Research Article

Auditory Forebrain Neurons Track Temporal Features of Time-Warped Natural Stimuli

ROSS K. MADDOX,^{1,2,3} KAMAL SEN,^{2,3} AND CYRUS P. BILLIMORIA^{2,3}

¹*Institute for Learning and Brain Sciences, University of Washington, 1715 NE Columbia Rd, Box 357988, Seattle, WA 98195, USA*

²*Hearing Research Center, Boston University, 44 Cummington Street, Boston, MA 02215, USA*

³*Center for Biodynamics, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215, USA*

Received: 9 January 2013; Accepted: 19 September 2013; Online publication: 16 October 2013

ABSTRACT

A fundamental challenge for sensory systems is to recognize natural stimuli despite stimulus variations. A compelling example occurs in speech, where the auditory system can recognize words spoken at a wide range of speeds. To date, there have been more computational models for time-warp invariance than experimental studies that investigate responses to time-warped stimuli at the neural level. Here, we address this problem in the model system of zebra finches anesthetized with urethane. In behavioral experiments, we found high discrimination accuracy well beyond the observed natural range of song variations. We artificially sped up or slowed down songs (preserving pitch) and recorded auditory responses from neurons in field L, the avian primary auditory cortex homolog. We found that field L neurons responded robustly to time-warped songs, tracking the temporal features of the stimuli over a broad range of warp factors. Time-warp invariance was not observed *per se*, but there was sufficient information in the neural responses to reliably classify which of two songs was presented. Furthermore, the average spike rate was close to constant over the range of time warps, contrary to recent modeling predictions. We discuss how this response pattern is surprising given current computational models of time-warp invariance and how such a response could be decoded downstream to achieve time-warp-invariant recognition of sounds.

Keywords: sensory systems, audition, perceptual invariance, time-warping, zebra finch, neurophysiology

INTRODUCTION

Understanding the neural basis for the discrimination and recognition of natural stimuli is an important goal in neuroscience. A challenging problem in this domain is perceptual invariance, i.e., the ability to classify natural stimuli despite changes in stimulus parameters. Perceptual invariance is of particular interest in the context of speech, where humans can recognize a word despite significant variations in its loudness (Shigenaga 1965; Denes and Pinson 1993; Versfeld and Dreschler 2002) or speed (Versfeld and Dreschler 2002). Auditory cortex is thought to play an important role in the recognition of natural sounds such as speech (Rauschecker 1998; Nelken 2004). Previous studies in the auditory cortex and its nonmammalian analogs, regions thought to be essential for processing natural sounds, have examined cortical discrimination of complex natural sounds in the absence of stimulus variations (Rauschecker 1998; Bar-Yosef et al. 2002; Nelken 2004; Wang et al. 2007) and have begun to investigate response invariance as it relates to sound intensity (Billimoria et al. 2008; Pienkowski and Eggermont 2011), but little is known about how the brain processes sounds to discriminate stimuli that have been sped up or slowed down. This issue—a phenomenon known as time-warp invariance—is a long-standing challenge in the field of machine learning (Tank and Hopfield 1987), and the neural mechanisms underlying

Correspondence to: Kamal Sen · Hearing Research Center · Boston University · 44 Cummington Street, Boston, MA 02215, USA. Telephone: (617) 353-5919; fax: (617) 353-6766; email: kamalsen@bu.edu

such invariance remain unknown. Neurons in the auditory periphery are typically very sensitive to timing cues (Carney and Yin 1988; Feng et al. 1991; Yost 2004; Moore 2006), making it difficult to explain such invariance based on peripheral mechanisms; it is therefore thought to emerge at central levels of processing. One study investigated cortical responses to time-warped calls in anesthetized marmosets, but no behavioral experiments were performed, leaving the relation of the neural results to perceptual invariance an open question (Wang et al. 1995). A few theoretical studies have investigated possible mechanisms for invariance to timing (Tank and Hopfield 1987; Hopfield and Brody 2000; Hopfield and Brody 2001; Barak and Tsodyks 2006; Gollisch 2008; Gutig and Sompolinsky 2009); however, relatively little remains known regarding the effect of these changes on cortical processing and discrimination.

Here, we examine the problem of time-warp invariance in songbirds (zebra finch), a model system which shows striking similarities to humans (Doupe and Kuhl 1999; Wang et al. 2010) and which has previously shown behavioral invariance to time warping (Nagel et al. 2010). We investigate neural responses in field L, a region homologous to mammalian primary auditory cortex, to variations introduced by time warping, where the stimulus is sped up or slowed down while preserving pitch.

MATERIALS AND METHODS

Electrophysiology

All procedures were done in accordance with the National Institutes of Health guidelines as approved by the Boston University Charles River Campus Institutional Animal Care and Use Committee. Recording methods followed the same procedures as in Billimoria et al. (2008). Anesthetized extracellular neural recordings were obtained from adult male zebra finches (*Taeniopygia guttata*). Birds were anesthetized with urethane, placed in a soft cloth restraining jacket, and positioned in a stereotactic assembly located within a double-walled sound-attenuated chamber (Industrial Acoustics). After performing a small craniotomy over the skull surface marked at 1.5 mm lateral and 1.2 mm anterior to the bifurcation point of the midsagittal sinus (Fortune and Margoliash 1992), tungsten electrodes (FHC, 1.2–4.0 M Ω) were advanced into the brain. None of the electrophysiological subjects were trained in the behavioral task, nor had they been exposed to the songs used as test stimuli. Invariant sites, after compensation, were distributed across all subregions of field L, using stereotactic coordinates and histological reconstruction. After recording, units were further isolated using a spike-sorting algorithm based on principal components analysis and cross-correlations of spike waveform shape. Using these spike-clustering methods, single units as

well as some multiunit clusters (which could not be separated into single units) were extracted and analyzed. A previous study from our lab shows similar discrimination performances for single and multiunit clusters (Wang et al. 2007).

Stimuli

The probing stimulus for seeking recording sites was a single representative zebra finch song at 75 dB sound pressure level (Radio Shack sound level meter #33-2055 with C-weighting). After identification of an auditory unit (identified using a paired *t* test comparing stimulus firing rate and background firing rate, $p \leq 0.05$), the full stimulus ensemble was presented. Time-warped stimuli were generated using two zebra finch songs as templates. They were each time-warped while preserving pitch to warp factors of 0.5 $\times$ (half speed), 0.75 $\times$, 0.95 $\times$, 0.99 $\times$, 1.01 $\times$, 1.05 $\times$, 1.5 $\times$, and 2 $\times$ (double speed) using a phase vocoding algorithm (Flanagan and Golden 1966; Ellis 2002). All songs were bandpass-filtered between 250 Hz and 8 kHz.

Data analysis

We used the van Rossum spike distance metric (van Rossum 2001) in conjunction with a template-matching scheme to calculate neural discrimination performance (Machens et al. 2003). First, to determine the dissimilarity between a pair of spike trains, both trains (represented as a string of zeros and ones) were convolved with exponential kernel $h(t)$ with time constant τ .

$$h(t) = u(t)e^{-t/\tau}$$

where $u(t)$ is the unit step function. The van Rossum distance D was then given by:

$$D = \frac{1}{\tau} \int_0^{T_0} (f(t)*h(t) - g(t)*h(t))^2 dt,$$

where $f(t)$ and $g(t)$ are the spike trains being compared, T_0 is the duration of the spike trains to use, and the star operator denotes convolution. To avoid using length of spike train as a cue, T_0 was taken here to be the length of the shortest stimulus, 830 ms. The time constant τ was varied from 1 ms, which emphasizes individual spike timing acting as a coincidence detector, to 1 s, which emphasizes the average spike rate and acts similarly to a rate discriminator, with time constants in between acting as a combination of timing and rate discriminator.

All pairwise combinations of spike train distances were then calculated for a site and stored in a distance matrix. The percent correct performance was calcu-

lated by finding the minimum distance between a selected spike train and the template spike trains and assigning the spike train to that template; the templates were the responses to the songs at 1× time warp. Using this method, a percent correct value was calculated for each time-warp value for all values of τ . For a given site, the optimal τ was chosen by computing the area under the performance vs. warp-factor curves (e.g., in Fig. 1C) for various values of τ and choosing the one that yielded the highest value.

To compute the invariance of a site's performance curve, a modified Riemann sum of the area between

the nontemplate percent correct values and the template percent correct value was calculated as:

$$I = 1 - \frac{\sum_{i=1}^{T-1} |y_T - y_i| (x_{i+1} - x_i) + \sum_{i=T+1}^N |y_T - y_i| (x_i - x_{i-1})}{(1 - \theta)(x_N - x_1)},$$

where y is percent correct, x is the warp factor, θ is chance performance (here, 50 %), N is the total number of versions of each song, and T is the index of the template version. The invariance value was normalized by the area above chance so that a site that had 100 % template performance but performed

FIG. 1. Neural discrimination of artificially time-warped songs compared to bird behavioral performance. **A** Spectrograms of five time-warped versions of the same song (only five of the nine tested speeds shown), with darker shades indicating higher power as a function of time and frequency. **B** Spike rasters in response to the stimuli in **A** from one recording site. **C** Performance curves for neural sites in response to the presentation of time-warped stimuli (color-coded with hotter colors for higher invariance), with chance line in dashed black. **D** A histogram of the invariance values for sites, with a median value of 0.35. **E** Behavioral performance of four birds in classifying the time-warped versions of two trained stimuli. Individual performance curves are gray, and the average curve ± 1 SEM is in black (chance performance is 50 % indicated by the dashed line). Birds were trained on normal-speed songs, and no feedback was given for time-warped stimuli. The invariance of the behavioral data (0.66) is denoted with a dotted line in **D**.

at chance at all other time warps had an invariance of 0, while a site with a completely flat percent correct curve had an invariance of 1.

A site with no information about which song was presented would perform at chance for all time warps. Such a site would have an invariance value of 1. To eliminate sites whose performance was low (causing them to appear spuriously invariant), an inclusion criterion was used: the nonwarped template performance must meet or exceed halfway between chance and perfect performance, or 75 % (the performance on warped stimuli was not considered when excluding sites). Of the excluded sites, all neurons were deemed auditory (though some weakly) using a *t* test ($\alpha \leq 0.05$) between the number of spikes in the 500 ms preceding and following stimulus onset. One of these sites had performance of 71 % and the rest were approximately evenly distributed between 49 % and 64 %. While the best of these sites had some discriminatory power (indicating some timing reliability from trial-to-trial), the stringent 75 % cutoff was needed to properly assess time-warp invariance, the major focus of this manuscript. As such, the underperforming neural sites are henceforth not considered.

Behavioral experiments

Four male zebra finches were trained to perform a recognition and classification task using a food reward-based system. Twice a day, 5 days a week, for 3–4 weeks, subjects were placed in a testing apparatus within a double-walled sound-attenuating chamber and subjected to a series of training phases that introduced them to the two songs they were to categorize. Each bird was first trained to associate one song with the left key by being presented only that song, pecking the key to respond and receiving a small food reward (or a lights-out punishment for pecking the incorrect key). After associating the first song with its key, the process was repeated for the second song and the right key, presenting only that song until the bird had associated it with the right key. Once the bird had associated both songs with their respective keys, the two songs were presented randomly interleaved, forcing the bird to decide for each trial which key to press and being rewarded or punished accordingly. When performance plateaued and was above 80 % (normal plateau was approximately 90 %), the bird was advanced to the experimental phase.

During the experiment phase, 80 % of trials were presentations of the unaltered training stimuli (warp factor of 1 $\times$), while the remaining 20 % of trials consisted of the eight different time-warped versions of those template stimuli. Presentations of the trained and time-warped stimuli were randomly interleaved. For trials in which the trained stimuli were presented, reward was based on the correctness of the response;

but for trials in which time-warped versions were presented, food was rewarded regardless of which key the bird pecked to avoid learning of the time-warped stimuli. Data were collected until the bird responded to at least 100 blocks of stimulus presentation, with each block consisting of 16 time-warped versions (i.e., one trial of each of the two songs at every nonunity warp factor) of the stimulus and 64 presentations of the trained versions.

Time-warp compensation

The time-warp compensation consisted of two components. The first was length compensation that aligned the time axes of the spike trains, effectively scaling them by the time-warp factor of their corresponding stimuli. This stretched the shorter trains and compressed the longer ones so that activity in the spike trains that were consistently timed with stimulus features lined up across time warps. Because the neural latency between stimulus onset and the start of response spike trains was constant across all stimuli, this latency was removed prior to length compensation to ensure poststretching response onset alignment. We then used the optimal value τ for each site to determine the neural discrimination percent correct. The median best τ was 10 ms (range 6–25 ms) and the median neural latency removed was 25 ms (16–84 ms).

The second method of compensation was an adjustment of the time constant of the exponential kernel used in calculating the distance between two spike trains. Since the average spike rates in the time interval corresponding to the stimulus presentation remained mostly constant across time warps (see “Results” section), the spike trains at the 0.5 $\times$ time warp had approximately twice as many spikes as the 1 $\times$ trains (since the stimulus interval was twice as long), and likewise, the spike trains at the 2 $\times$ time warp had approximately half as many spikes as the 1 $\times$ spike trains. To compensate for these differences in spike count, we employed a τ compensation that scaled the value of the exponential kernel of the distance metric by the time warp of the stimulus that produced the spike train. For the faster songs (with fewer spikes), this made the kernels longer and hence have a proportionally higher area, and the opposite effect occurred for slower songs (with more spikes). In order to make comparisons between all pairs of time warps, the normalization term in the van Rossum metric ($1/\tau$) was held fixed.

RESULTS

Time-warp experiments

To investigate the neural response to artificially time-warped versions of two songs, we used phase vocoding to stretch and compress the songs while preserving

their pitch over a range of half speed ($0.5\times$) to double speed ($2\times$; Fig. 1A). We then recorded 39 field L sites (after clustering) in five birds and found that, in some sites, neural firing remains time-locked to the stimulus for all warp factors (Fig. 1B), but appears compressed or stretched relative to the firing observed for the template ($1\times$) time warp. This caused neural discrimination performance to suffer at time warps of more than a few percent (Fig. 1C), which, in turn, resulted in low values of invariance for all sites (Fig. 1D).

Behavioral experiments with four birds showed that they could generalize over a range of time warps (Fig. 1E). Even at the most extreme compression ($2\times$), performance was significantly better than chance ($T_3=4.9$, $P=0.0083$, one-sided t test). Performance at all other warps was at least as significantly above chance, with the exception of the maximal dilation ($0.5\times$), which trended ($T_3=2.14$, $P=0.061$).

Time-warping causes the evoked spikes not to line up in time (Fig. 2A), making it difficult to compare the responses to the fine structure of responses at different warp factors. To compensate for this, we stretched or compressed spike trains evoked in response to compressed or stretched songs, respectively, to match the nonwarped length (Fig. 2B). This realignment produced firing at points in time that corresponded to equivalent features in the time-warped stimuli. Figure 2C shows an example of the response after such alignment. After this compensation, the overall shapes of the peristimulus time histograms (PSTHs) for this site were very consistent across the time-warp factors, indicating responsiveness to the spectrotemporal features of the stimulus regardless of the speed of the song. To get a general sense of how well they matched, we computed the grand correlations

(appending both songs and responses from all neurons) of the PSTHs at each nonunity time warp with the template unity PSTH. The correlations were high: 0.89, 0.91, 0.92, 0.92, 0.92, 0.92, 0.88, and 0.82, from $0.5\times$ to $2\times$ (excluding unity), indicating strong similarity of the compensated PSTHs at all of the time warps.

Average spike rates in the stimulus presentation interval were not proportional to time warp (Fig. 3). While there were slight increases in spike rate for most sites with increasing warp factor, they were closer to constant than to scaling directly with spike rate. This is illustrated by the slope of the dashed line, which shows the spike rate vs. warp factor of a hypothetical unit that exhibits this proportional scaling. The slopes of all sites are less than that of the hypothetical unit.

A modified classification scheme for time-warped stimuli

We previously used a classification scheme for song discrimination based on the van Rossum spike distance metric and a nearest neighbor template-matching scheme. However, such a classification scheme failed with time-warped stimuli because even though the recording sites responded to the same stretched or compressed features in the time-warped stimuli, time-warping caused the evoked spikes not to line up in time, resulting in large van Rossum distances to the template ($1\times$ speed) spike trains. To address this, we modified the classification scheme by rescaling the time axis of the responses by the respective warp factor, effectively realigning the responses. Although length compensation helped restore the temporal alignment of neural responses, temporally rescaling the spike trains caused a corresponding a

FIG. 2. Time realignment of spike trains evoked in response to songs across a range of time-warps. **A** Spike rasters from one site in response to stimuli at five time-warp factors (same site as Fig. 1B). **B** The same rasters as in **A**, but with time axes adjusted according to the time-warp factor of the stimulus (note the different time axes). **C** PSTHs

on the same time axes as in **B**. A 50-ms rectangular smoothing window was used for the nonwarped PSTH, and the kernel length was scaled inversely by the time warp to create the same level of detail for each PSTH. The solid black lines show the average spike rates over the duration of the stimulus.

FIG. 3. Average spike rates do not scale with warp factor. Each recording site's average spike rate within the stimulus presentation interval is plotted against the time-warp factor (gray lines, black dots). A hypothetical unit (black dashed line) whose spike rate increases at the same rate as warp factor is shown for comparison. All units' slopes were less than the hypothetical unit's slope on the log-log plot.

change in the effective firing rates. To address this second issue, we modified the classification scheme by introducing a second form of compensation to correct for the change in the effective firing rate. This was done by scaling the time constant of the filter (τ parameter), in the van Rossum distance calculations proportionally to the time warp of the corresponding stimulus, which compensated for the spike density mismatch via a proportional increase (for speeds faster than template) or decrease (for speeds slower than template) in convolved spike area.

These two simple modifications to the classification scheme (length compensation followed by τ compensation), led to a dramatic improvement in the neural performance across sites ($T_{22}=4.72$, $P=0.000052$, one-sided paired t test). While uncompensated data yielded high van Rossum spike train distances within the same song across time warps, shown for an example site in Figure 4A, length and τ compensation yielded low intrasong distances across all time warps (Fig. 4B). Multidimensional scaling (Fig. 4C, D) further illustrates that song separation improved between the uncompensated and fully compensated regimes in the same recording site. The significant overlap of the data points in the uncompensated case, with little to no clustering in the song identity domain, was fixed by compensation, allowing significant segregation based on song identity. Neural classification performances improved across all the 23 recording sites that met the inclusion criterion (Fig. 4E, F). Due to this compensation, the median performance value across recording sites improved from 53.6 % (range 47.0–67.4 %) to 87.3 % (range 56.0–100 %). The invariance of every recording site also improved, some substantially (Fig. 4G, H); the uncompensated median invariance of 0.32 (range from 0.12 to 0.75) improved to a median invariance of 0.86 (range from 0.45 to 1.00).

FIG. 4. Compensation yields time-warp-invariant discrimination performance using neural responses. The two columns demonstrate the results of using uncompensated (left column) and compensated (right column) algorithms in distance calculations. **A, B** The van Rossum spike distance matrices for one recording site using the two algorithms, with cooler colors representing higher similarity. **C, D** The multidimensional scaling scatter plots of these data, with song identity designated by marker shape, and time warp by rainbow order from red (2.0 $\times$) to blue (0.5 $\times$). **E, F** Discrimination performance curves for different recording sites (color coded for increasing invariance from blue to red); the example unit is highlighted by the dotted black curve, and chance (50 %) is the dashed black line. **(G, H)** Histograms of the invariance values, where median invariance increased from 0.32 (range of 0.12–0.75) to 0.86 (range of 0.45–1.0) with compensation.

DISCUSSION

Previous studies have implicated the primary auditory cortex and its analogs in the processing of natural sounds (Rauschecker 1998; Bar-Yosef et al. 2002; Nelken 2004; Wang et al. 2007; Narayan et al. 2007), but relatively few have studied the impact of speed

variations on cortical processing. One previous study found a correlation between speech comprehension of compressed speech waveforms and cortical activity (Ahissar et al. 2001), but the measure of cortical activity used was a magnetoencephalographic signal with a generic head model, making it difficult to infer the underlying single neuron responses. Another study found, similar to the present avian results, that neurons in marmoset primary auditory cortex track the temporal repetitions in time-warped twitter calls (Wang et al. 1995). Perceptual invariance was not discussed in that paper, and as such, it did not include behavioral experiments.

The natural range of timing variations in adult zebra finches, the species studied here, is relatively small, suggesting that neural discrimination performance of natural variants of song is likely to be high. Our results using natural variants of song confirmed this (data not shown). However, the small variations in adult song raise the question of whether birds would be capable of discriminating time-warped stimuli over a larger range. In behavioral experiments, we found that birds were able to discriminate time-warped stimuli up to a factor of two significantly over chance level. These data are consistent with a recent behavioral study in female birds (Nagel et al. 2010). This behavioral capacity motivated the search for an underlying neural basis for time-warp invariance.

Due to training restrictions, in both this study and a previous behavioral study (Nagel et al. 2010), only two songs were used. It is possible subjects relied on a single feature in each of the songs, rather than the entire song, to make their choice. Nevertheless, the neural results reveal that the entirety of each song was well coded. We used the same pairs of songs for all subjects here, which leads to a possibility of pseudoreplication. However, in the previous study (Nagel et al. 2010), a different set of songs was used and behavioral results were strikingly consistent with those here (for the subset of time warps tested).

Our results indicate the ability of neural responses at the cortical level to robustly follow changes in sounds over a broad range of speeds, as indicated by the similarity of time-rescaled PSTHs for the different warp factors. This result is consistent with the previous study employing MEG signals, but provides a more detailed picture at the neural level, revealing an essentially perfect rescaling of the PSTH for some neurons.

The change in song speed presents an immediate problem for previously employed methods for neural discrimination of songs that rely on metrics, such as the van Rossum metric, that are highly sensitive to timing. With time-warped stimuli, the timing of features in the sound to be recognized changes dramatically with speed. Not surprisingly, the previous neural discrimination method failed when tested on

time-warped songs. However, the similarity of the time-rescaled PSTHs for the different time warps motivated modifications to the van Rossum metric-based classification. We found that two simple modifications to this metric were able to compensate for changes in stimulus speed: (1) rescaling the spike trains according to the warp factor of the stimulus, then (2) scaling the time constant of the kernel used in the spike distance metric by the warp factor. The compensation we use here is for analytical purposes—neural computations with similar outcomes but likely different mechanisms may be carried out downstream, presumably in higher level songbird auditory areas such as caudal mesopallium, a candidate area for the recognition of conspecific songs (Gentner and Margoliash 2003). It cannot be concluded from this study alone if the preserved PSTH shape and nearly constant average spike rate across time warps are emergent properties of field L or if more peripheral neurons would show the same properties. The former is likely, as precise timing in responses is also found in earlier stages of auditory processing. However, the preservation of average spike rate across time warps, a property of field L neurons discovered in this study, needs to be tested directly in earlier stages.

The most recent computational model of time-warp invariance (Gutig and Sompolinsky 2009) assumed afferent neurons whose spike rates scaled according to warp factor. Specifically, neurons tracked the amplitude envelope of spectral bands with an array of center frequencies and spiked either at an upward (onset) or downward (offset) threshold crossing. Due to the nearly consistent spike rate across time warps observed here, it seems that field L neurons do not respond in this way. While we did not observe time-invariant responses in field L, the results of this study will improve future models by constraining the properties of input neurons.

Although the compensation scheme suggests computational steps necessary for robust time warp invariance, it does not specify how this could be achieved in a biologically plausible way. Many modeling studies have theorized potential mechanisms (Tank and Hopfield 1987; Hopfield and Brody 2000; Hopfield and Brody 2001; Barak and Tsodyks 2006; Gollisch 2008; Gutig and Sompolinsky 2009). Previously, we have proposed biologically plausible computational model for song recognition (Larson et al. 2009, 2010). One of these models (Larson et al. 2009) uses a van Rossum-like scheme to compare an incoming spike train to a large set of memorized ones and chooses the most likely song among them. We found that a simple extension to the model by incorporating multiple memory banks at different speeds enables time-warp-invariant recognition on the experimental data reported here (data not shown), suggesting that neural responses in field L to time-warped stimuli could form the basis for time-warp-

invariant recognition of natural sounds. Future experiments may reveal if our findings are a general property shared by other auditory systems.

ACKNOWLEDGMENTS

The authors wish to thank Elizabeth M McClaine and Gilberto David Graña for help with behavioral testing and histology.

REFERENCES

- AHISSAR E, NAGARAJAN S, AHISSAR M ET AL (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci U S A* 98:13367–13372. doi:10.1073/pnas.201400998
- BARAK O, TSODYKS M (2006) Recognition by variance: learning rules for spatiotemporal patterns. *Neural Comput* 18:2343–2358. doi:10.1162/neco.2006.18.10.2343
- BILLIMORIA CP, KRAUS BJ, NARAYAN R ET AL (2008) Invariance and sensitivity to intensity in neural discrimination of natural sounds. *J Neurosci* 28:6304–6308. doi:10.1523/JNEUROSCI.0961-08.2008
- CARNEY LH, YIN TC (1988) Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model. *J Neurophysiol* 60:1653–1677
- DENES PB, PINSON EN (1993) The speech chain. Macmillan
- DOUPE AJ, KUHL PK (1999) Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22:567–631. doi:10.1146/annurev.neuro.22.1.567
- ELLIS DPW (2002) A phase vocoder in MATLAB. <http://www.ee.columbia.edu/~textasciitildedpwe/resources/matlab/pvoc/>
- FENG AS, HALL JC, SIDDIQUE S (1991) Coding of temporal parameters of complex sounds by frog auditory nerve fibers. *J Neurophysiol* 65:424–445
- FLANAGAN JL, GOLDEN RM (1966) Phase Vocoder. *Bell System Tech J* 45:1493–1509
- FORTUNE ES, MARGOLIASH D (1992) Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (*Taenopygia guttata*). *J Comp Neurol* 325:388–404
- GENTNER TQ, MARGOLIASH D (2003) Neuronal populations and single cells representing learned auditory objects. *Nature* 424:669–674
- GOLLISCH T (2008) Time-warp invariant pattern detection with bursting neurons. *New J Phys* 10:015012
- GUTIG R, SOMPOLINSKY H (2009) Time-warp-invariant neuronal processing. *PLoS Biol* 7:e1000141. doi:10.1371/journal.pbio.1000141
- HOPFIELD JJ, BRODY CD (2000) What is a moment? “Cortical” sensory integration over a brief interval. *Proc Natl Acad Sci U S A* 97:13919–13924
- HOPFIELD JJ, BRODY CD (2001) What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. *Proc Natl Acad Sci U S A* 98:1282–1287
- LARSON E, BILLIMORIA CP, SEN K (2009) A biologically plausible computational model for auditory object recognition. *J Neurophysiol* 101:323–331. doi:10.1152/jn.90664.2008
- LARSON E, PERRONE BP, SEN K, BILLIMORIA CP (2010) A robust and biologically plausible spike pattern recognition network. *J Neurosci* 30:15566–15572. doi:10.1523/JNEUROSCI.3672-10.2010
- MACHENS CK, SCHUTZE H, FRANZ A ET AL (2003) Single auditory neurons rapidly discriminate conspecific communication signals. *Nat Neurosci* 6:341–342
- MOORE BC (2006) An introduction to the psychology of hearing. Elsevier, Oxford, UK
- NAGEL KI, MCLENDON HM, DOUPE AJ (2010) Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *J Neurophysiol* 104:1426–1437. doi:10.1152/jn.00028.2010
- NARAYAN R, BEST V, OZMERAL E ET AL (2007) Cortical interference effects in the cocktail party problem. *Nat Neurosci* 10:1601–1607. doi:10.1038/nn2009
- NELKEN I (2004) Processing of complex stimuli and natural scenes in the auditory cortex. *Curr Opin Neurobiol* 14:474–480
- PIENKOWSKI M, EGGERMONT JJ (2011) Sound frequency representation in primary auditory cortex is level tolerant for moderately loud, complex sounds. *J Neurophysiol* 106:1016–1027. doi:10.1152/jn.00291.2011
- RAUSCHHECKER JP (1998) Cortical control of the thalamus: top-down processing and plasticity. *Nat Neurosci* 1:179–180. doi:10.1038/625
- SHIGENAGA S (1965) The constancy of loudness and of acoustic distance. *Bull Faculty Lit Kyushu Univ* 9:289–333
- TANK DW, HOPFIELD JJ (1987) Neural computation by concentrating information in time. *Proc Natl Acad Sci U S A* 84:1896–1900
- VERSFELD NJ, DRESCHLER WA (2002) The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners. *J Acoust Soc Am* 111:401–408. doi:10.1121/1.1426376
- WANG L, NARAYAN R, GRANA G ET AL (2007) Cortical discrimination of complex natural stimuli: can single neurons match behavior? *J Neurosci* 27:582–589. doi:10.1523/JNEUROSCI.3699-06.2007
- WANG X, MERZENICH MM, BEITEL R, SCHREINER CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 74:2685–2706
- WANG Y, BRZOZOWSKA-PRECHTL A, KARTEN HJ (2010) Laminar and columnar auditory cortex in avian brain. *Proc Natl Acad Sci U S A* 107:12676–12681. doi:10.1073/pnas.1006645107
- YOST WA (2004) Determining an Auditory Scene. In: Gazzaniga MS (ed) *The cognitive neurosciences*. MIT Press, Cambridge, MA, pp 385–410