSTH of neurons in primary auditory area, L2a, reveal synchronous responses to song (bin size: 5ms). In this figure, the average PSTH of all neurons is shown.

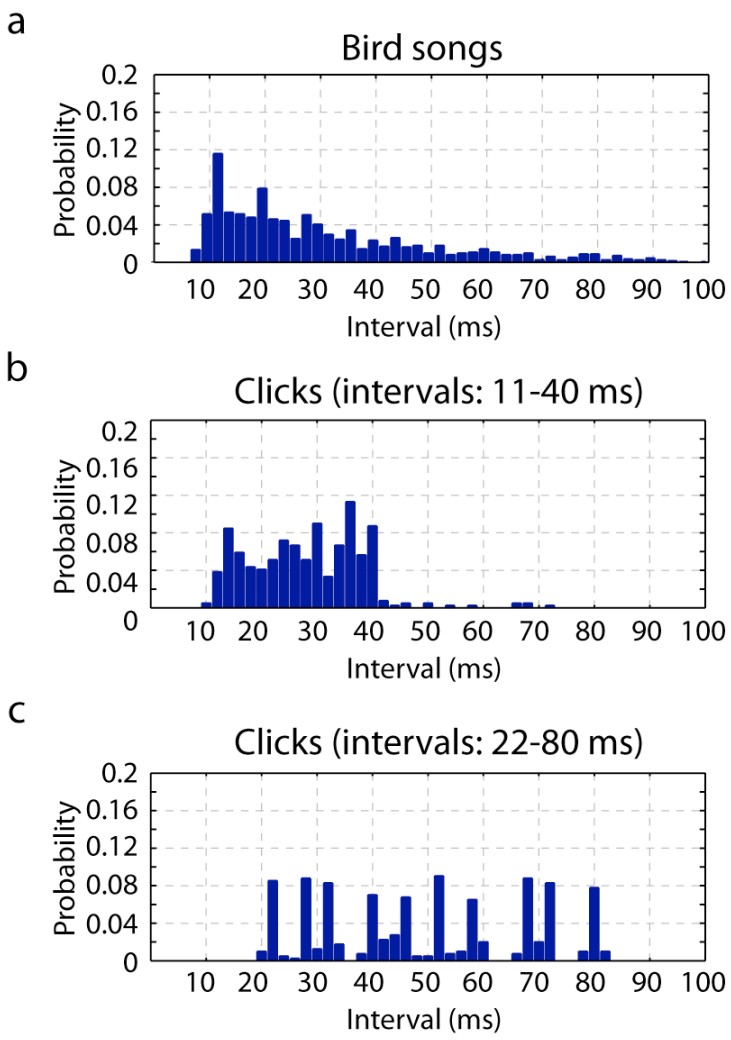


Figure 2. Timescales of neural responses in the primary auditory area, L2a.

(a) Interval histogram of peaks in the PSTH of neurons in the primary auditory area, L2a, in response to bird songs. The population PSTH contains intervals distributed from 10-40ms.

(b,c) Interval histogram of peaks in the PSTH of neurons in the primary auditory area, L2a, in response to click sequences. For the click patterns, we applied two different timescales for the click intervals. In the first timescale, the click sequence evokes PSTH intervals in the range of 10-40ms. The slower set of stimuli evokes PSTH intervals in the range of 20-80ms.

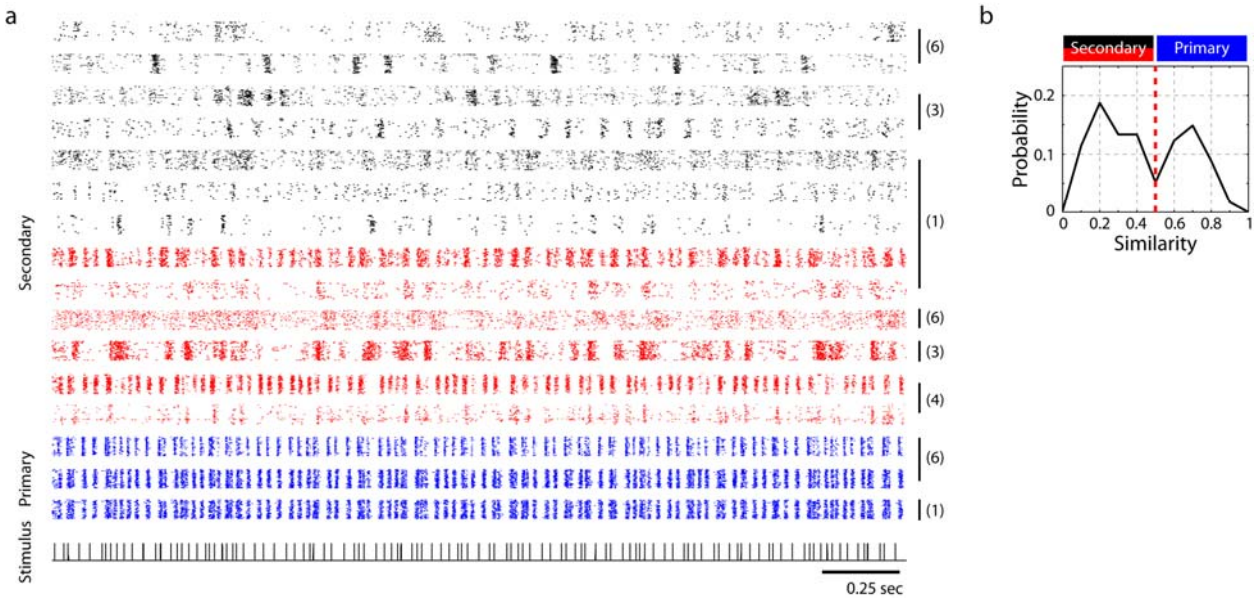


Figure 3. Neural responses to click sequences in primary and secondary auditory areas.

(a) Example of neural responses in primary and secondary auditory areas. Units from individual birds are grouped (black vertical bars and corresponding bird indices are shown on the right of the rasters). Red and black rasters mark two classes of cells in secondary auditory areas defined by spike-width (For red, spike width is less than 250μs and for black, greater than 250μs). Blue rasters are cells in the primary auditory area, L2a.

(b) Histogram of cross correlation scores between the click stimulus and the PSTH response. The discrimination line between two peaks (at 0.5 similarity score) also segregates cells spatially (Figure 1 - figure supplement 1), confirming the classification of neurons as residing in spatially separated areas – either L2a, or L3/L2b.

The following figure supplements are available for Figure 3:

Figure 3 - figure supplement 1. All click sequences used for neural recordings and operant training.

Figure 3 - figure supplement 2. Combined single and multi-unit responses to click sequence 1 and sequence 2

Figure 3 – figure supplement 3. Example of spike waveforms and their corresponding click responses.

Figure 3 – figure supplement 4. PSTH of neurons in response to click sequence.

Figure 3 – figure supplement 5. Latency of neural responses to click sequences in the primary auditory area, L2a.

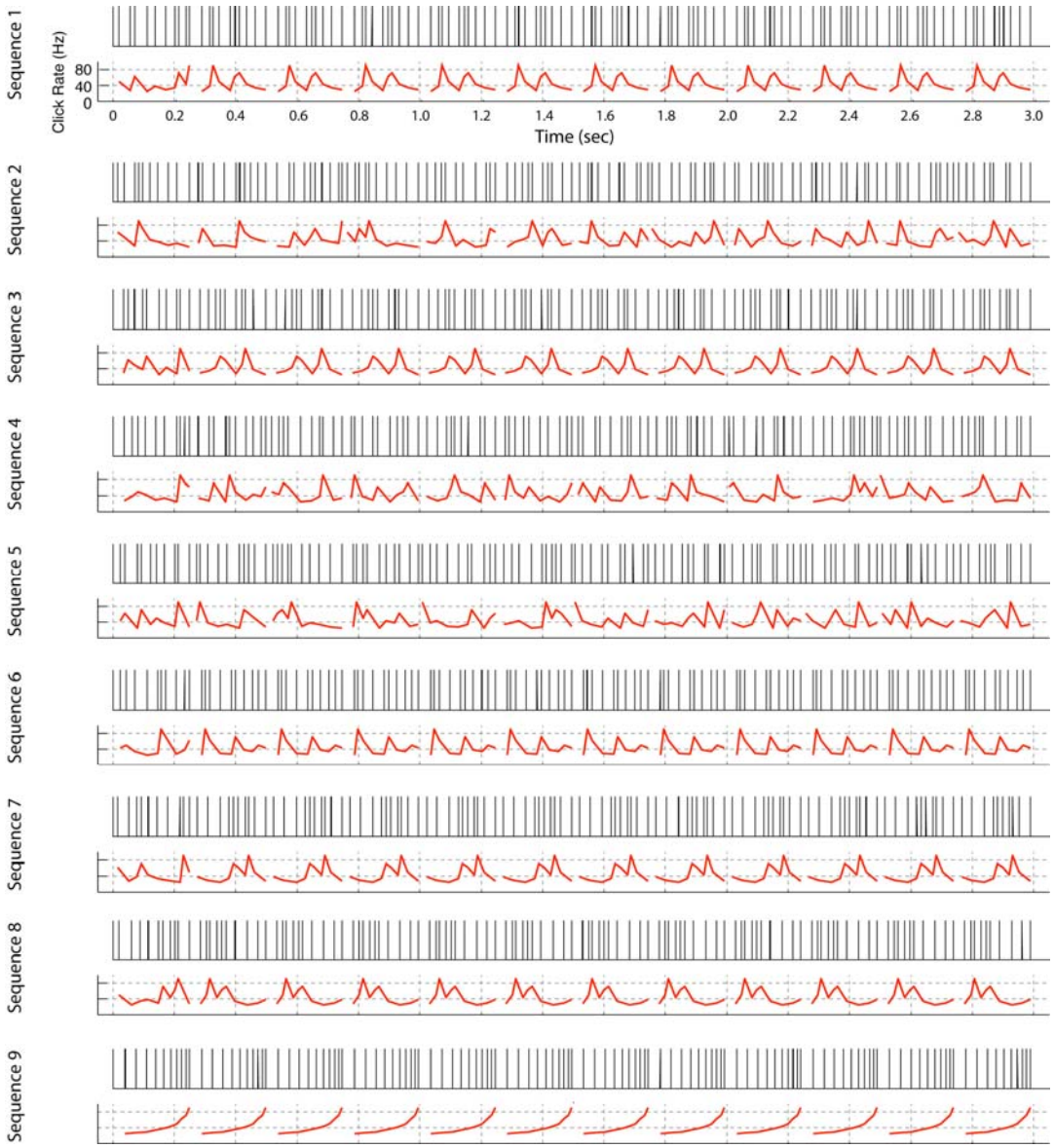
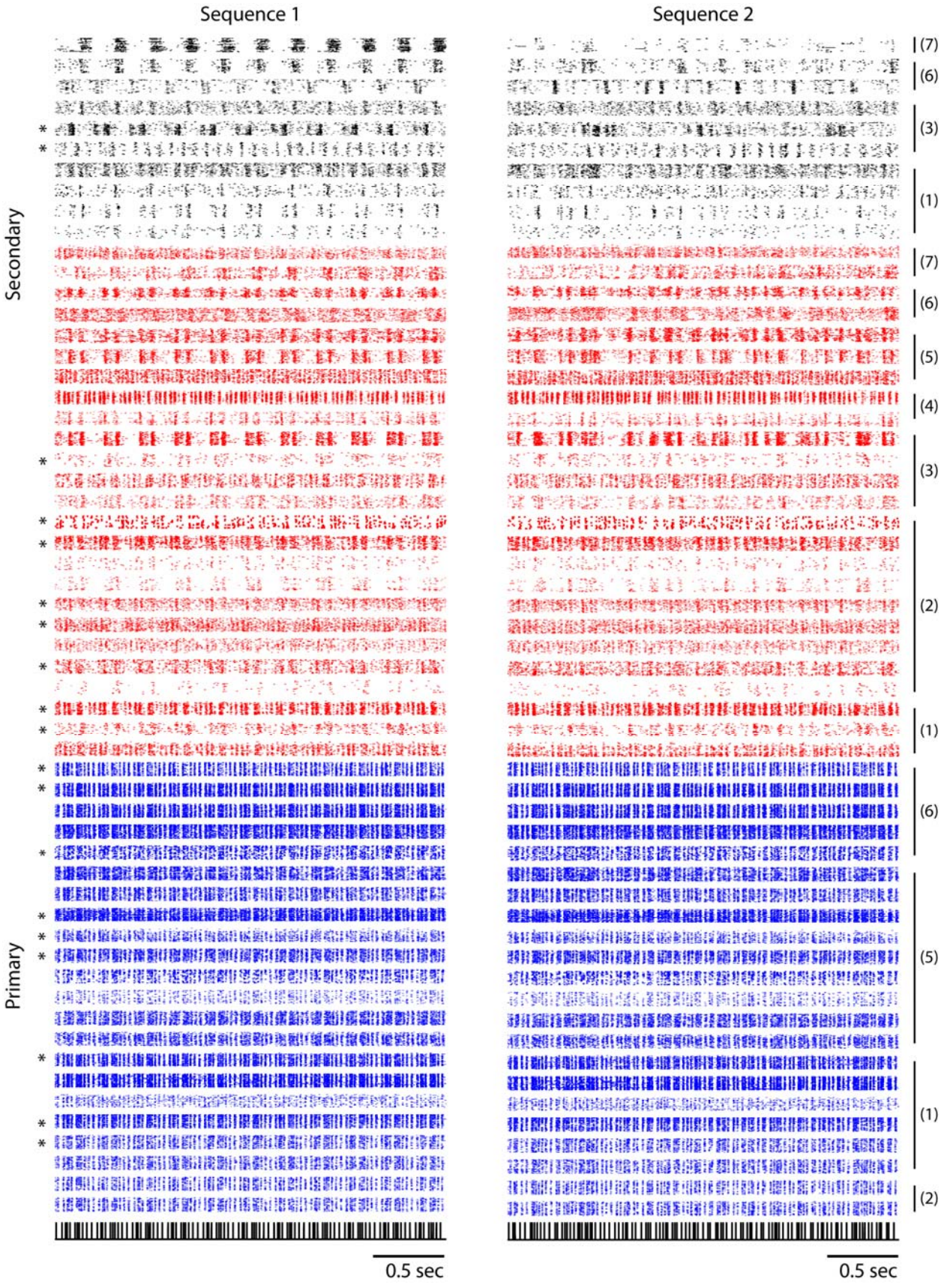


Figure 3 - figure supplement 1. All click sequences used for neural recordings and operant training.

Click sequences were repeating or non-repeating temporal patterns. Each temporal pattern is 249ms long and the total length of the sequence is 3sec. For sequence 1,3,6,7,8, and 9, a single fixed temporal pattern repeats 11 times; the other sequences are composed with 11 different non repeating patterns. We also built some sequences in reverse order (Seq. 1 vs Seq. 3, Seq. 2 vs Seq. 4, Seq. 7 vs Seq. 8). Sequences 1–8

were used for neural recording and sequences 1, 2, and 9 were used for the operant training experiment. An audio file for each click sequence is provided (Supplementary file 1).


Sequence 1
Sequence 2
Secondary
Primary
(7)
(6)
(3)
(1)
(7)
(6)
(5)
(4)
(3)
(2)
(1)
(6)
(5)
(1)
(2)
0.5 sec
0.5 sec

862

Figure 3 - figure supplement 2. Combined single and multi-unit responses to sequence 1 and sequence 2.

Responses in the primary auditory area, L2a, are shown in blue and in secondary areas, L2b/L3 are shown in red and black. Multi-unit responses as opposed to single units are indicated by asterisk marks on the left. Responses from a single bird are grouped by a black vertical bar with the corresponding bird index on the right. Two different classes of neurons in the secondary auditory areas (red and black rasters) are classified based on the peak-to-peak width of spike waveform following the conventions of Figure 3a.

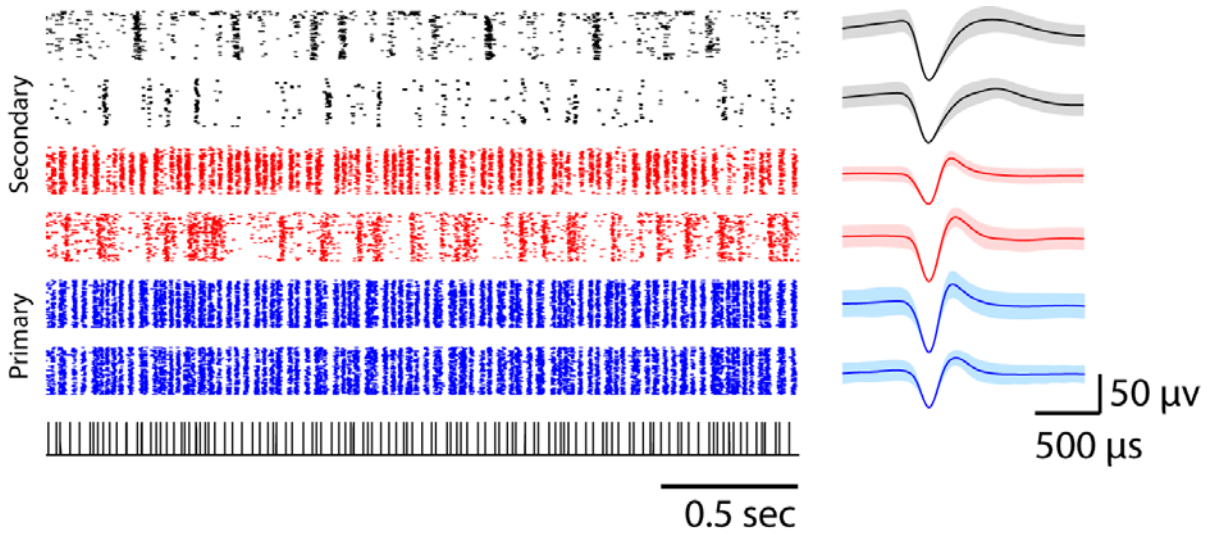


Figure 3 – figure supplement 3. Example spike waveforms corresponding click responses shown in raster form.

Each row of the raster plot represents single unit responses to a click sequence (sequence 2); the corresponding spike waveform is shown on the right. The shaded error bars represent the standard deviation of waveforms. Primary L2a neurons are shown in blue. Narrow and broad spiking units in L2b or L3 are shown in red and black respectively.

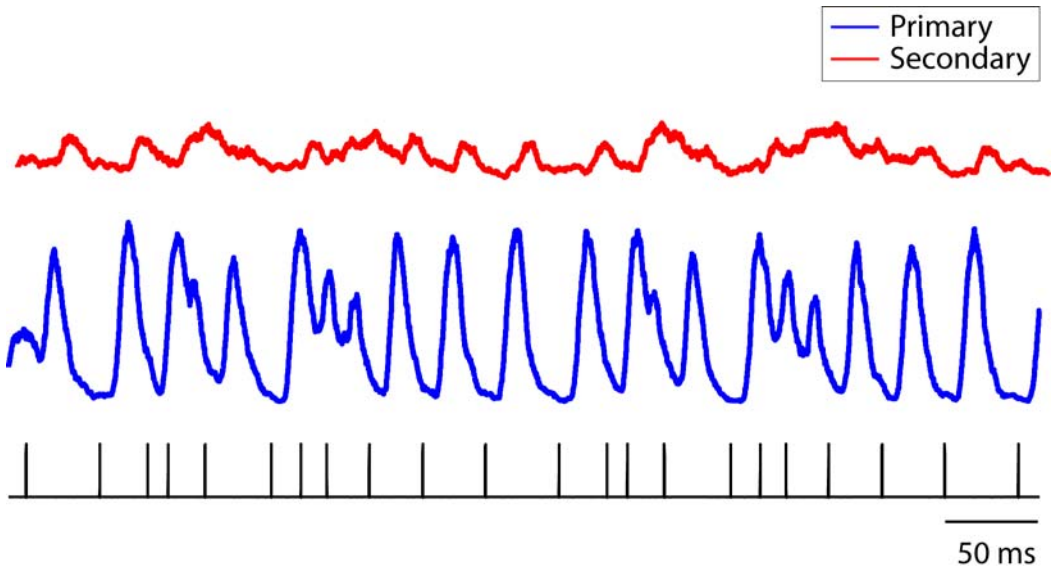


Figure 3 – figure supplement 4. Population PSTH of neurons in response to click sequences.

The combined population PSTH of neurons in the primary auditory area, L2a, is deeply modulated, a result of synchronous responses to the click sequence (blue trace, bin size: 5ms). The combined population PSTH of neurons in secondary areas (L2b and L3) is shown in red. The bottom tick marks show the waveform of the click stimulus (click sequence 1).

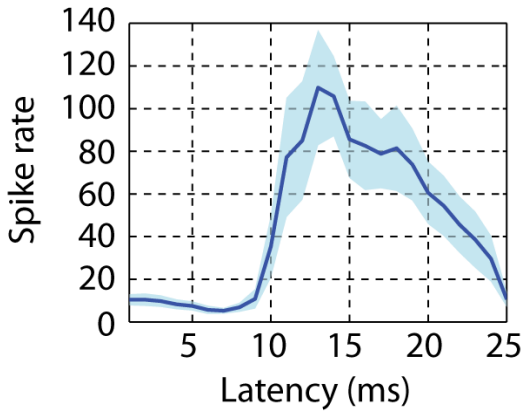


Figure 3 – figure supplement 5. Latency of neural responses to click sequences in primary auditory area, L2a.

To calculate the latency in the primary auditory area, a click triggered histogram of single unit responses is generated. The origin of this plot corresponds to the onset time of each click. The solid line represents the mean latency histogram and the shaded error bar is standard deviation of latency.

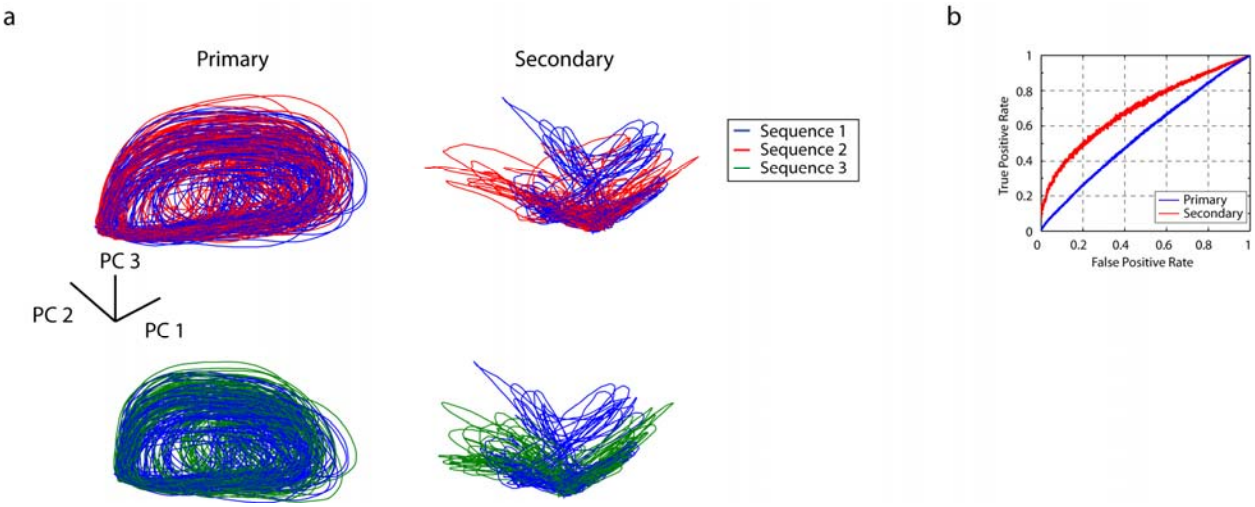


Figure 4. Temporal sequences are transformed to distinct population vectors in secondary auditory areas, L2b and L3.

(a) For different stimuli, ensemble state-space trajectories are discriminable in secondary auditory areas but not in the primary auditory area, L2a. For each trace, the bin size for the ensemble state space was 5ms. Each trace is smoothed by rectangular windows (10ms) for visualization.

(b) ROC analysis reveals enhanced discriminability of click sequences in secondary auditory areas, L2b and L3, relative to primary auditory area, L2a.

The following source data and figure supplements are available for Figure 4:

Figure 4 – Source data 1. Source data for ROC curve

Figure 4 – figure supplement 1. Short click sequence discriminability analysis

Figure 4 – Source data 1. Source data for ROC curve

This zip file contains spike timing data used for the ROC analysis shown in Figure 4b. Spike times of 10 different cells recorded in primary or secondary auditory areas are included in folders with corresponding names. For simple visualization of spike rasters, Matlab source code (DataLoad.m) is also provided.

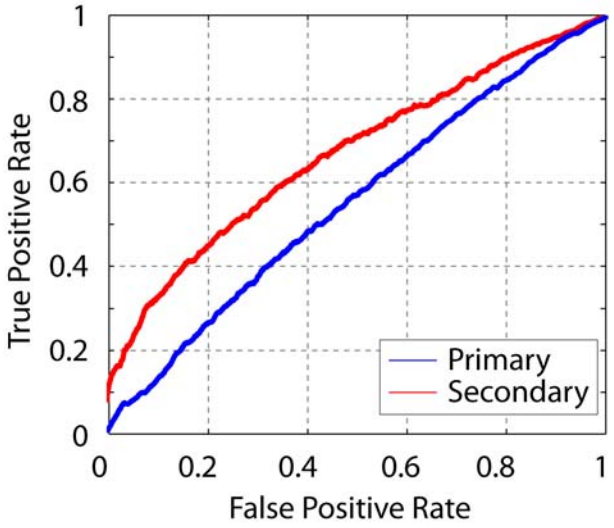


Figure 4 – figure supplement 1. Short click sequence discriminability analysis. ROC analysis shows that the sequence discriminability in secondary auditory areas is maintained even when considering only the first 500ms of the neural response.

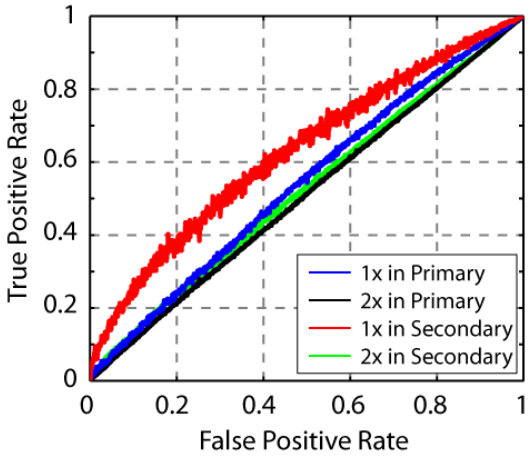


Figure 5. Neural sequence discriminability depends on the timescale of the click sequence.

ROC analysis reveals that the discriminability of the click sequences is constrained by the interval distribution of the click stimuli. When the sequence is slowed by a factor of two, the discriminability of click sequences is lost in the secondary auditory area (shown in green).

The following figure supplements are available for Figure 5:

Figure 5 – figure supplement 1. Spike rate of cells in response to click sequences with different timescales

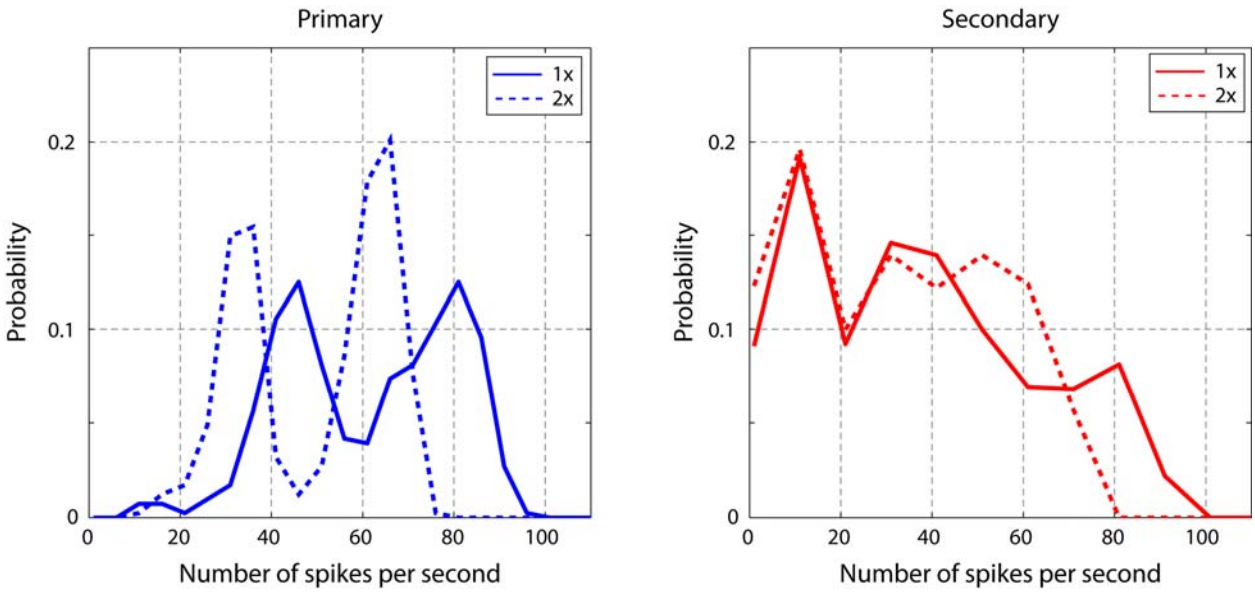


Figure 5 – figure supplement 1. Spike rate of cells in response to click sequences with different timescales.

Slower click sequences evoke a lower spike rate in primary and secondary auditory areas. For the secondary auditory areas, this reduction in spike rate is relatively small. This analysis was based on data used in Figure 5.

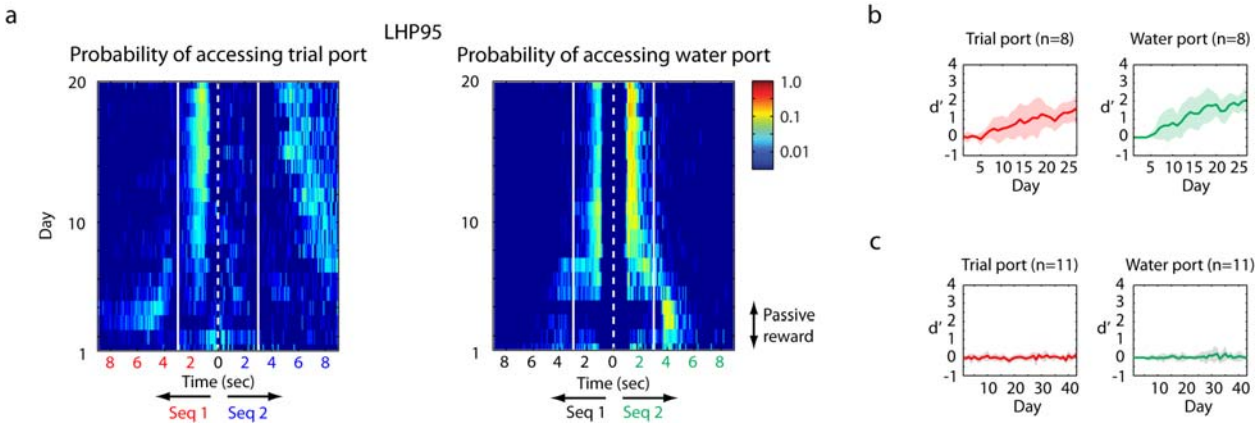


Figure 6. Operant training with click sequences.

(a) Example of training by the single-stage behavioral shaping method. The probability distribution of accessing trial port and water port is illustrated in log scale. The white dotted line represents the start of sequence playback and white solid line is the termination time of the stimulus (We show two stimuli back to back with mirrored time axes. Asymmetry in this image between the solid lines indicates learning). Over the course of training, this bird started to interrupt playback of the non-rewarding sequence by accessing the trial port before sequence 1 (the unrewarded sequence) stopped playing. The bird also learned to access the water port selectively during the playback of the rewarded sequence (sequence 2).

(b) Learning curve for birds exposed to the single-stage training method (n=8 birds). With the single-stage training method, most birds start to show differentiated responses (d' is around 1) after two weeks of training; that is, they interrupt and reset sequence 1 playback and access the water port for sequence 2 playback.

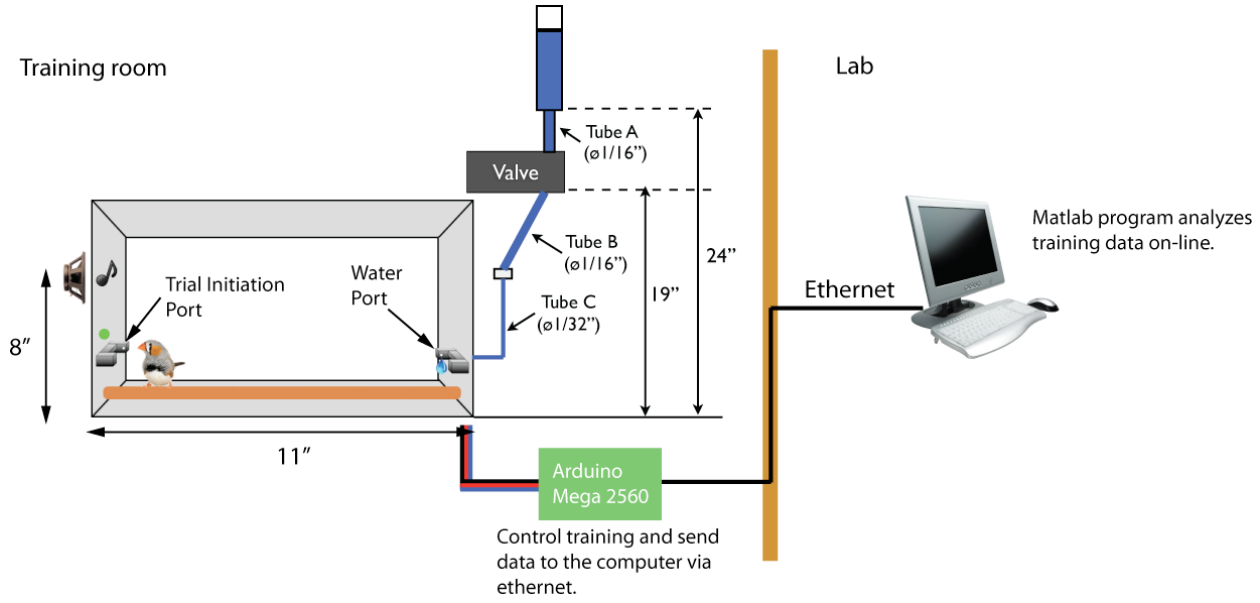
(c) When the click intervals are slowed by a factor of two, all trained birds (n=11 in the single-stage method) were unable to discriminate the temporal sequences; d' is around 0.

The following figure supplements and source data are available for Figure 6:

Figure 6 - figure supplement 1. Operant training setup.

Figure 6 - figure supplement 2. Result of operant training.

Figure 6 – source data 1. Summary of training.



964 **Figure 6 - figure supplement 1. Operant training setup.**

965 There are two infrared switches, a green LED (trial indicator), and a water spout in the
966 training cage. An Arudino microprocessor monitors the timing of port access, plays stimuli,
967 and delivers water rewards. The water reservoir is located 24 inches above the floor of
968 cage. The water valve is opened for a fixed duration, just long enough to produce a drop of
969 water that is consistently 1-5µl in volume. During operant training, data collected by the
970 Arduino is sent to another computer over Ethernet and analyzed in real-time.

971

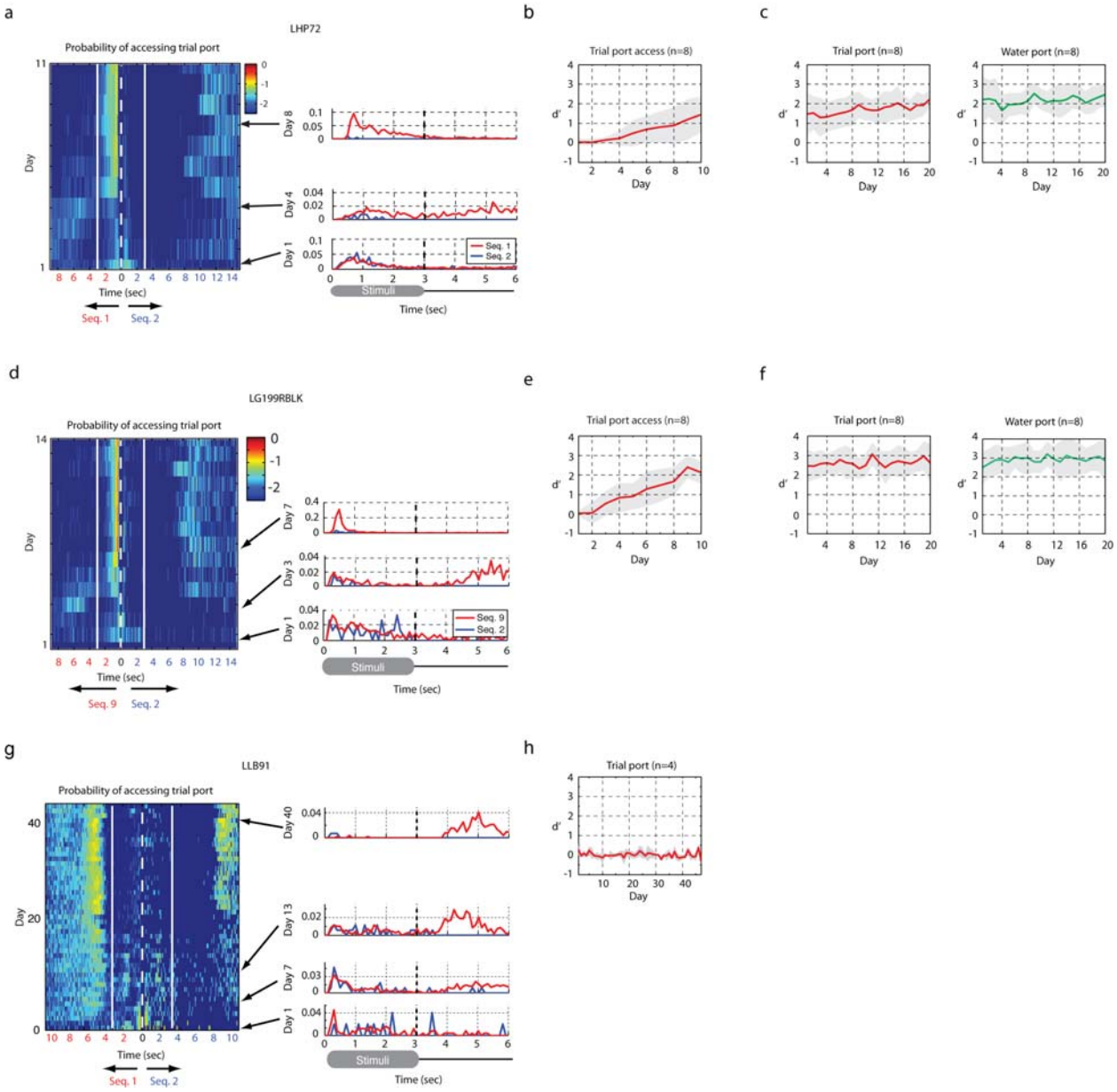


Figure 6 - figure supplement 2. Result of operant training.

(a,b,c) Two-stage training, example of a bird learning sequence 1. **(a)** The probability distribution of the bird accessing the trial port is shown for the entire training period of the first stage of training (left). The white dotted line represents the start of sequence playback and the white solid line shows the termination time of the stimulus. Any asymmetry between the dotted and solid lines indicates learning (asymmetry implies different behaviors for rewarded and non-rewarded sequences). This bird started to interrupt non-rewarding trials

around day 5. Individual rows (specific days in panel (a)) are plotted to the right to illustrate detail. **(b)** Learning curve for the first stage. Mean d-prime (± s.d.) after ten days of training is shown (n=8 birds). **(c)** Learning curve of training after passive reward is switched off (the second stage of training). This transition resulted in a minimal change in behavior.

(d,e,f) Example of two-stage training for another bird learning a distinct sequence (sequence 9). **(d)** The probability of accessing the trial port during the first stage of training (left) and three sample days (right). **(e)** Learning curve at the first stage (n=8 birds). **(f)** Learning curve at the second stage (n=8 birds).

(g,h) Example of two-stage training for a sequence whose intervals were slowed by a factor of two. **(g)** Probability distribution of accessing the trial port during the first stage of training. This bird usually reinitiated trials immediately after the presentation of the click sequence or after drinking water for rewarded trials (note the increased probability of accessing trial port around 10sec). The absence of asymmetry between the dotted and solid lines indicates an absence of learning. **(h)** Learning curve during the first stage of training. No birds (n=4 birds in two-stage training) learned to discriminate the slowed click sequences over the course of 40 days of training.

Figure 6 – source data 1. Summary of training.

The success of operant training was determined based on the d-prime score. When d' is over 1, the bird was deemed successful in learning the task. In this table, the number of birds that succeeded in operant training for click sequence discrimination (d'>1) out of total number of birds is shown. For example, 8 out of 10 birds succeeded in two-stage training to discriminate sequence 9 and 2.

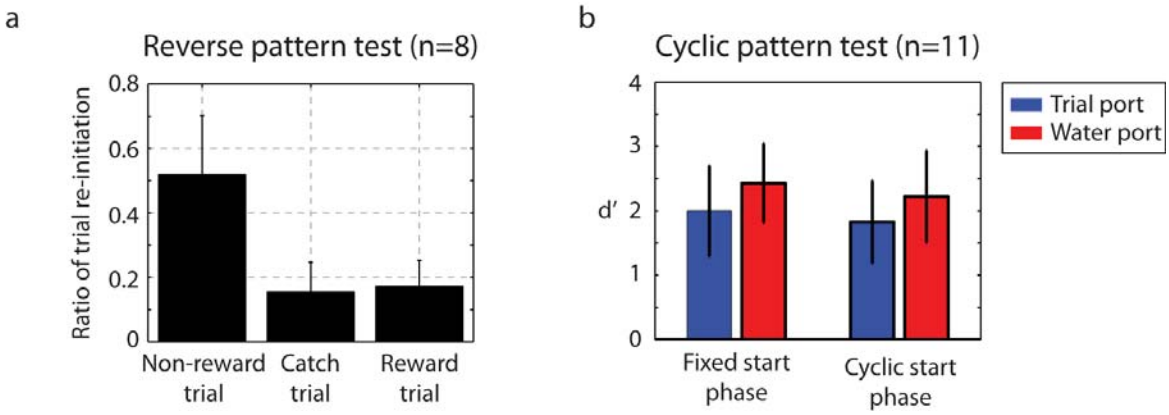


Figure 7. Catch trial analysis.

(a) During catch trial analysis, for 10% of non-rewarding trials, we presented reverse patterns to eight birds. The birds did not show any recognition of the reverse pattern (catch trials). Only the familiar non-rewarded sequence led to the adaptive behavior of resetting playback. Mean ± s.d. of trial interruption ratio is shown.

(b) In this cyclic permutation catch trial analysis, playback of the click sequence started at a random interval in the repeating sequence on each trial (a phase shift in the stimulus order); all birds (n = 11) maintained performance. This indicates that discriminations were based on patterns of intervals regardless of the absolute time of any specific click relative to trial onset.

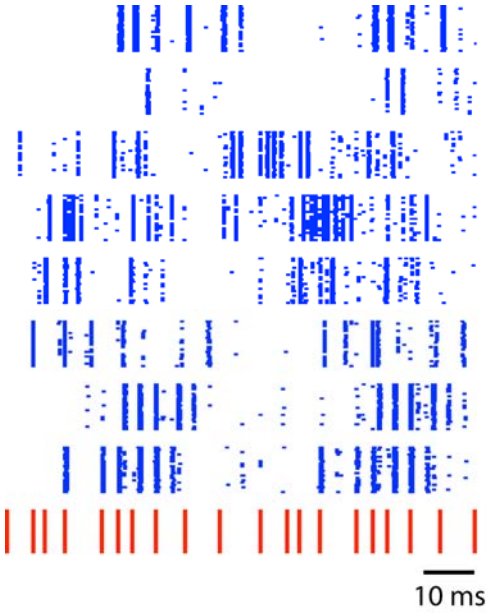
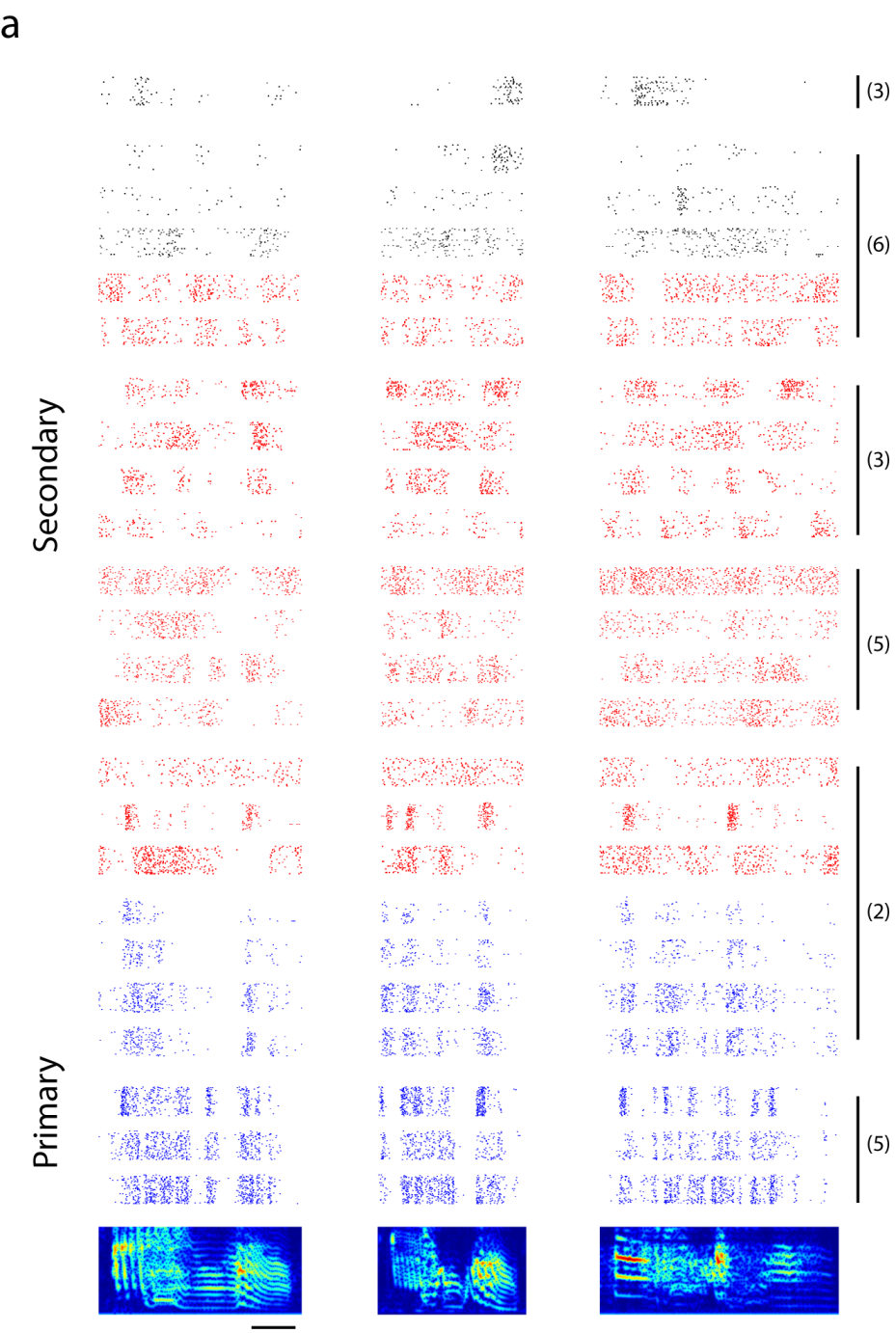


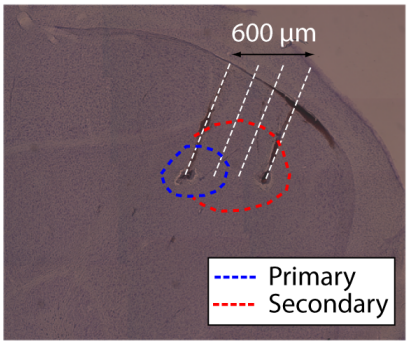
Figure 8. Illustration: sequence selective responses in a critically tuned linear dynamical system.

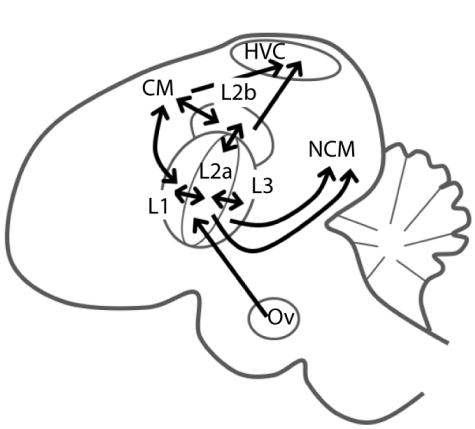
Each blue row represents simulated neural responses of a simple linear model. The input stimulus (red) has a temporal pattern similar to the click sequences used in this study. This toy model illustrates a temporal to spatial transformation arising from simple linear dynamics in a recurrent system.

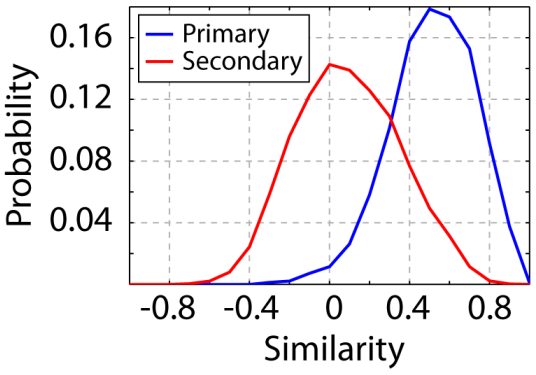
Supplementary file 1. Click sequence audio files.

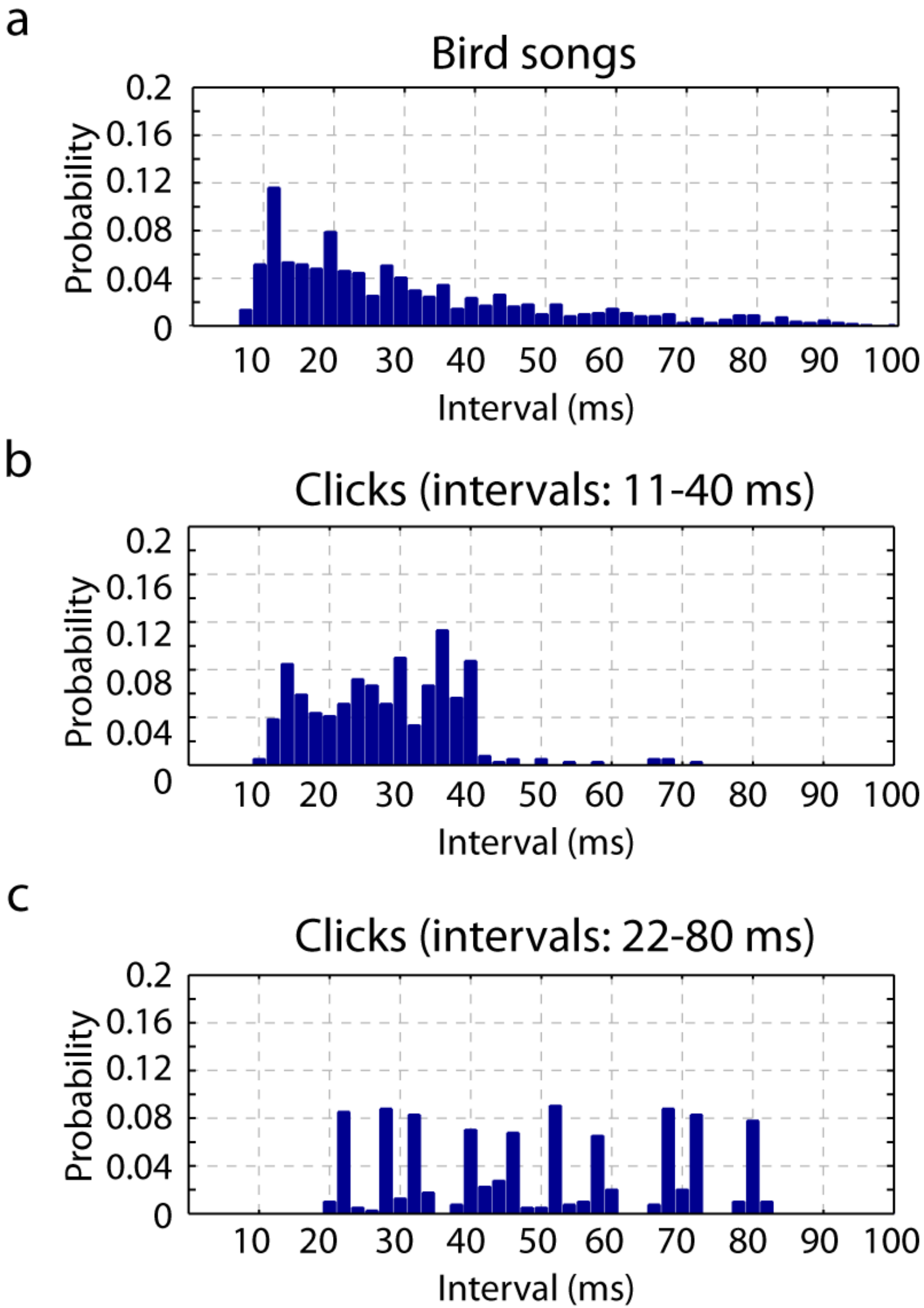
We provide audio files of all the click sequences used in this study in wav format. The last number of the file name corresponds to the index of click sequence. For example, Clk_Sequence_1.wav contains audio data for sequence 1.

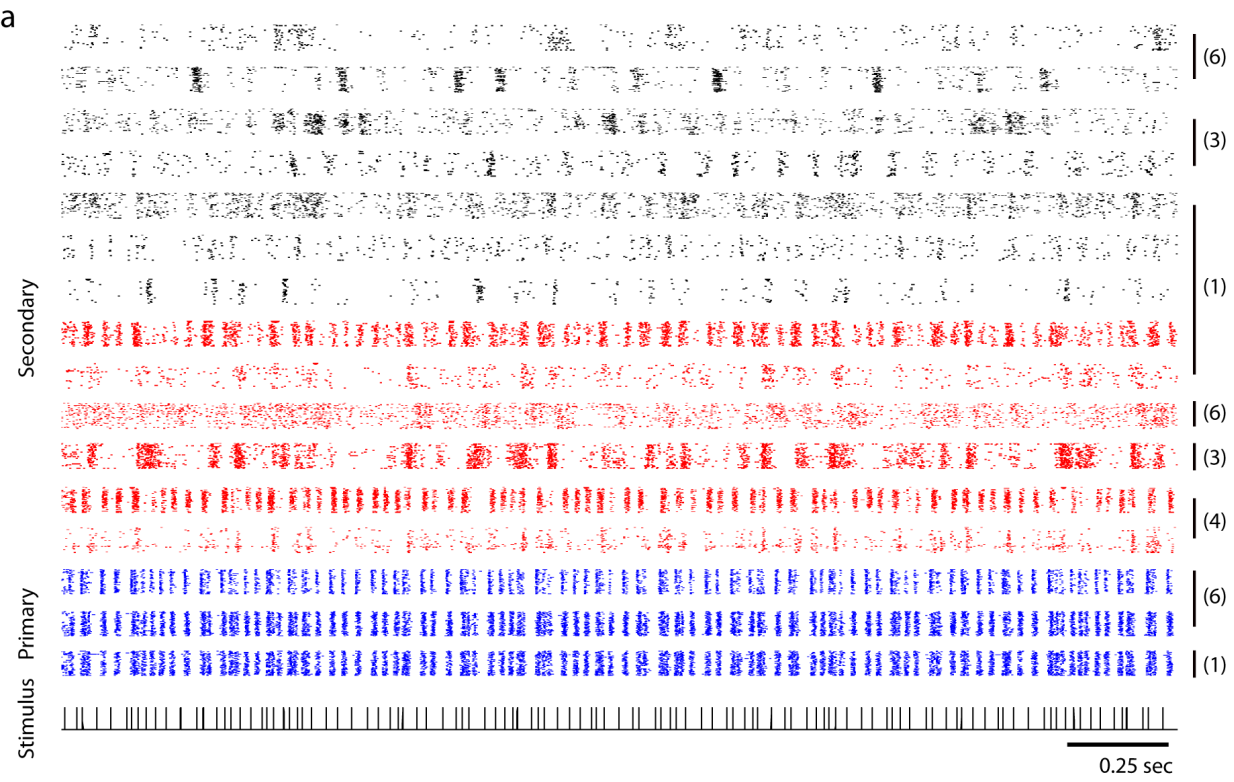

a
Secondary
Primary
(3)
(6)
(3)
(5)
(2)
(5)

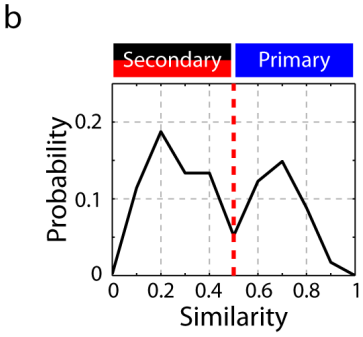

b
600 µm
Primary
Secondary


c
HVC
CM
L2b
NCM
L2a
L3
L1
Ov

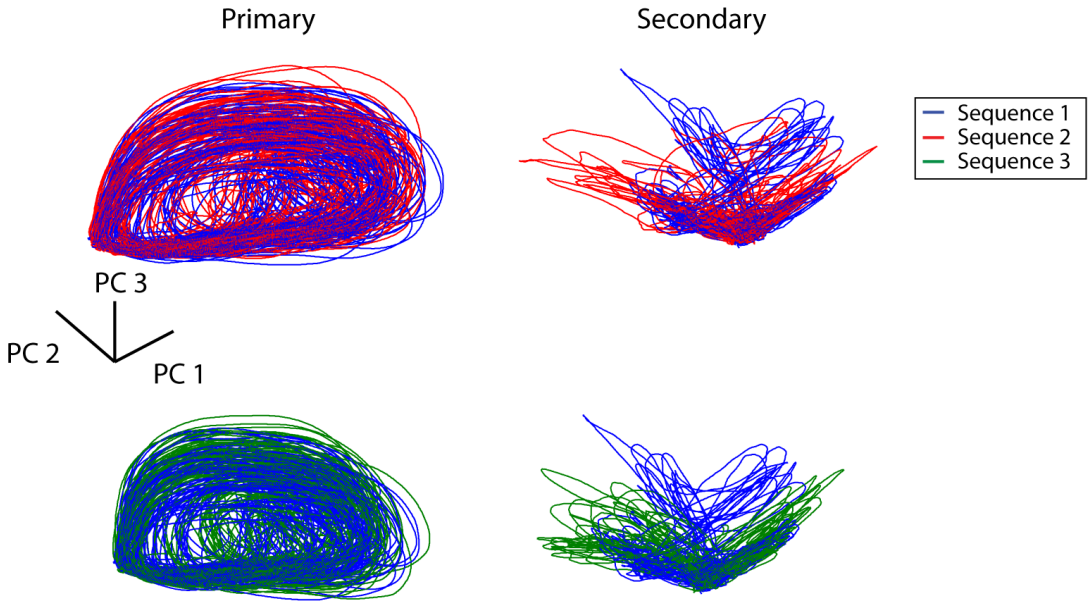

d
Primary
Secondary
Probability
0.16
0.12
0.08
0.04
-0.8
-0.4
0
0.4
0.8
Similarity


a
Bird songs
Probability
0.2
0.16
0.12
0.08
0.04
0
10 20 30 40 50 60 70 80 90 100
Interval (ms)
b
Clicks (intervals: 11-40 ms)
Probability
0.2
0.16
0.12
0.08
0.04
0
10 20 30 40 50 60 70 80 90 100
Interval (ms)
c
Clicks (intervals: 22-80 ms)
Probability
0.2
0.16
0.12
0.08
0.04
0
10 20 30 40 50 60 70 80 90 100
Interval (ms)

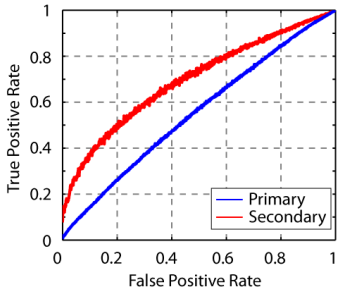

a
(6)
(3)
(1)
Secondary
(6)
(3)
(4)
(6)
Primary
(1)
Stimulus
0.25 sec

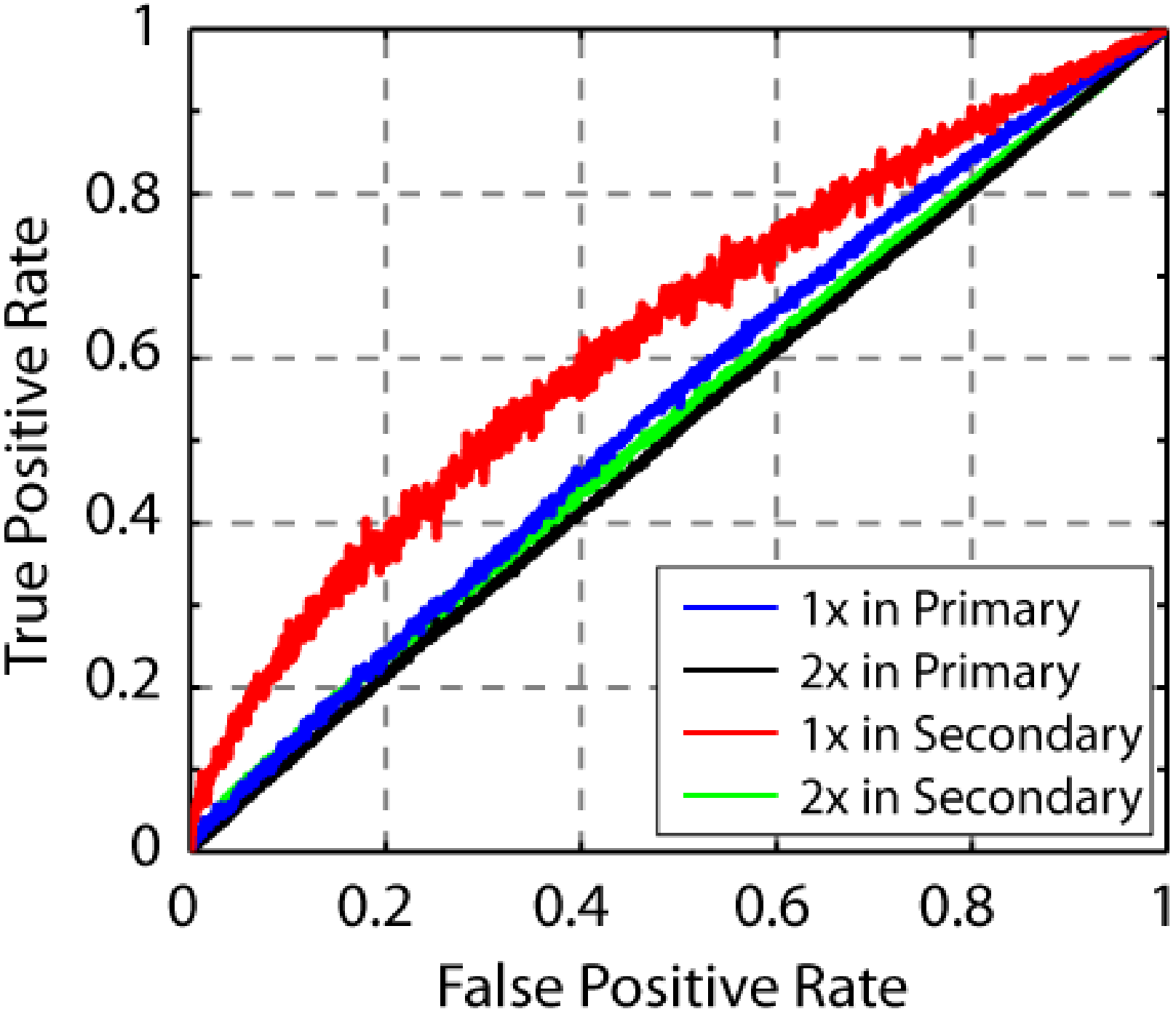

b
Secondary
Primary
Probability
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
Similarity

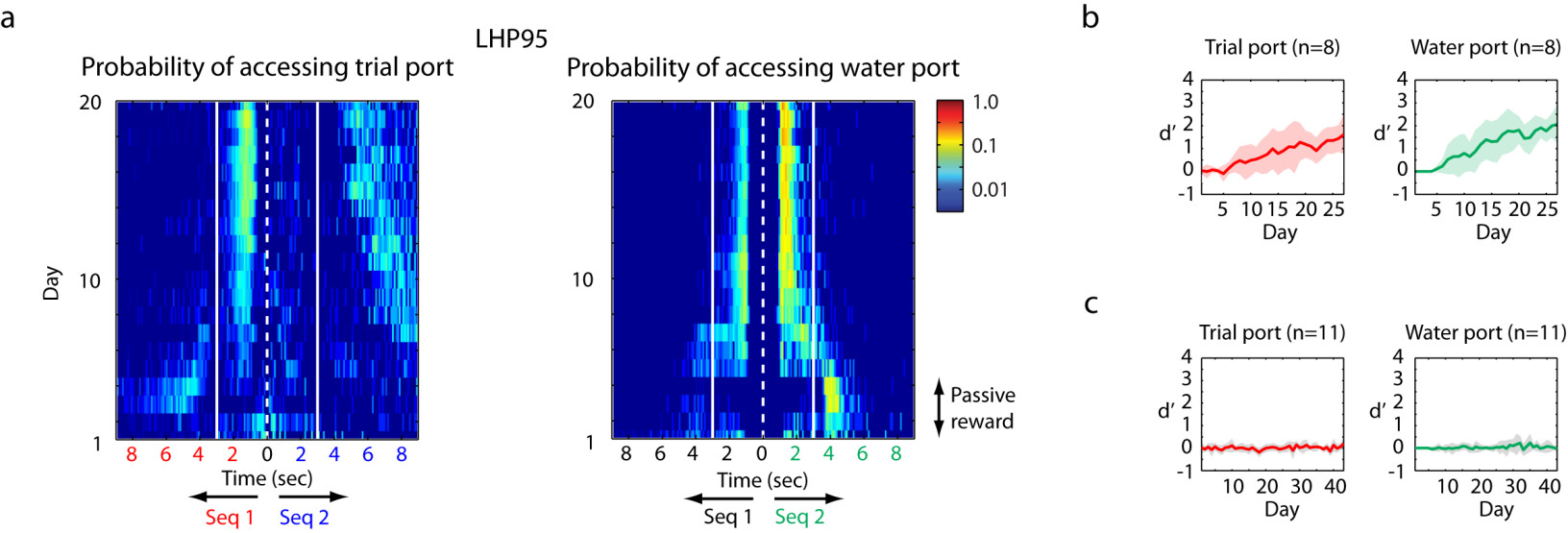
a


Primary
Secondary
Sequence 1
Sequence 2
Sequence 3
PC 3
PC 2
PC 1

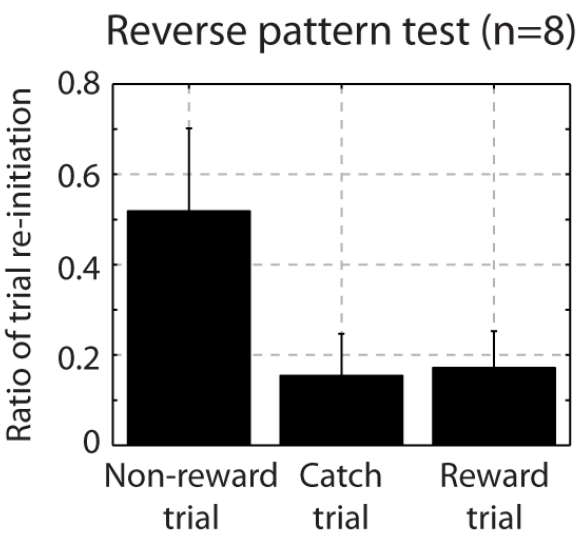
b


True Positive Rate
False Positive Rate
0
0.2
0.4
0.6
0.8
1
Primary
Secondary

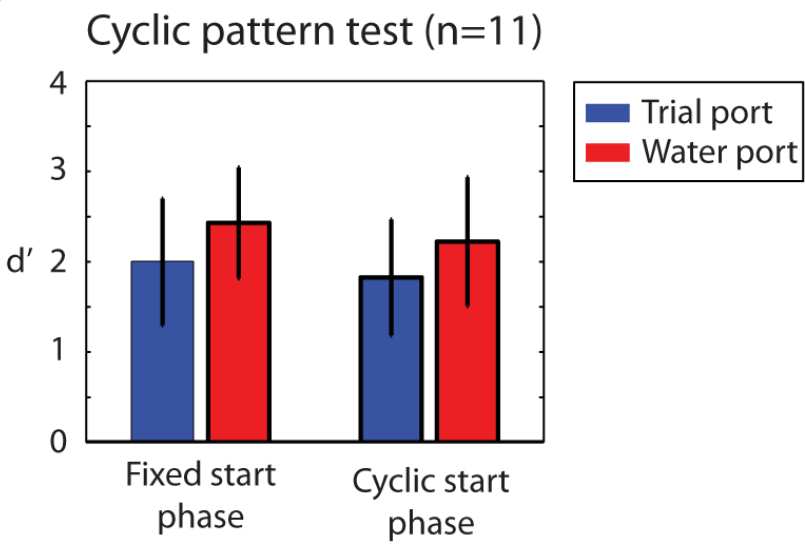

True Positive Rate
False Positive Rate
0
0.2
0.4
0.6
0.8
1
1x in Primary
2x in Primary
1x in Secondary
2x in Secondary

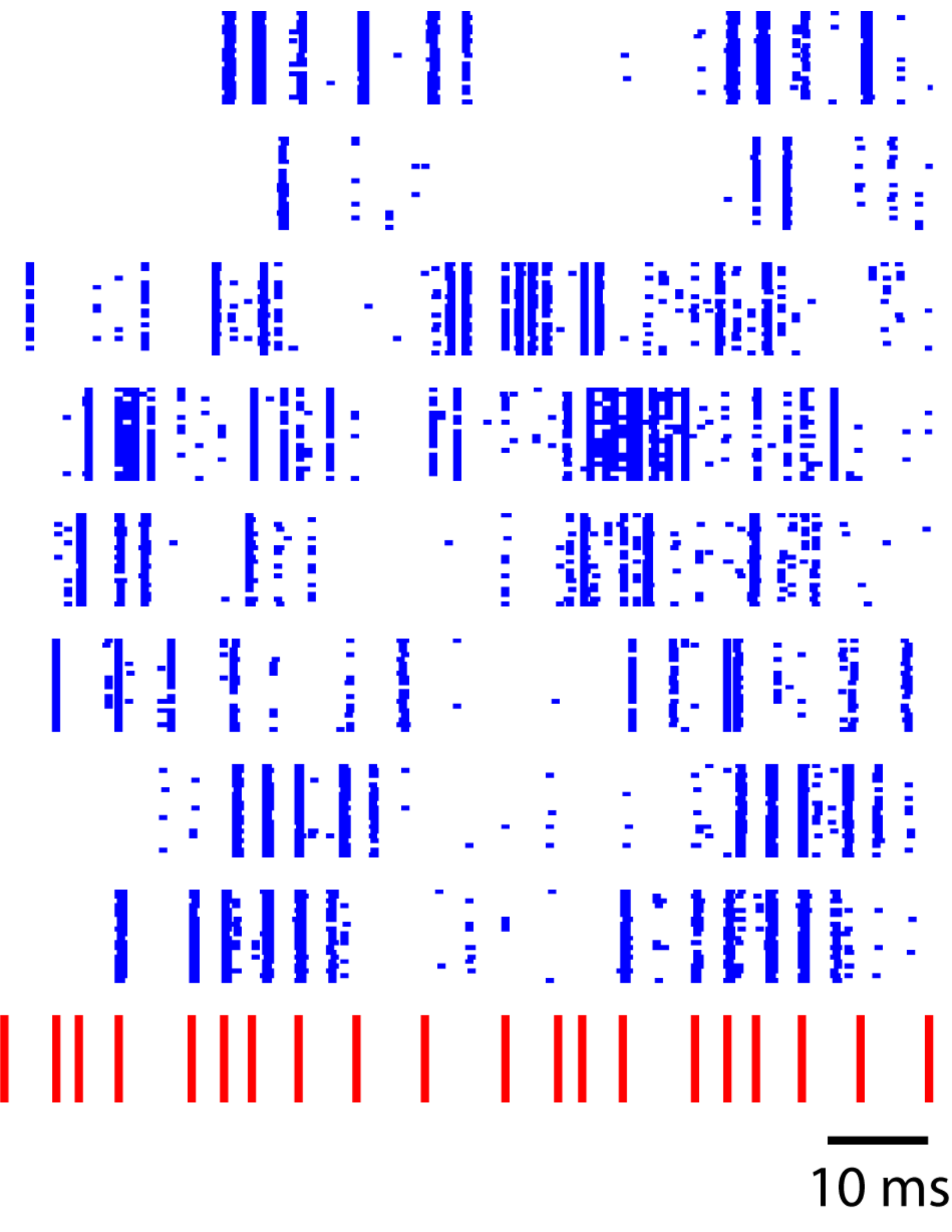

a
LHP95
Probability of accessing trial port
Probability of accessing water port
Day
20
10
1
8 6 4 2 0 2 4 6 8
Time (sec)
Seq 1
Seq 2
1.0
0.1
0.01
Passive reward
b
Trial port (n=8)
Water port (n=8)
d'
Day
c
Trial port (n=11)
Water port (n=11)

a

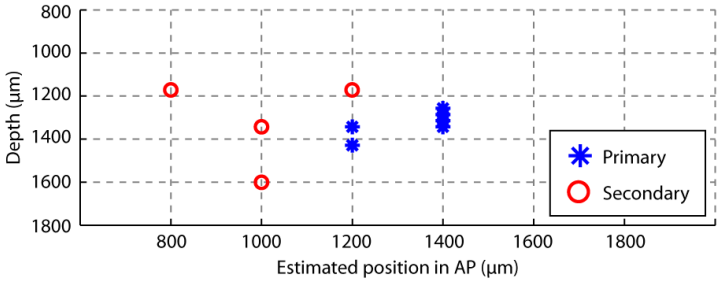

Reverse pattern test (n=8)
Ratio of trial re-initiation
0.8
0.6
0.4
0.2
0
Non-reward trial
Catch trial
Reward trial

b

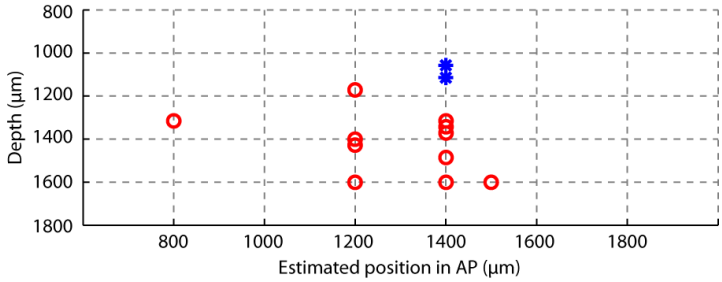

Cyclic pattern test (n=11)
d'
4
3
2
1
0
Fixed start phase
Cyclic start phase
Trial port
Water port


10 ms

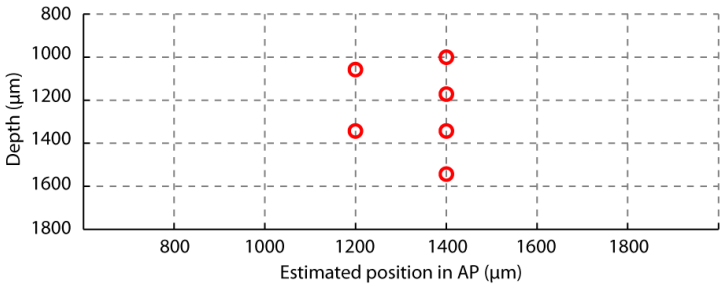
(1) LPI24



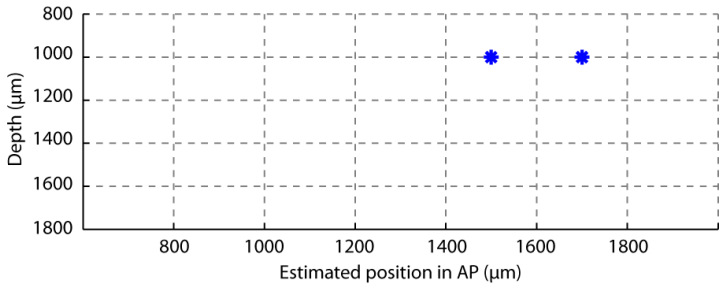
(2) LG344RBLK



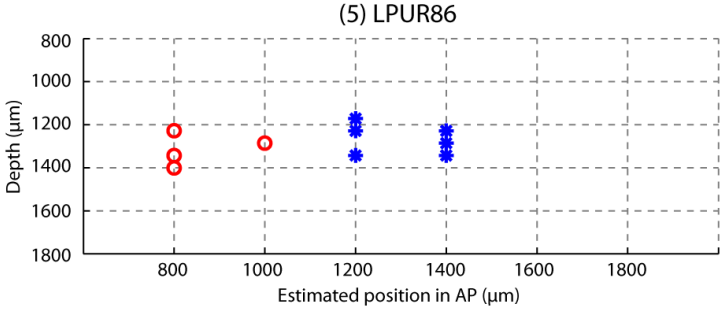
(3) LG304RBLK



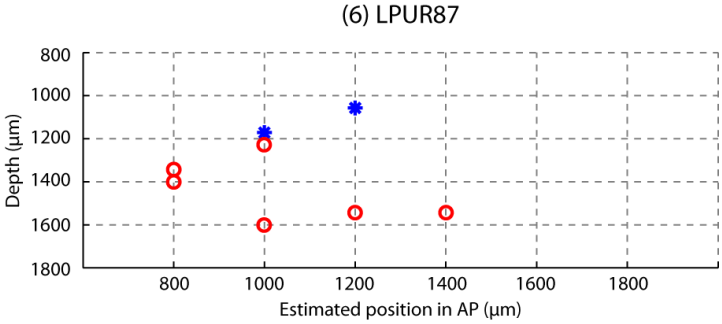
(4) LHP98



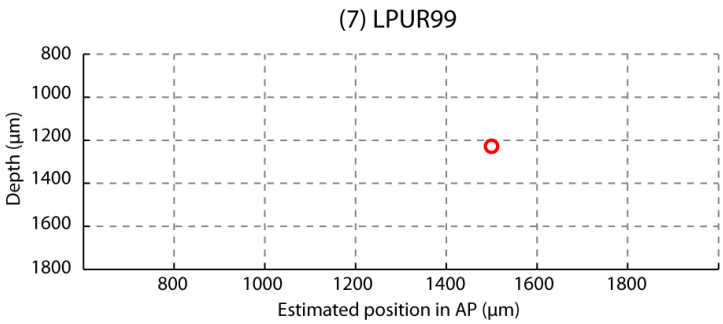
(5) LPUR86



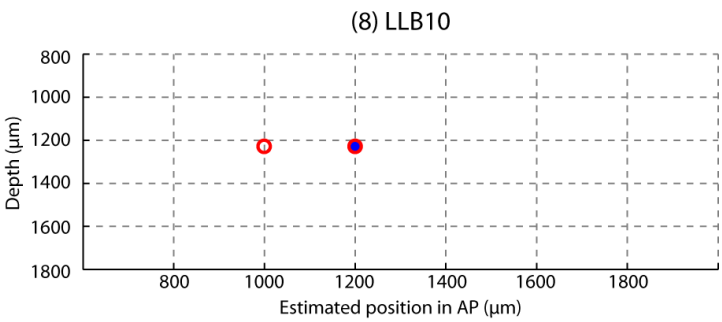
(6) LPUR87



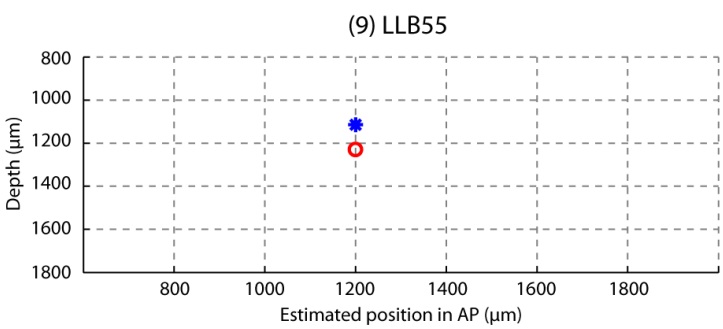
(7) LPUR99



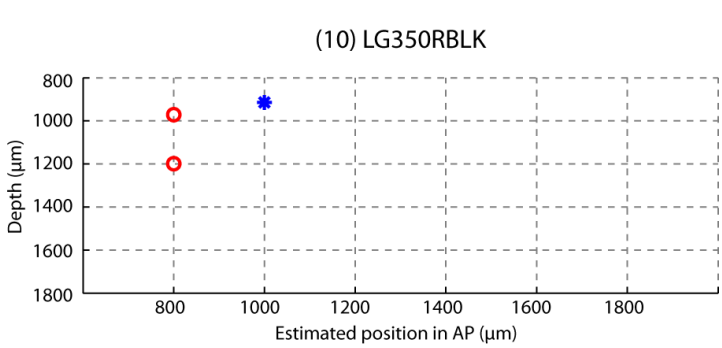
(8) LLB10



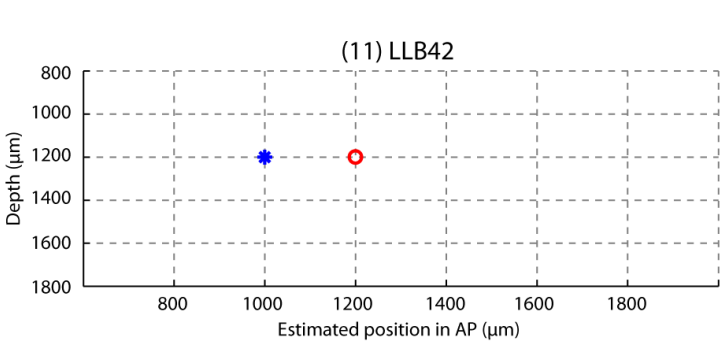
(9) LLB55

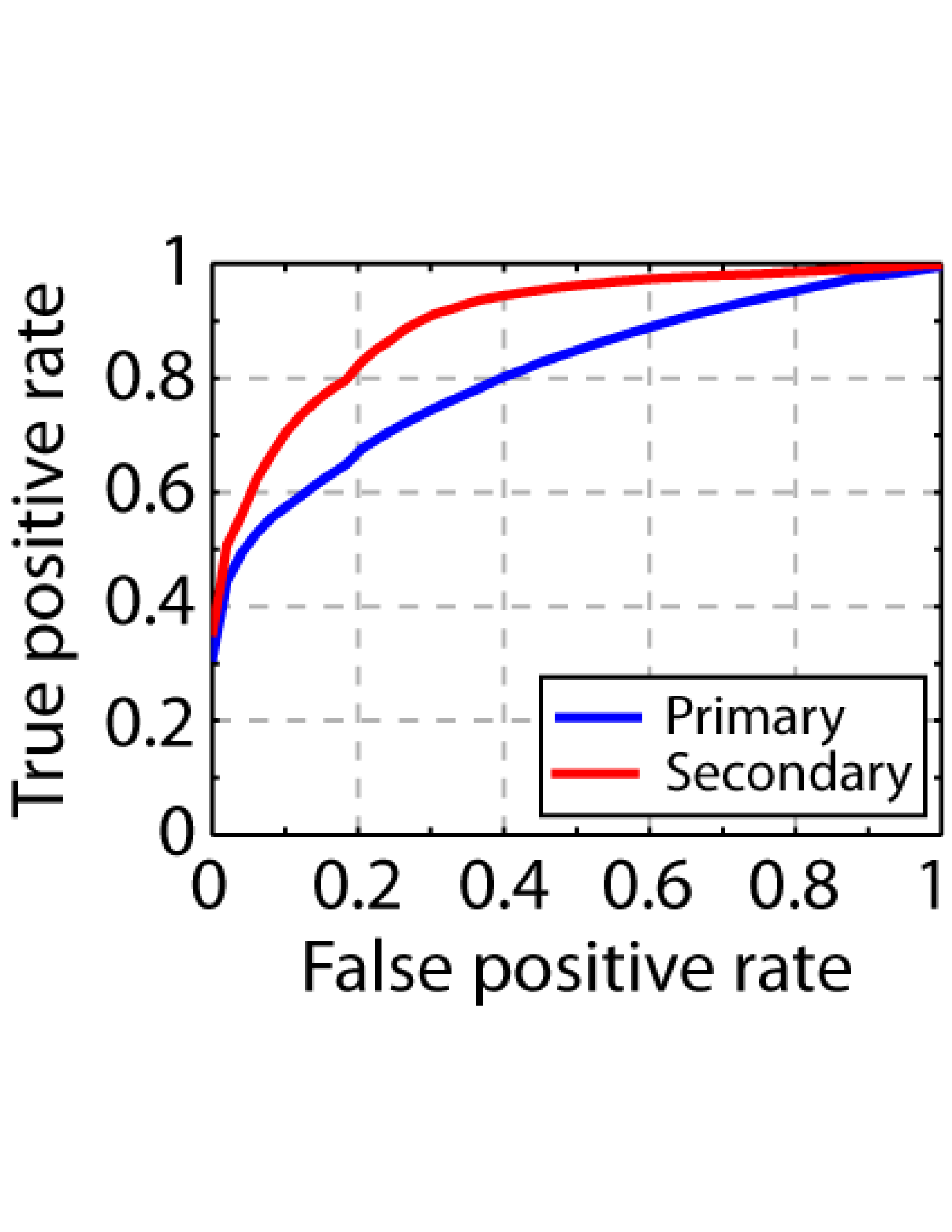


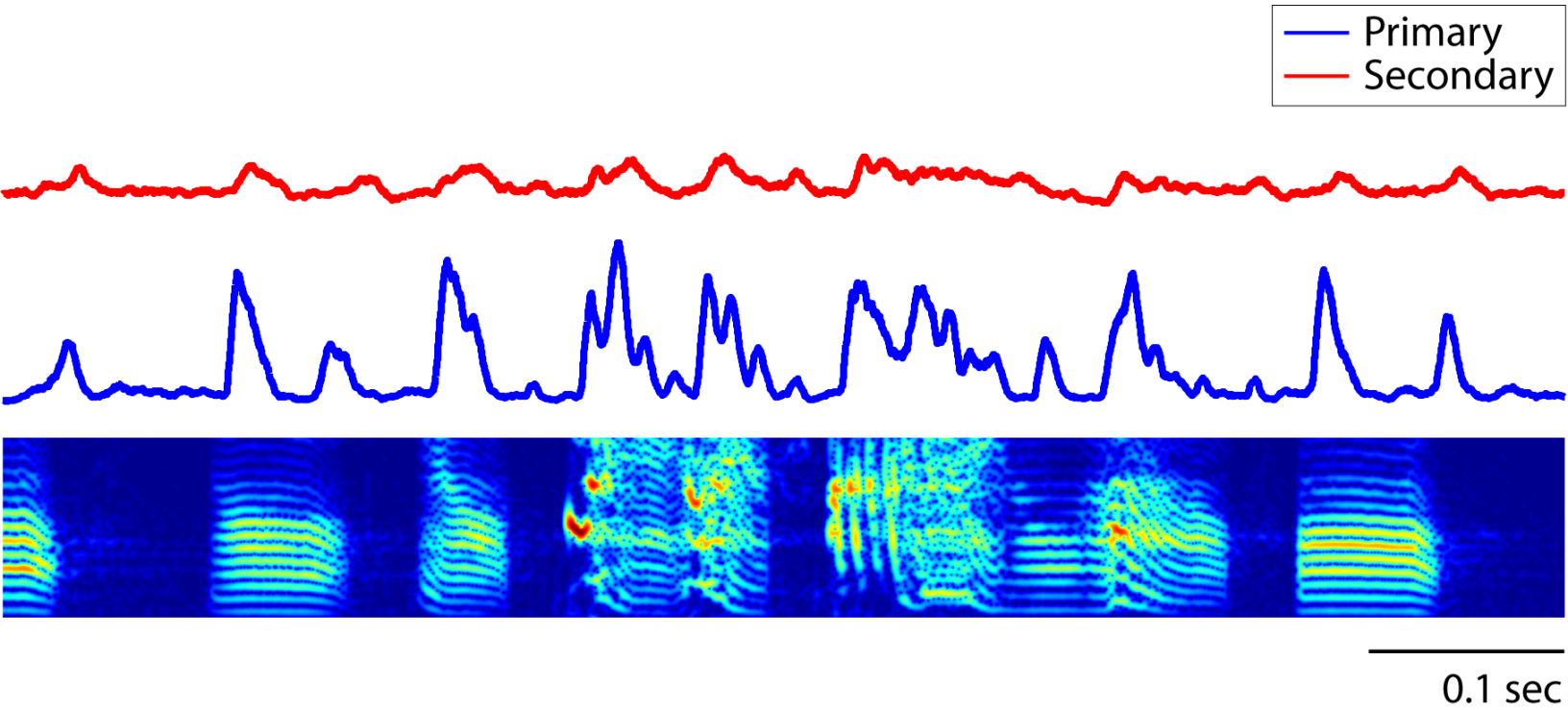
(10) LG350RBLK



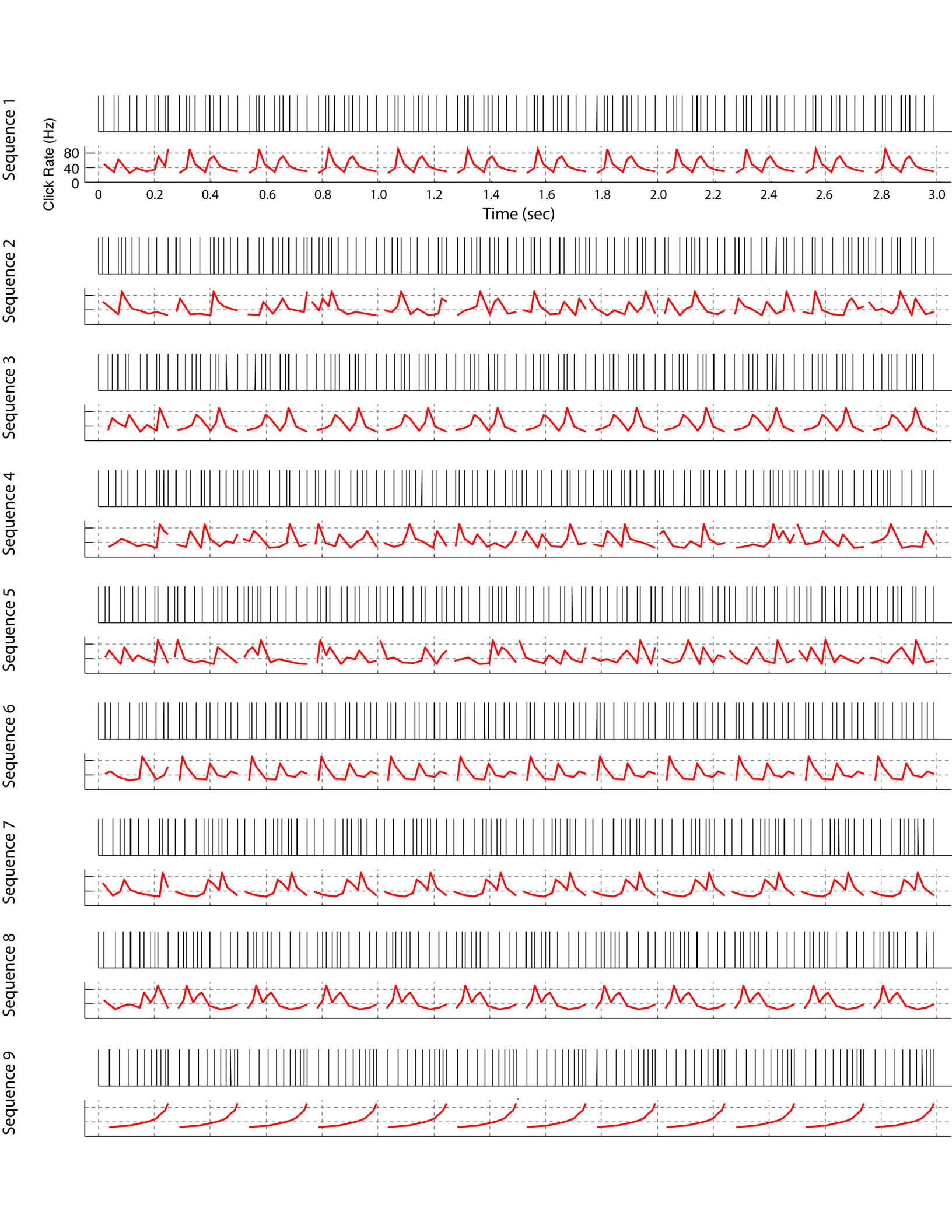
(11) LLB42

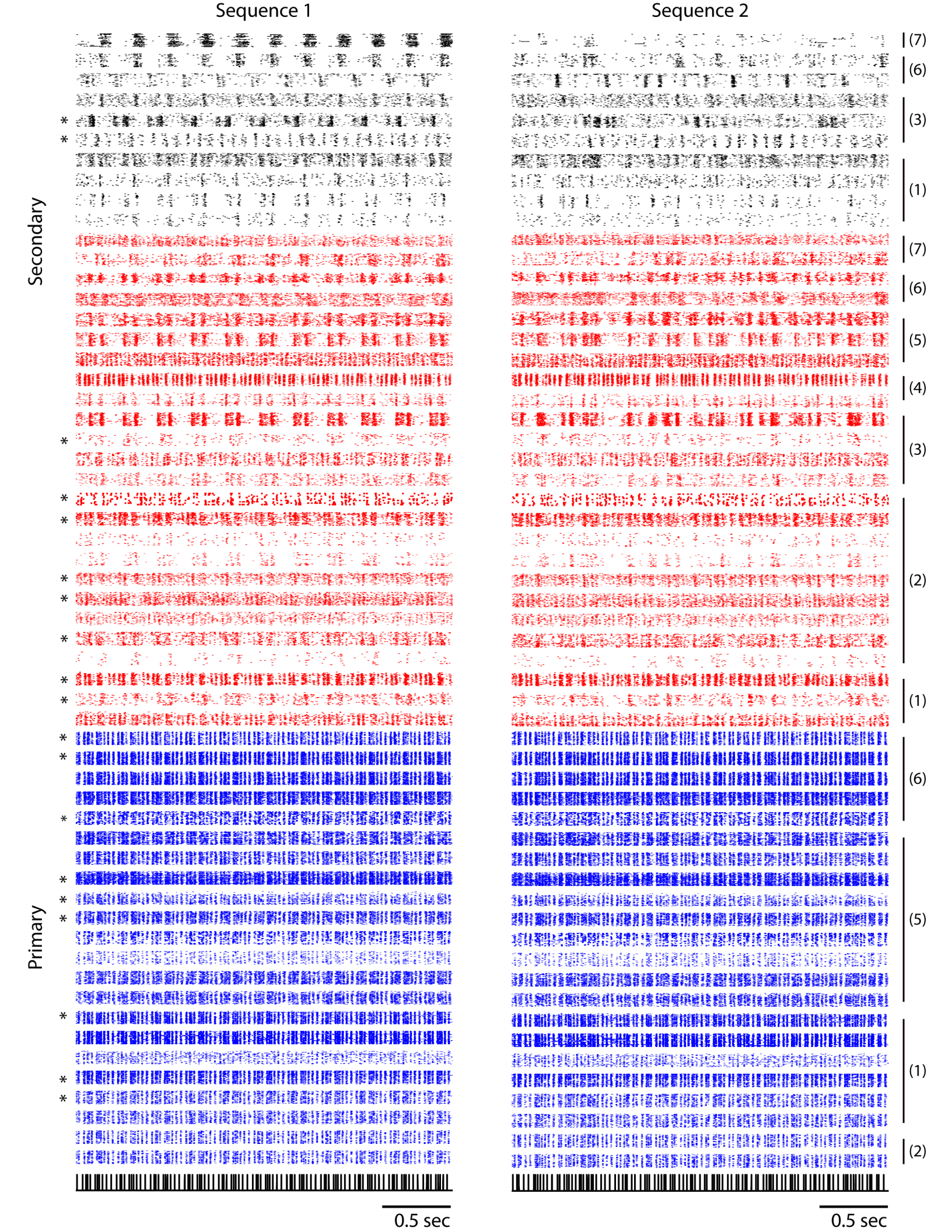


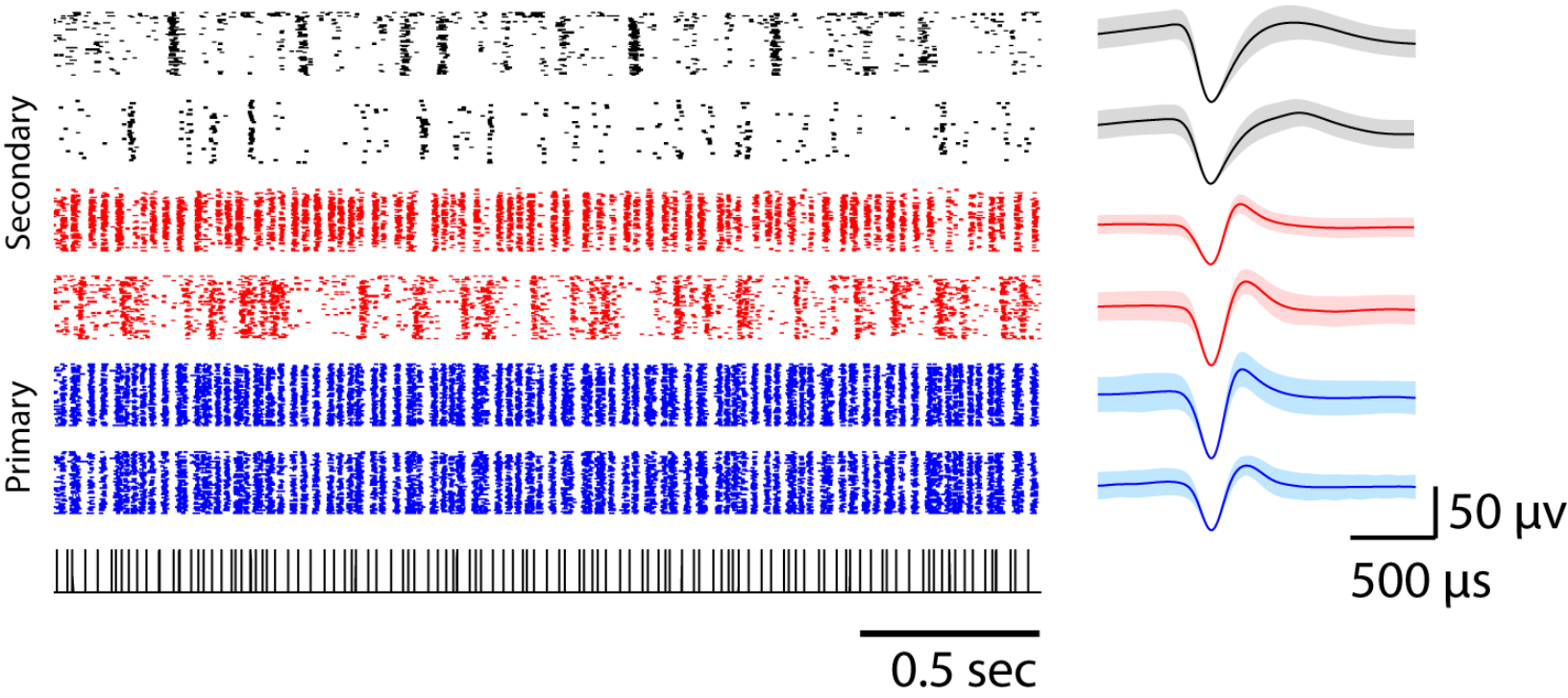

True positive rate
False positive rate
0
0.2
0.4
0.6
0.8
1
Primary
Secondary

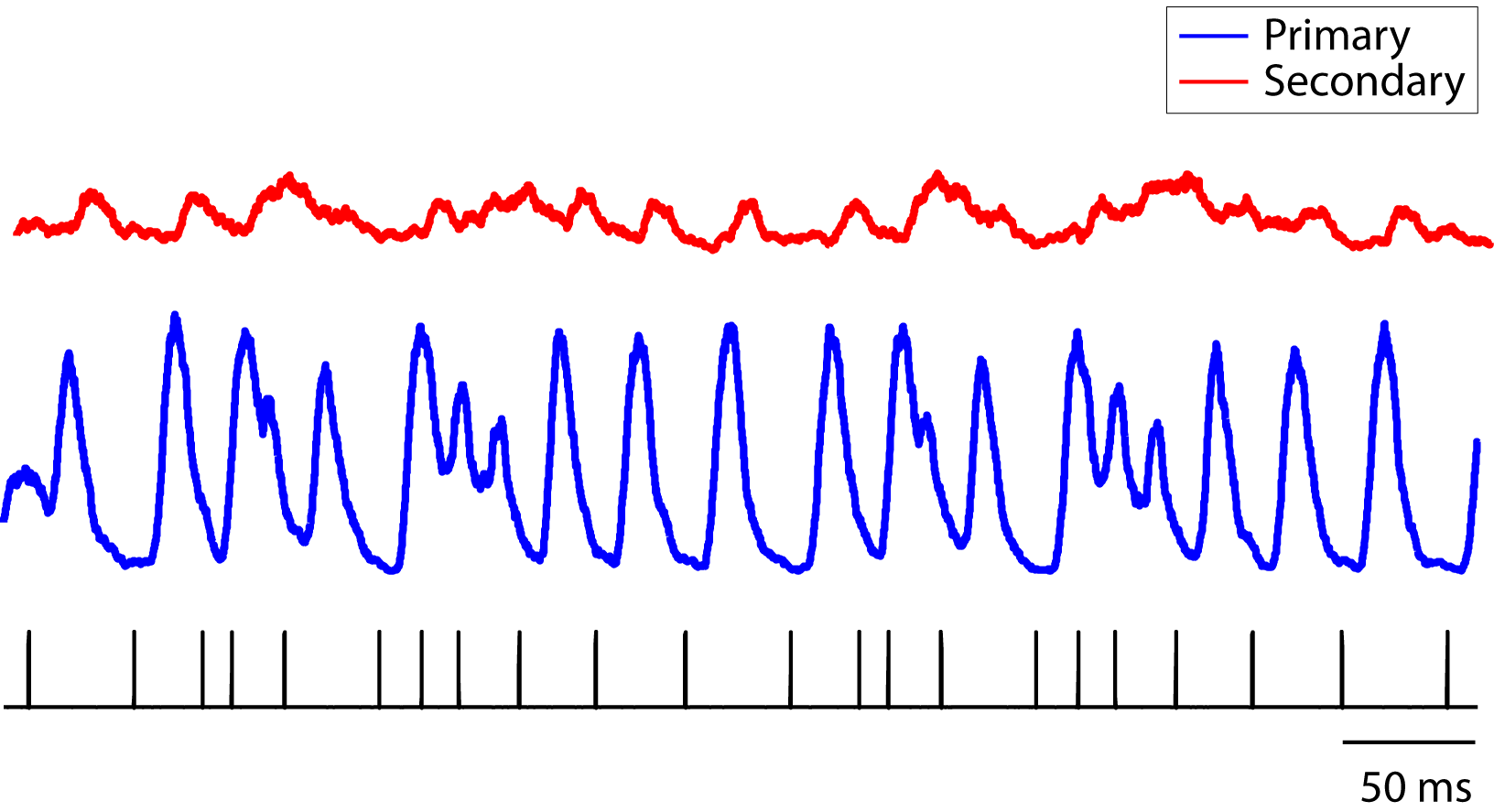


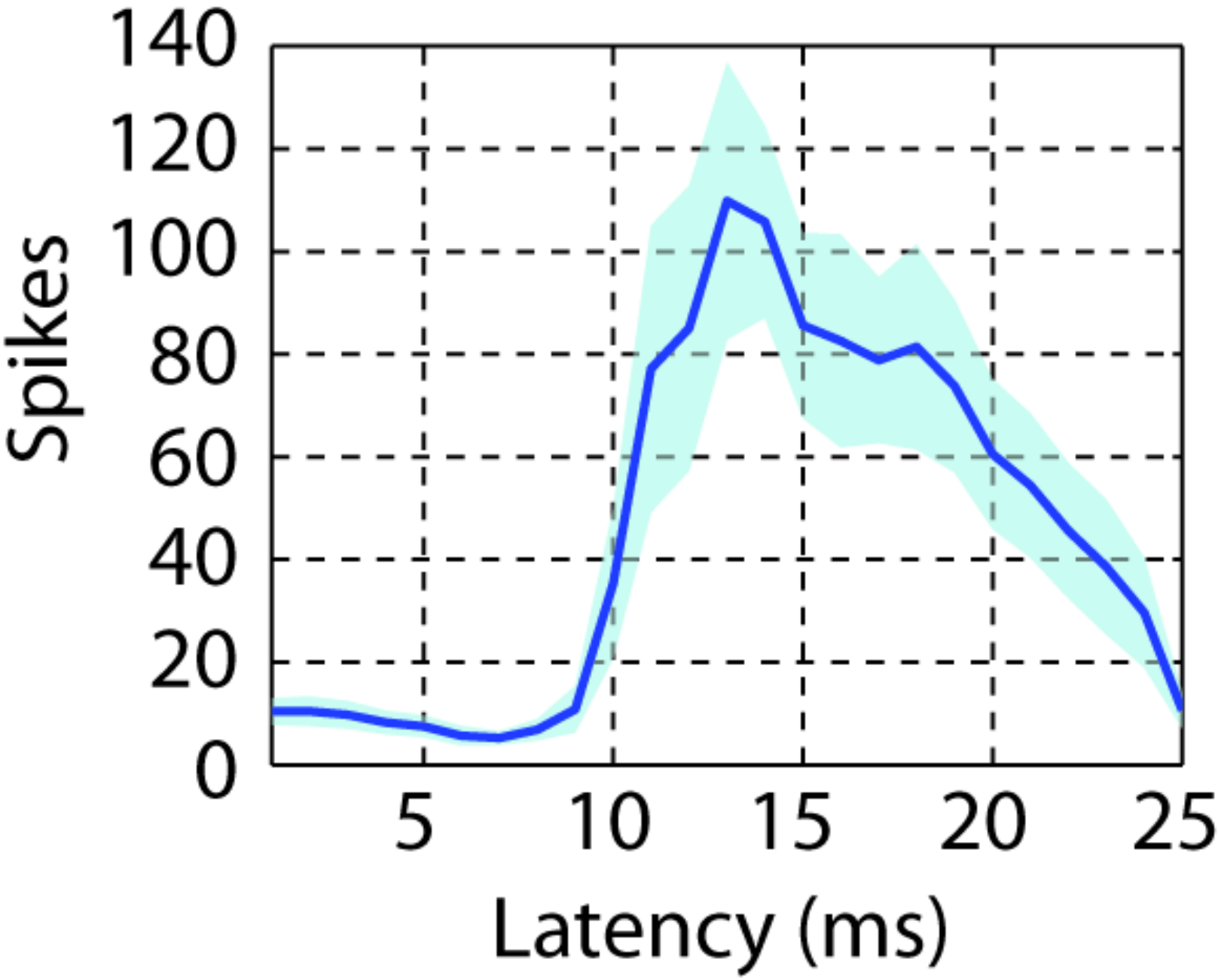
0.1 sec

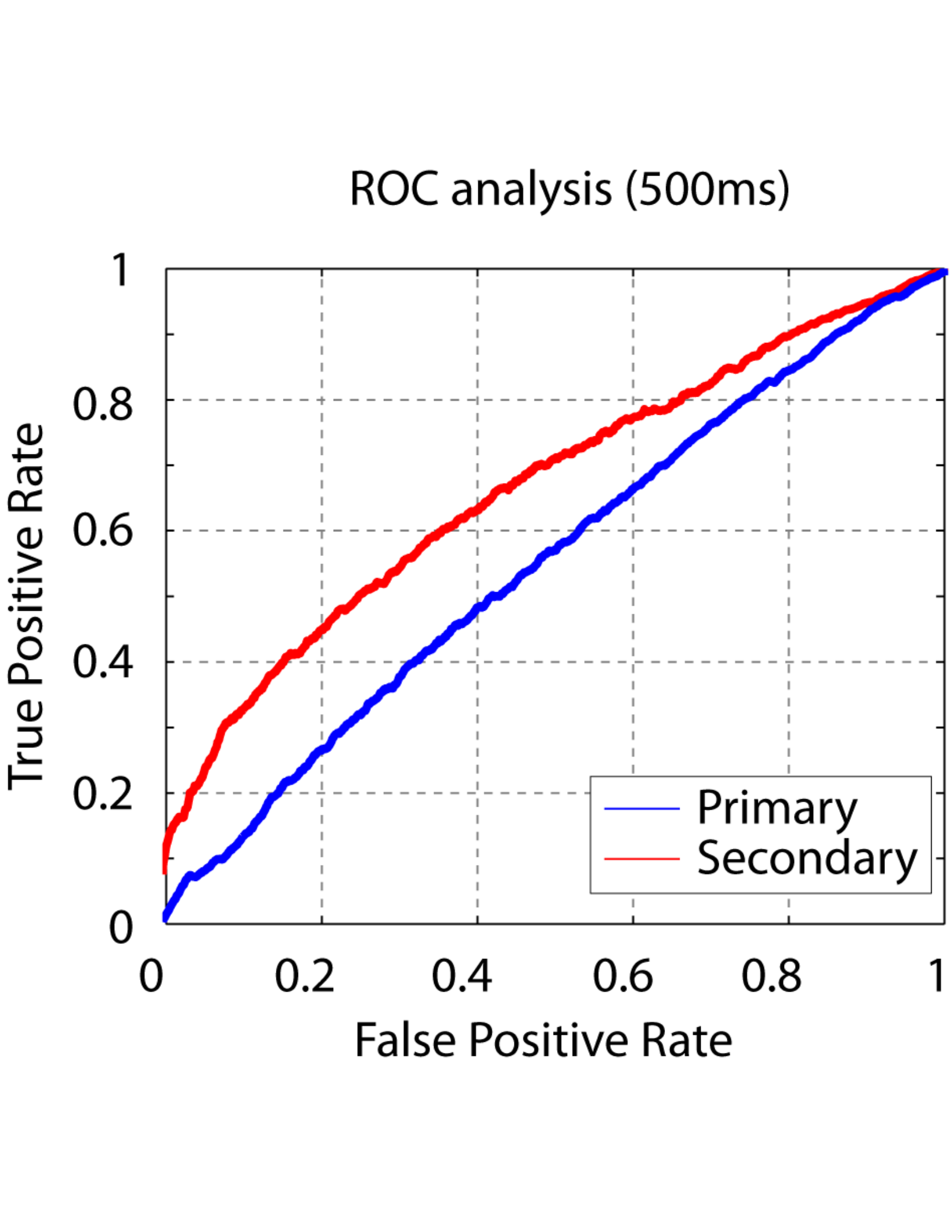

Sequence 1
Sequence 2
Sequence 3
Sequence 4
Sequence 5
Sequence 6
Sequence 7
Sequence 8
Sequence 9
Click Rate (Hz)
80
40
0
0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
Time (sec)

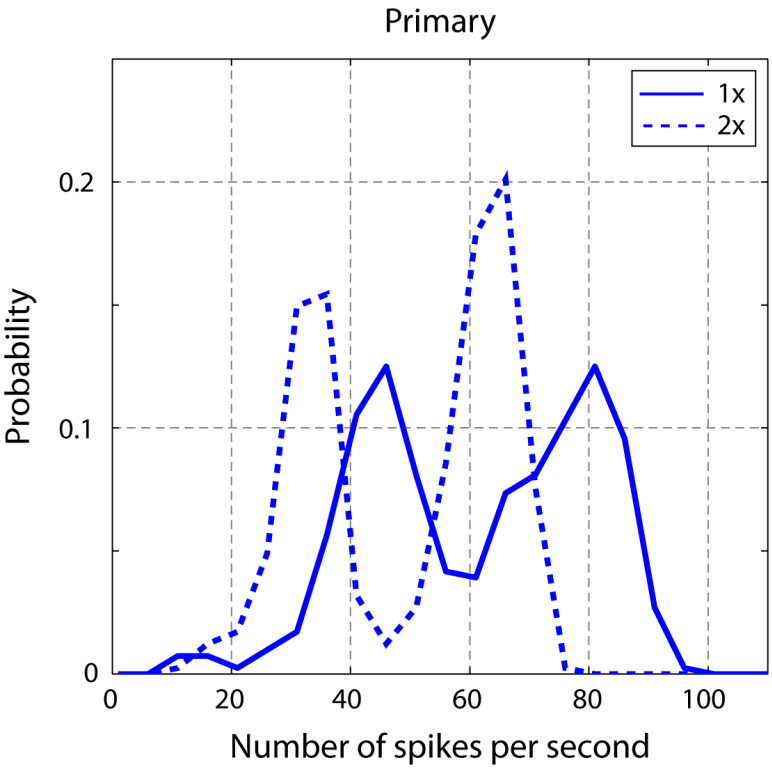

Sequence 1
Sequence 2
Secondary
Primary
(7)
(6)
(3)
(1)
(7)
(6)
(5)
(4)
(3)
(2)
(1)
(6)
(5)
(1)
(2)
0.5 sec
0.5 sec

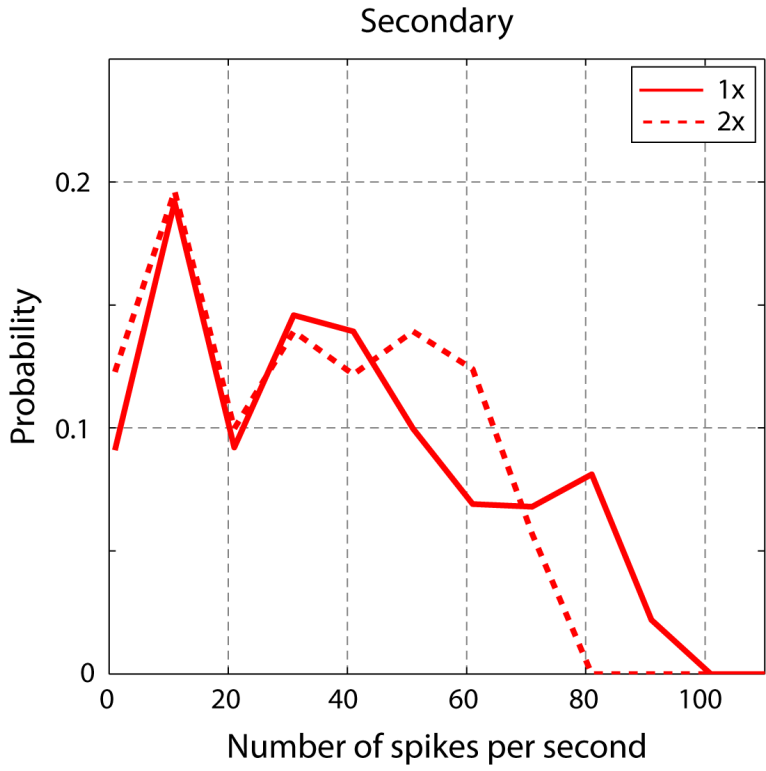

Secondary
Primary
0.5 sec
50 µv
500 µs

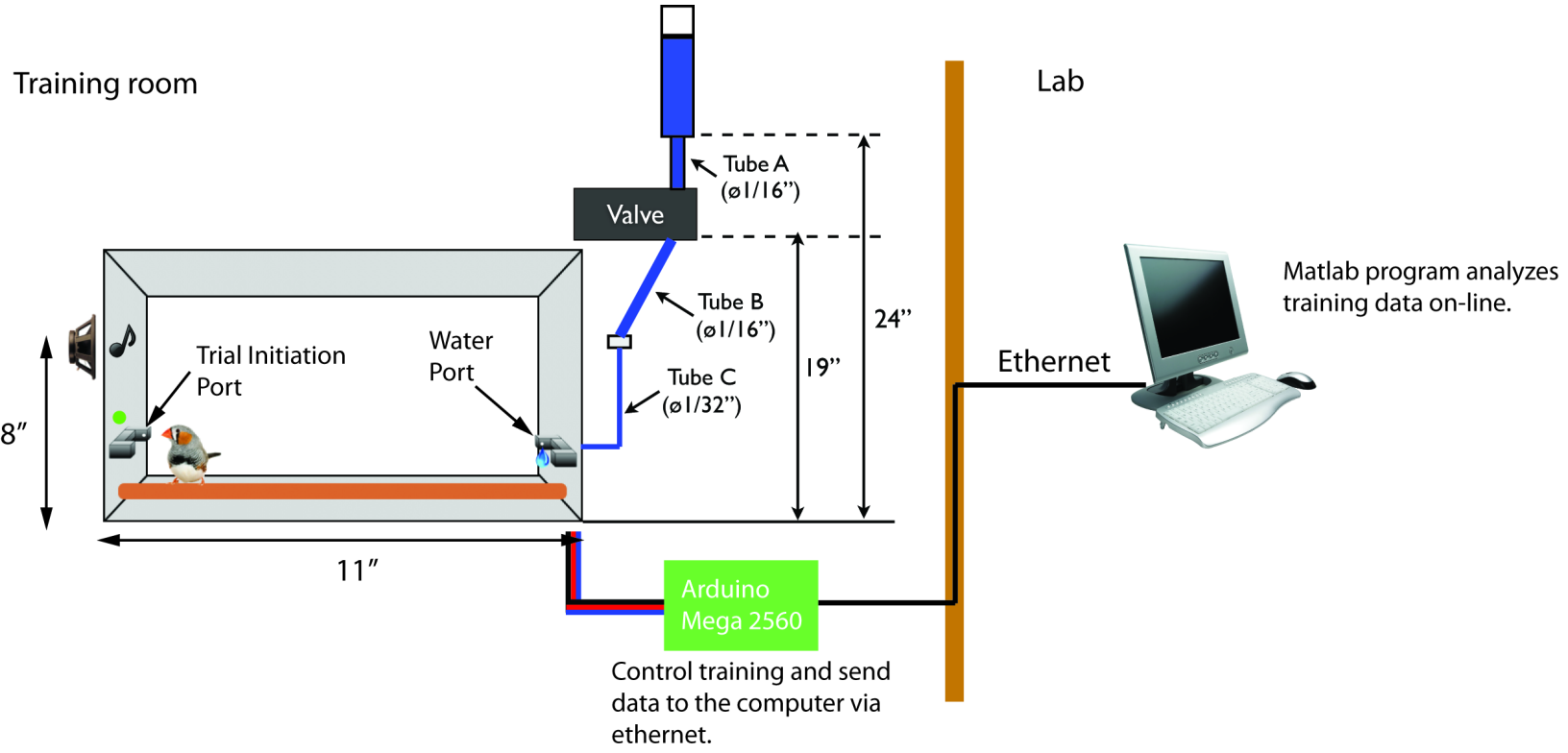

Primary
Secondary
50 ms

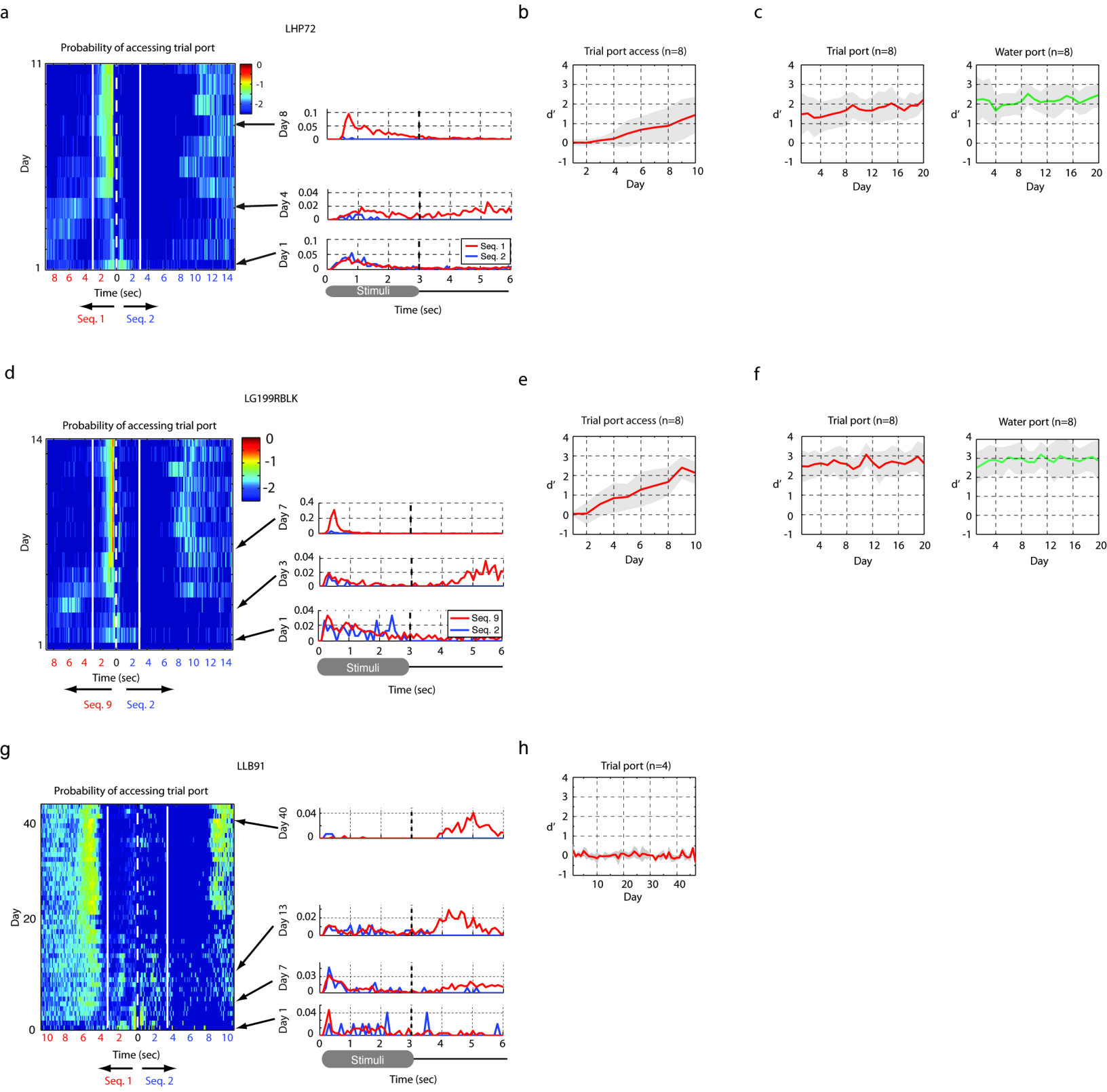

140
120
100
80
60
40
20
0
Spikes
5
10
15
20
25
Latency (ms)


ROC analysis (500ms)
True Positive Rate
False Positive Rate
0
0.2
0.4
0.6
0.8
1
Primary
Secondary


Primary
1x
2x
0.2
0.1
0
Probability
0
20
40
60
80
100
Number of spikes per second


Secondary
1x
2x
0.2
0.1
0
Probability
0
20
40
60
80
100
Number of spikes per second


Training room
Lab
Tube A
(ø1/16")
Valve
Tube B
(ø1/16")
Tube C
(ø1/32")
24"
19"
8"
11"
Trial Initiation
Port
Water
Port
Ethernet
Matlab program analyzes
training data on-line.
Arduino
Mega 2560
Control training and send
data to the computer via
ethernet.


a
LHP72
Probability of accessing trial port
Day
Time (sec)
Seq. 1
Seq. 2
Day 8
Day 4
Day 1
Stimuli
b
Trial port access (n=8)
d'
Day
c
Trial port (n=8)
Water port (n=8)
d
LG199RBLK
Probability of accessing trial port
Seq. 9
Seq. 2
e
Trial port access (n=8)
f
Trial port (n=8)
Water port (n=8)
g
LLB91
Probability of accessing trial port
Day 40
Day 13
Day 7
Day 1
h
Trial port (n=4)