Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Neurobiology

Computational models of auditory perception from feature extraction to stream segregation and behavior

James Rankin[1] and John Rinzel[2,3]

Audition is by nature dynamic, from brainstem processing on sub-millisecond time scales, to segregating and tracking sound sources with changing features, to the pleasure of listening to music and the satisfaction of getting the beat. We review recent advances from computational models of sound localization, of auditory stream segregation and of beat perception/generation. A wealth of behavioral, electrophysiological and imaging studies shed light on these processes, typically with synthesized sounds having regular temporal structure. Computational models integrate knowledge from different experimental fields and at different levels of description. We advocate a neuromechanistic modeling approach that incorporates knowledge of the auditory system from various fields, that utilizes plausible neural mechanisms, and that bridges our understanding across disciplines.

Addresses
[1] College of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Rd, Exeter EX4 4QF, UK
[2] Center for Neural Science, New York University, 4 Washington Place, 10003 New York, NY, United States
[3] Courant Institute of Mathematical Sciences, New York University, 251 Mercer St, 10012 New York, NY, United States

Corresponding author: Rankin, James (james.rankin@gmail.com)

Current Opinion in Neurobiology 2019, **58**:xx-yy

This review comes from a themed issue on **Computational neuroscience**

Edited by **Máté Lengyel** and **Brent Doiron**

https://doi.org/10.1016/j.conb.2019.06.009

Introduction

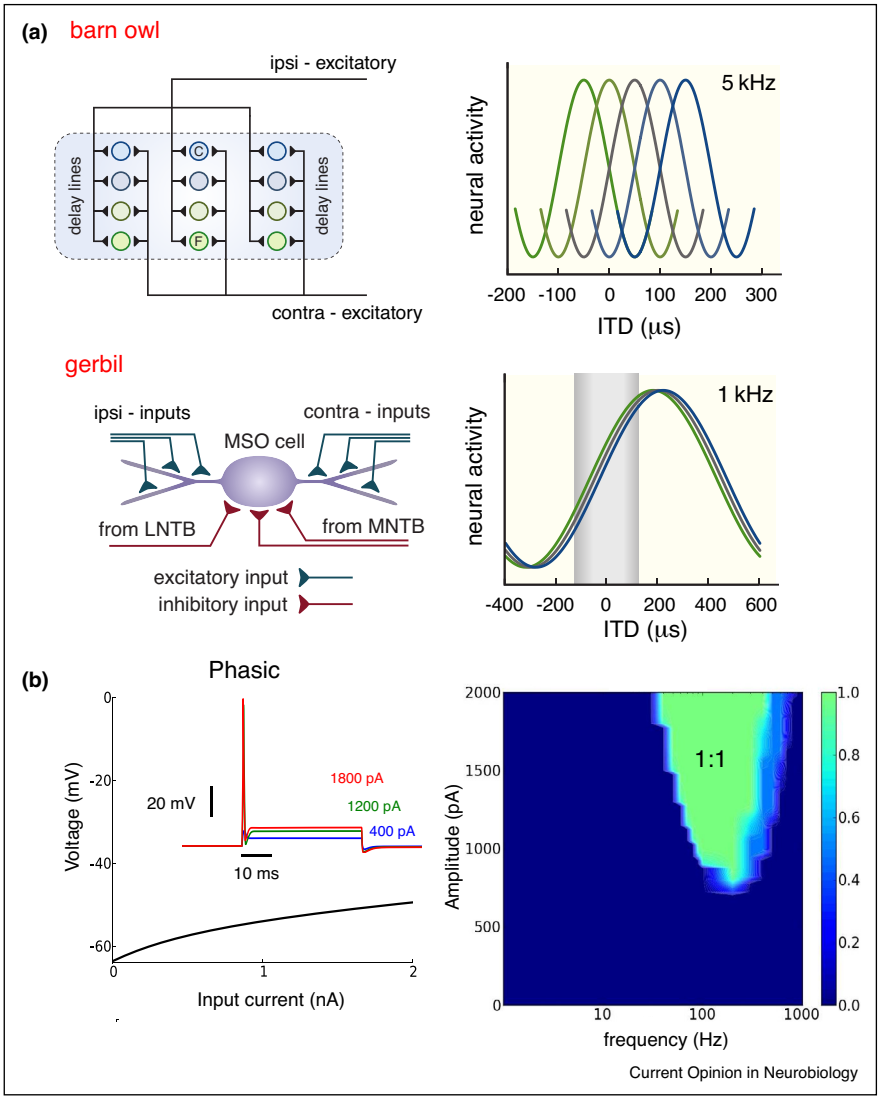
In a crowded bar, people chatter away and glasses clink, but from the corner stage we pick out the repetitive snap of a snare drum and start to tap along. All this relies on the extraction of multiple auditory features from a rich soundscape. The separation of features, such as pitch and location, along with timing cues, allows for the segregation of individual streams like a voice or melody. Once identified, a stream can be predicted in order to drive motor behavior like tapping along to the beat. This review focuses on computational modeling, especially neuromechanistic approaches, of the dynamics of auditory processing, that is the representation and perception of how we hear the world. Among recent research developments our review highlights: the biophysics underlying neuronal computation with exceptional temporal precision — on the order of tens of microseconds — for sound localization, the emergence of stream segregation and subsequent perceptual bistability for ambiguous sounds and the continuation of a learned beat after stimulus offset by a neural oscillator. The relatively mature topic of sound localization, having benefited from longstanding interplay between modeling and experiments, is presented first. We propose that the less developed fields of stream segregation and beat perception will profit from a similar interplay albeit with new challenges arising from the experimental constraints in studying higher-level, cognitive processes.

Sound localization

Localization of a sound source involves detecting interaural time differences, ITD, for low frequency sounds (say, <1.5 kHz) or interaural level differences, ILD, for high frequency sounds. These neuronal computations are performed early in the auditory pathway where inputs from the two ears converge: in mammals, the superior olivary complex, SO. According to the 'duplex' theory, the medial portion, MSO, computes ITD while the lateral portion, LSO, handles ILD [1,2]. Theoretical research, including biophysically based and neural coding models, has aligned closely with quantitative neurophysiological experiments (*in vitro* and *in vivo* [3•]) in reaching substantial mechanistic understanding, whilst several challenges remain.

Behavioral and neuronal-MSO ITD tuning curves show discriminability with an astonishing temporal resolution, tens of microseconds. Various biophysical specializations underlie this extraordinary and essentially single neuron computation: sub-millisecond membrane time constants, fast subthreshold nonlinear conductance mechanisms underlying onset firing, strong phase-locking, and brief synaptic conductances segregated to bipolar dendrites [3•]. An MSO neuron's onset responsiveness supports coincidence detection. An MSO neuron behaves as a differentiator, responding only to fast change, as with nearly coincident inputs, but not to slow inputs [4,5] (Figure 1b). Spiking follows feed-forward summation of relatively few inputs per dendrite [6]; spikes are generated downstream of the soma with almost no back propagation [7,8,9]. Dendritic cable modeling demonstrates why single-sided inputs rarely fire a cell [10].

Figure 1



Physiology, tuning and onset response for sound localization. **(a)** Schematic of physiological architecture (left) for the neuronal computation of ITD tuning (right); comparison of bird (barn owl, upper) and mammal (gerbil, lower); adapted from [13]. The barn owl exemplifies the Jeffress conceptual model: labeled delay lines, one set from the ipsilateral ear and one from the contralateral ear, providing excitatory input to an array of coincidence-detector neurons. The neurons along the array with the highest firing rate correspond to the ITD. The collection of tuning curves span the physiological range, as determined by head size. In the gerbil (lower A panels) MSO neurons receive excitatory and inhibitory input from ipsilateral and contralateral ears; an ITD tuning curve has maximum firing for ITD that lies outside the physiological range (shaded) [14•]. The ITD computation is thought to involve the difference between the oppositely sloping tuning curves in the two brain hemispheres, the 'two-channel' hypothesis of [1,15]. Various models have been proposed to account for 'slope-based' encoding: precise and fast timing of the contralateral inhibition to disfavor firing for ITD <0 [14•]; difference in EPSP slopes for ipsi/contra inputs [16]; and asymmetrical emergence of axon from soma/dendrite [17]. **(b)** MSO principal neurons fire phasically, only to fast rising inputs such as step current (left). They do not fire in response to slowly varying input as shown here (right) with a model [18]: for rectified sinusoidal current input the model fires once per cycle (green) for a stimulus frequency range (approximately 100–350 Hz, for adequate strength input); no firing occurs for lower frequency (dark blue), phase-locking but with cycle-skipping may occur for higher frequency (light blue). *In vitro* experiments and biophysically based modeling together reveal dynamic, but fast, subthreshold mechanisms that preclude spike generation if depolarizing input is too slow. To get a spike, depolarization should be fast enough to out-race the activation of a low-voltage-threshold potassium current, I_{klt}, [4,18] and the substantial and fast inactivation of the sodium current [19]. If the conductance of I_{klt} is frozen at its resting level the model converts to tonic firing, Type 3 to Type 2 excitability, while phase-locking and ITD-sensitivity suffer [5].

Recent findings help to focus further questions about the role for dendrites. Since dendritic and synaptic conductances counteract temporal broadening [11] and provide somatic EPSP amplitude equalization [12], we might feel satisfied that single soma-dendrite compartment models succeed in addressing some questions about MSO

processing. On the other hand, further insights are likely if we understand more about the spatio-temporal patterning of inputs to MSO dendrites [12].

The conceptual Jeffress model [20] for localization applies to barn owl anatomy and physiology [13] (Figure 1a) but not directly to mammals [1,3•]. Both excitatory input and fast temporally precise inhibitory input shapes ITD-tuning in gerbils [14•] (although just how fast to avoid temporal summation is under question [21]). Further, and unexpected according to Jeffress, the ITD for maximal firing can lie beyond the physiological range (Figure 1a), as determined by head-size. A coding theory approach, involving Fisher information, offers an explanation that optimal ITD estimation is based on tuning function slope (not peak) [22,23]. With regard to the delay lines of Jeffress, anatomical evidence of explicit axonal delays (Figure 1a) is lacking for the gerbil and alternate explanations remain under consideration, including cochlear disparities, mismatch of inputs from cochlea to MSO, preceding inputs influencing spike threshold, and dependence on stimulus properties [24,3•,2].

In the classical view LSO performs rate-based encoding rather than timing-based encoding as in MSO [1,2]. Yet recent studies have found timing-based biophysical mechanisms, namely some LSO neurons are not just simple integrators but have resonance properties [25,26] with frequency preferences comparable to those reported in MSO [27] and some LSO neurons show onset behavior and/or ITD sensitivity [24,25,2]. The classical lines of the 'duplex' theory continue to blur and hypotheses are being proposed about how ITD information from MSO and LSO may be combined for sound localization [15].

Auditory streaming, ambiguity and bistability

How does the brain extract auditory objects and track their cues and features? This so-called 'cocktail party problem' involves isolating separate voices in a dynamic environment and attending to one speaker. Initially we hear an integrated mixture of sound/voices but then our auditory system distinguishes separate streams (Auditory Scene Analysis; recent reviews [28,29]). A valued paradigm from Van Noorden [30] for studying auditory streaming involves segregating two interleaved sequences of A tones and B tones, separable by a perceived difference in pure tone frequency and timing (Figure 2a). Initially heard in one stream (integrated, Figure 2b), the probability of hearing two streams (segregated, Figure 2b) gradually builds up over several to tens of seconds. Build-up occurs more rapidly with a large difference in tone frequency (DF) between A and B (Figure 2h) and at faster presentation rates. The first perceptual switch, typically from integrated to segregated, is followed by persistent alternations between the two interpretations [31] (Figure 2c). Imaging approaches have shed light on, for example, the network of brain areas involved in streaming with fMRI [32], the effects of attention on neural representations of streams with MEG [33] or magnetic resonance spectroscopy [34], and the role of oscillations in encoding streams with EEG [35] (comprehensive review: [29]).

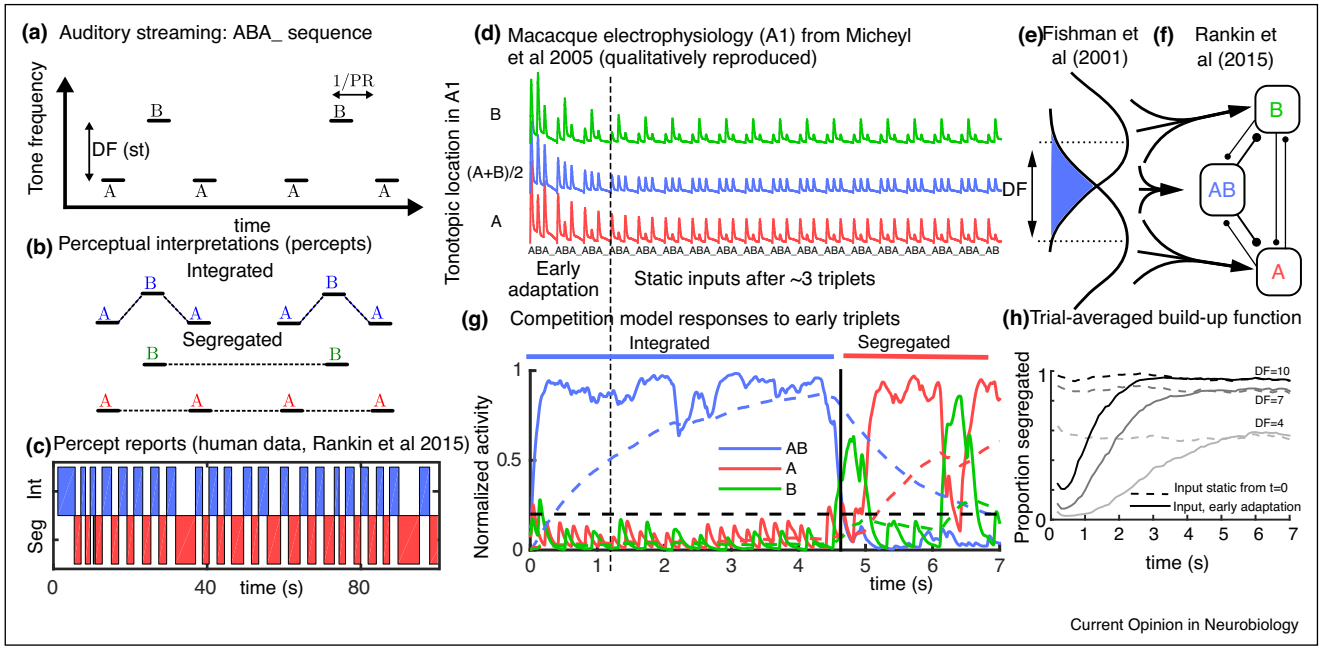
Most existing computational models of auditory streaming (recent review: Szabó et al. [39•]) focused on reproducing the dependence of perceptual bias, and/or the dynamics of build-up, on DF and presentation rate. Models of build-up are posed in a range of frameworks: signal processing [40], temporal coherence [41], tonotopic organization [42] or neural oscillations [43]. A complete theoretical framework for streaming should account for build-up and later alternations (build-up converges to the long-term probability of bistable alternations (Figure 2h)).

Several recent models focused on post-build-up alternations (auditory bistability) with competition dynamics [44,36] or probabilistic switching schemes [45,46]. The statistical properties of percept durations share features across a range of bistable perceptual phenomena: typically described by log-normal distributions [31,47,48]. The statistical model of [45], based on an alternating renewal process, reproduces the main features of build-up and later alternations, but not observed switch time correlations. A Bayesian model for alternations using an evidence accumulation process [46] succeeds in reproducing correlations. In these models the initial bias to integration is set by specifying a priori initial conditions [45,46].

Competition-based models proposed for visual bistability (e.g. binocular rivalry) incorporate mutual inhibition, slow adaptation and noise [49]. In competition-based dynamics a slow adaptation process sets durations and produces switch correlations [45]. The phenomenological model presented in [44] treats build-up and subsequent bistability separately. The pattern discovery stage addresses algorithmically the formation of the different perceptual patterns during build-up and the initial bias to integration emerges from this process (albeit without a link to neural computations). Abstracted units assigned to each perceptual pattern (once discovered) enter into competition through mechanisms similar to those described above.

Our recent study introduced the first neuromechanistic competition model of auditory bistability [36], a departure from percept-based rivalry models. Dynamic inputs are linked to sensory features by mimicking the neuronal responses from electrophysiologically recorded A1 [37] (Figure 2d). On the basis of a theoretical description proposed in [38] (Figure 2e), the model considers competition downstream of A1 (Figure 2f). It captures

Figure 2



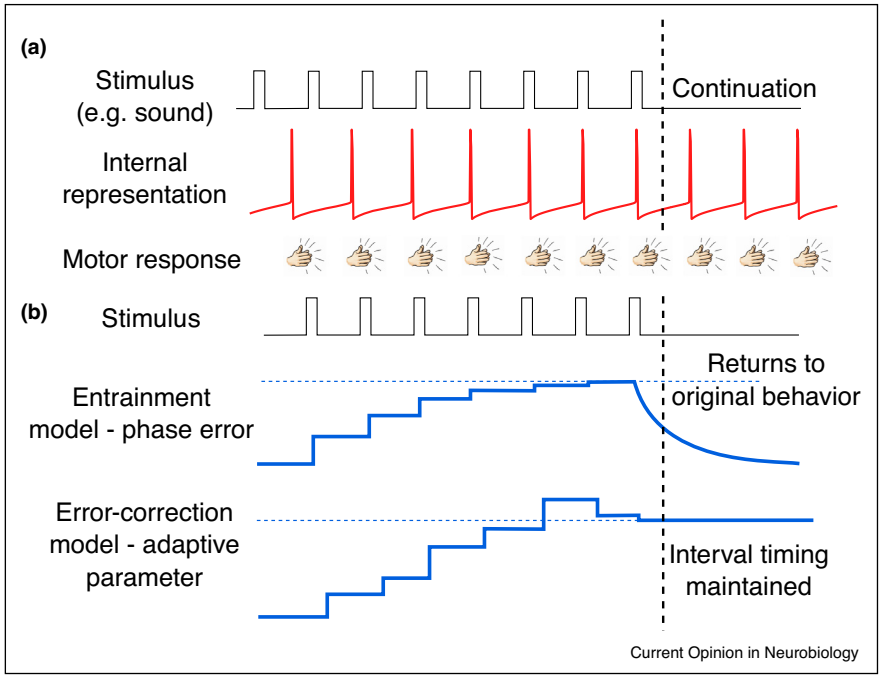
Dynamics and competition for auditory streaming. **(a)** Stimulus paradigm where low A tones, high B tones (separated by the difference in tone frequency DF) and silences (_) each of 100 ms repeat in an ABA_ triplet pattern. **(b)** Stimulus is perceived as either one integrated stream ABA_ABA_ . . . or two segregated streams A_A_A_A_ . . . and _B___B__, ... **(c)** Perceptual reports for 90 s of a single trial [36]. Initial percept is integrated (bias to integration) followed by a switch to segregated within first ~10 s (build-up phase). Subsequently perception alternates every ~2–5 s between integrated and segregated (bistability) if DF is not too large or too small. **(d)** Neural responses in primary auditory cortex (A1) to repeating triplet stimulus at three tonotopic locations with best frequency A (red), B (green) or in between at (A + B)/2 (blue). Time axis as in panel G, vertical offset for visualization only. Responses mimick trial-averaged firing rates from [37] capturing qualitative characteristics: rapid early adaption of overall amplitude (timescale 500 ms), initially responses are broad across tonotopy with similar responses to all tones at each location and tonotopic dependence emerges after early adaptation with full responses to the A (B) tones at the A (B) location and reduced responses to each tone at the intermediate location (A+B)/2. **(e)** Schematic of the population separation model proposed in [38]. Tonotopic spread of responses to A and B tones gives significant overlap (blue shaded region) if DF is sufficiently small. Interpretation: For small DF joint responses to both A and B tones centered at the location (A + B)/2 presumably leads to the integrated percept. For large DF minimal or no overlap leads to the segregated percept. At intermediate values both percepts are possible, resulting in build-up to segregation (which increases gradually after trial averaging) followed by bistability. **(f)** The three-unit competition model proposed in [36] pools inputs from the three tonotopic locations in panel D. The model's competition stage shown here is assumed to be downstream of (and taking input from) A1, with mutual inhibition between units, adaptation and noise driving competition. **(g)** One model simulation showing the activation threshold (horizontal dashed), and each population's excitation variable (solid) and adaptation variable (dashed). When the central AB unit is active (integrated), the peripheral units are suppressed through mutual inhibition. Rising adaptation for AB increases the probability of noise inducing a switch; when units A or B become active and dominant after ~4.5 s (segregated), the integrated (AB) unit is suppressed. **(h)** Averaging across many behavioral trials (or many simulations), the smooth build-up [37] in the probability of segregation and dependence on DF (faster for larger DF) is captured by the model when early adaptation, as shown in panel (d), is included (solid curves). Without early adaptation (dashed curves) the responses only reflect the probability of segregation for post-build-up alternations.

the switching statistics of bistable auditory perception for long stimulus presentations and their dependence on DF [36]. The work was recently extended to account for early bias to integration and build-up (Figure 2g–h) [50]. Our model demonstrates that broader tonotopic responses in A1 before rapid adaptation on a timescale of 500 ms biases towards integration, whilst the slower timescale of build-up (~10 s) emerges from competition downstream. Our model is the first treatment — through a direct link to observed neurophysiological responses — to explain both the initial bias for integration and the apparent disparity between adaptation timescales.

Auditory beat perception, beat generation and sensorimotor synchronization

Humans have a remarkable ability to perceptually track complicated sensory patterns and synchronize movement, even predicting upcoming events [51–54] as investigated behaviorally in finger tapping experiments [54] (Figure 3a). A recent review of imaging experiments investigating musical rhythm and timing proposes that the perception and production of rhythm relies on similar mechanisms involving sensory and motor areas [55]. Indeed, perception of simple musical rhythms (without movement) involves auditory and motor regions, as shown

Figure 3



Sensorimotor synchronization (SMS) and beat generation, entrainment models and error-correction models. **(a)** Repetitive stimulus with regular intervals between events (black). An internal representation predicts the onset of the next event (spikes in red). Internal representation drives motor responses, for example finger tapping or clapping, timed to match the stimulus. The timing entrains to match the stimulus after 5–8 cycles. This process is faster than say long term potentiation or depression requiring hundreds of repetitions. The learned interval time is maintained after stimulus offset (vertical black) and the listener continues to clap in time (beat generation). **(b)** In entrainment models, the difference in phase between the stimulus and its internal representation monotonically increases towards 0 (horizontal blue). At stimulus offset (vertical black) the phase error drifts away as the oscillator returns to its intrinsic frequency of oscillation (unless this matches the stimulus exactly). In error-correction models, a parameter is adapted in discrete steps to reduce the error between the stimulus and its internal representation. Whilst this approach can overshoot, the stimulus interval timing is learned and maintained at stimulus offset, allowing for continuation of the beat with correct timing.

in combined behavioral and fMRI [56] or EEG [57,58] experiments. However, with these approaches distinguishing perceptual from sensory signals is challenging and the necessary trial averaging compromises timing information.

Models of rhythm and beat perception are geared towards understanding how temporally structured stimuli generated patterns of neural activation (Figure 3a, internal representation) from which perceptual experience is derived. A hierarchical auditory and motor oscillator model can explain the perception of musical pulse at frequencies without spectral energy through entrainment [59•]. The oscillator model describes an array of canonical oscillators organized by natural frequency where responses gradually entrain with the stimulus (Figure 3b). A recent extension with Hebbian learning for tuning intrinsic oscillator frequencies [60] allows for a large dimensionality reduction [61]. Recently, imaging experiments identified entrained neural activity linked to the perception of a missing pulse in only auditory (not motor) areas [62], suggesting a reassessment of the hierarchy in [59•]. Elsewhere, predictive coding models explain some aspects of processing for more complicated rhythms (e.g. syncopation) [63,64], however, these models focus only on spectral profiles rather than event timing information (recent review [65]).

Models of beat generation focus on frameworks that adapt to timed intervals of an incoming signal and learn a matching pattern (which continues after the input). The framework proposed in [66] depends on an error correction mechanism [67] that samples input–output differences and makes predictions from an internal model (weak anticipation [66]). Tested against behavioral experiments [68], the model best accounts for tempo changes when both correction and prediction mechanisms are incorporated. A recent dynamical model of beat generation [69•] exploits plausible neural mechanisms in an error correction framework so that the neuronal beat generator learns the period and timing of a rhythmic input and continues the beat after input offset. The model, robust for tempo changes and to noise, implements a plasticity rule with gamma oscillations as a timekeeper measuring differences in spike times between inputs and beat generation [70] (plausible neural implementation

the learning rule in [61]). Elsewhere, tap interval timing as effected by noise effects are studied in drift-diffusion models [71,72], but without scope to learn and continue time interval production.

Perspectives

Theoretical advances on sound localization encoding have benefited from close links to neurophysiological experiments early in the auditory pathway in animal models. In a relatively mature field, a long-established and mutually beneficial interplay between theory and experiments has driven significant progress [3•]. Achieving similar advances for auditory streaming and beat perception/generation will depend on such interplay, but with different challenges. Tasks involving perceptual reports and behavior are limited in animal models and these functions involve a network of multiple cortical areas [32,55]. Whilst [36] successfully bridged between available neurophysiological data from macaque A1 [37,38] and behavior in humans [31], future insights are likely to be informed by a closer link to imaging work in humans (three papers exploring attention [33–35], as yet unexplored in models). A prime example on beat perception is a recent imaging study [62] linked to and informative for related modeling work [59•].

There is potential for convergence between models of beat perception/generation and of auditory streaming. The former are in some recent cases adaptive to event timing [61,69•]. Such processes allow a common population of neurons to learn sequences with a range of timing properties with a simple model structure. A similar process is likely at play for streaming but with separate populations entraining oscillations to different streams [35]. In this case the importance of temporal coherence [41] as a cue for binding of events and therefore, for integration, would be emergent. As suggested in [69•], beat generation with more complex stimuli could lead to bistability. In both fields, a drive towards more dynamic environments with slowly varying cues and timing would bring us closer to real-world situations.

Conflict of interest statement

Nothing declared.

Funding

Rankin acknowledges support from an Engineering and Physical Sciences Research Council (EPSRC) New Investigator Award (EP/R03124X/1) and from the EPSRC Centre for Predictive Modelling in Healthcare (EP/N014391/1). This is a review study, and as such did not generate any new data.

Acknowledgements

The authors thank Aine Byrne, Amit Bose, Nace Golding and Jason Mikiel-Hunter for their valuable feedback on the manuscript.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest

1. Grothe B, Pecka M, McAlpine D: **Mechanisms of sound localization in mammals**. *Physiol Rev* 2010, **90**:983-1012.

2. Joris P, van der Heijden M: **Early binaural hearing: the comparison of temporal differences at the two ears**. *Annu Rev Neurosci* 2019, **42**.

3. • Grothe B, Leibold C, Pecka M: **The medial superior olivary nucleus: meeting the need for speed**. *The Oxford Handbook of the Auditory Brainstem*. Oxford University Press; 2018 http://dx.doi.org/10.1093/oxfordhb/9780190849061.013.9.
A valuable review of experimental and computational studies that tested hypotheses and led to understanding of mechanisms that contribute to the neuronal computation of ITD and the dynamics of spatial processing. The phylogenetic context is touched upon; see also Grothe and Pecka (2014) for fascinating considerations.

4. Svirskis G, Kotak V, Sanes DH, Rinzel J: **Enhancement of signal-to-noise ratio and phase locking for small inputs by a low-threshold outward current in auditory neurons**. *J Neurosci* 2002, **22**:11019-11025.

5. Meng X, Huguet G, Rinzel J: **Type III excitability, slope sensitivity and coincidence detection**. *Discrete Contin Dyn Syst Ser A* 2012, **32**:2729.

6. Couchman K, Grothe B, Felmy F: **Medial superior olivary neurons receive surprisingly few excitatory and inhibitory inputs with balanced strength and short-term dynamics**. *J Neurosci* 2010, **30**:17111-17121.

7. Lehnert S, Ford MC, Alexandrova O, Hellmundt F, Felmy F, Grothe B, Leibold C: **Action potential generation in an anatomically constrained model of medial superior olive axons**. *J Neurosci* 2014, **34**:5370-5384.

8. Goldwyn JH, Remme MW, Rinzel J: **Soma-axon coupling configurations that enhance neuronal coincidence detection**. *PLoS Comput Biol* 2019, **15**:e1006476.

9. Scott LL, Mathews PJ, Golding NL: **Posthearing developmental refinement of temporal processing in principal neurons of the medial superior olive**. *J Neurosci* 2005, **25**:7887-7895.

10. Agmon-Snir H, Carr CE, Rinzel J: **The role of dendrites in auditory coincidence detection**. *Nature* 1998, **393**:268.

11. Mathews PJ, Jercog PE, Rinzel J, Scott LL, Golding NL: **Control of submillisecond synaptic timing in binaural coincidence detectors by Kv1 channels**. *Nat Neurosci* 2010, **13**:601-609.

12. Winters BD, Jin S-X, Ledford KR, Golding NL: **Amplitude normalization of dendritic EPSPs at the soma of binaural coincidence detector neurons of the medial superior olive**. *J Neurosci* 2017, **37**:3138-3149.

13. Ashida G, Carr CE: **Sound localization: Jeffress and beyond**. *Curr Opin Neurobiol* 2011, **21**:745-751.

14. • Brand A, Behrend O, Marquardt T, McAlpine D, Grothe B: **Precise inhibition is essential for microsecond interaural time difference coding**. *Nature* 2002, **417**:543.
These authors showed that blocking inhibition had a significant asymmetric effect on ITD-tuning of MSO neurons, resulting in the slope-based coding hypothesis, a major deviation from the Jeffress model that identifies the peak of the tuning curve with ITD and sound source location. Simulations with a biophysically based model assuming fast and precisely timed inhibition was used to demonstrate the observed electrophysiological data.

15. Lingner A, Pecka M, Leibold C, Grothe B: **A novel concept for dynamic adjustment of auditory space**. *Sci Rep* 2018, **8**.

16. Jercog PE, Svirskis G, Kotak VC, Sanes DH, Rinzel J: **Asymmetric excitatory synaptic dynamics underlie interaural time difference processing in the auditory system**. *PLoS Biol* 2010, **8**:e1000406.

17. Zhou Y, Carney LH, Colburn HS: **A model for interaural time difference sensitivity in the medial superior olive: interaction of excitatory and inhibitory synaptic inputs, channel dynamics, and cellular morphology**. *J Neurosci* 2005, **25**:3046-3058.

18. Rothman JS, Manis PB: **Kinetic analyses of three distinct potassium conductances in ventral cochlear nucleus neurons**. *J Neurophysiol* 2003, **89**:3083-3096.

19. Scott LL, Mathews PJ, Golding NL: **Perisomatic voltage-gated sodium channels actively maintain linear synaptic integration in principal neurons of the medial superior olive**. *J Neurosci* 2010, **30**:2039-2050.

20. Lloyd Jeffress A: **A place theory of sound localization**. *J Comp Physiol Psychol* 1948, **41**:35.

21. Roberts MT, Seeman SC, Golding NL: **A mechanistic understanding of the role of feedforward inhibition in the mammalian sound localization circuitry**. *Neuron* 2013, **78**:923-935.

22. Harper NS, McAlpine D: **Optimal neural population coding of an auditory spatial cue**. *Nature* 2004, **430**:682.

23. Harper NS, Scott BH, Semple MN, McAlpine D: **The neural code for auditory space depends on sound frequency and head size in an optimal manner**. *PLOS ONE* 2014, **9**:e108154.

24. Franken TP, Roberts MT, Wei L, Golding NL, Joris PX: **In vivo coincidence detection in mammalian sound localization generates phase delays**. *Nat Neurosci* 2015, **18**:444.

25. Remme MW, Donato R, Mikiel-Hunter J, Ballestero JA, Foster S, Rinzel J, McAlpine D: **Subthreshold resonance properties contribute to the efficient coding of auditory spatial cues**. *Proc Natl Acad Sci U S A* 2014, **111**:E2339-E2348.

26. Fischer L, Leibold C, Felmy F: **Resonance properties in auditory brainstem neurons**. *Front Cell Neurosci* 2018, **12**:8.

27. Mikiel-Hunter J, Kotak V, Rinzel J: **High-frequency resonance in the gerbil medial superior olive**. *PLoS Comput Biol* 2016, **12**: e1005166.

28. Bendixen A: **Predictability effects in auditory scene analysis: a review**. *Front Neurosci* 2014, **8**:60.

29. Snyder JS, Elhilali M: **Recent advances in exploring the neural underpinnings of auditory scene perception**. *Ann N Y Acad Sci* 2017 http://dx.doi.org/10.1111/nyas.13317.

30. van Noorden LPAS: **Temporal coherence in the perception of tone sequences**. *PhD Thesis*. Eindhoven University; 1975.

31. Pressnitzer D, Hupé J: **Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization**. *Curr Biol* 2006, **16**:1351-1357.

32. Kashino M, Kondo H: **Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations**. *Philos Trans R Soc Lond Ser B: Biol Sci* 2012, **367**:977-987.

33. Billig AJ, Davis MH, Carlyon RP: **Neural decoding of bistable sounds reveals an effect of intention on perceptual organization**. *J Neurosci* 2018:3022-3117.

34. Kondo HM, Pressnitzer D, Shimada Y, Kochiyama T, Kashino M: **Inhibition–excitation balance in the parietal cortex modulates volitional control for auditory and visual multistability**. *Sci Rep* 2018, **8**:14548.

35. Costa-Faidella J, Sussman ES, Escera C: **Selective entrainment of brain oscillations drives auditory perceptual organization**. *NeuroImage* 2017, **159**:195-206.

36. Rankin J, Sussman E, Rinzel J: **Neuromechanistic model of auditory bistability**. *PLoS Comput Biol* 2015, **11**:e1004555.

37. Micheyl C, Tian B, Carlyon R, Rauschecker J: **Perceptual organization of tone sequences in the auditory cortex of awake macaques**. *Neuron* 2005, **48**:139-148.

38. Fishman Y, Reser D, Arezzo J, Steinschneider M: **Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey**. *Hear Res* 2001, **151**:167-187.

39. • Szabó BT, Denham SL, Winkler I: **Computational models of auditory scene analysis: a review**. *Front Neurosci* 2016, **10** http://dx.doi.org/10.3389/fnins.2016.00524.
A comprehensive review of computational models of auditory scene analysis that proposes an integrative approach that exploits complementary modeling frameworks.

40. Beauvois M, Meddis R: **Computer simulation of auditory stream segregation in alternating-tone sequences**. *J Acoust Soc Am* 1996, **99**:2270-2280.

41. Krishnan L, Elhilali M, Shamma S: **Segregating complex sound sources through temporal coherence**. *PLoS Comput Biol* 2014, **10**:e1003985.

42. Almonte F, Jirsa V, Large E, Tuller B: **Integration and segregation in auditory streaming**. *Physica D* 2005, **212**:137-159.

43. Wang D, Chang P: **An oscillatory correlation model of auditory streaming**. *Cogn Neurodyn* 2008, **2**:7-19.

44. Mill R, Böhm T, Bendixen A, Winkler I, Denham S: **Modelling the emergence and dynamics of perceptual organisation in auditory streaming**. *PLoS Comput Biol* 2013, **9**:e1002925.

45. Steele S, Tranchina D, Rinzel J: **An alternating renewal process describes the buildup of perceptual segregation**. *Front Comput Neurosci* 2015, **8**:1-13.

46. Barniv D, Nelken I: **Auditory streaming as an online classification process with evidence accumulation**. *PLOS ONE* 2015, **10**:e0144788.

47. Cao R, Pastukhov A, Mattia M, Braun J: **Collective activity of many bistable assemblies reproduces characteristic dynamics of multistable perception**. *J Neurosci* 2016, **36**:6957-6972 http://dx.doi.org/10.1523/JNEUROSCI.4626-15.2016.

48. Denham SL, Farkas D, Van Ee R, Taranu M, Kocsis Z, Wimmer M, Carmel D, Winkler I: **Similar but separate systems underlie perceptual bistability in vision and audition**. *Sci Rep* 2018, **8**.

49. Laing C, Chow C: **A spiking neuron model for binocular rivalry**. *J Comput Neurosci* 2002, **12**:39-53.

50. Rankin J, Osborn Popp P, Rinzel J: **Stimulus pauses and perturbations differentially delay or promote the segregation of auditory objects: psychoacoustics and modeling**. *Front Neurosci* 2017, **11** http://dx.doi.org/10.3389/fnins.2017.00198.

51. Bendixen A, SanMiguel I, Schröger E: **Early electrophysiological indicators for predictive processing in audition: a review**. *Int J Psychophysiol* 2012, **83**:120-131.

52. Repp BH, Su Y-H: **Sensorimotor synchronization: a review of recent research (2006–2012)**. *Psychon Bull Rev* 2013, **20**:403-452.

53. Todd NP, Lee CS: **The sensory-motor theory of rhythm and beat induction 20 years on: a new synthesis and future perspectives**. *Front Hum Neurosci* 2015, **9**:444.

54. Iversen JR, Balasubramaniam R: **Synchronization and temporal processing**. *Curr Opin Behav Sci* 2016, **8**:175-180.

55. Grahn JA: **Neural mechanisms of rhythm perception: current findings and future perspectives**. *Top Cogn Sci* 2012, **4**:585-606.

56. Bengtsson SL, Ulln F, Henrik Ehrsson H, Hashimoto T, Kito T, Naito E, Forssberg H, Sadato N: **Listening to rhythms activates motor and premotor cortices**. *Cortex* 2009, **45**:62-71 http://dx.doi.org/10.1016/j.cortex.2008.07.002.

57. Nozaradan S: **Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging**. *Philos Trans R Soc Lond Ser B: Biol Sci* 2014, **369**:20130393 http://dx.doi.org/10.1098/rstb.2013.0393.

58. Nozaradan S, Peretz I, Keller PE: **Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization**. *Sci Rep* 2016, **6**:20612.

59. • Large E, Herrera J, Velasco M: **Neural networks for beat perception in musical rhythm**. *Front Syst Neurosci* 2015:159 http://dx.doi.org/10.3389/fnsys.2015.00159.
Using a network of coupled canonical oscillators, the authors were able to show, through entrainment mechanisms, the emergence of network

activity (matching perception in behavioral experiments) corresponding to the musical pulse at frequencies without spectral energy.

60. Righetti L, Buchli J, Ijspeert AJ: **Dynamic Hebbian learning in adaptive frequency oscillators**. *Physica D: Nonlinear Phenom* 2006, **216**:269-281.

61. Lambert AJ, Weyde T, Armstrong N: **Adaptive frequency neural networks for dynamic pulse and metre perception**. *International Society for Music Information Retrieval (ISMIR) Conference* 2016:60-66 http://m.mr-pc.org/ismir16/website/articles/228_Paper.pdf.

62. Tal I, Large EW, Rabinovitch E, Wei Y, Schroeder CE, Poeppel D, Golumbic EZ: **Neural entrainment to the beat: the missing-pulse phenomenon**. *J Neurosci* 2017, **37**:6331-6341.

63. Vuust P, Witek MA: **Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music**. *Front Psychol* 2014, **5**:1111.

64. Vuust P, Dietz MJ, Witek M, Kringelbach ML: **Now you hear it: a predictive coding model for understanding rhythmic incongruity**. *Ann N Y Acad Sci* 2018, **1423**:19-29 Wiley Online Library.

65. Koelsch S, Vuust P, Friston K: **Predictive processes and the peculiar case of music**. *Trends Cogn Sci* 2019, **23**:63-77.

66. van der Steen MC, Keller PE: **The adaptation and anticipation model (ADAM) of sensorimotor synchronization**. *Front Hum Neurosci* 2013, **7**:253.

67. Mates J: **A model of synchronization of motor acts to a stimulus sequence**. *Biol Cybern* 1994, **70**:463-473.

68. van der Steen MM, Jacoby N, Fairhurst MT, Keller PE: **Sensorimotor synchronization with tempo-changing auditory sequences: modeling temporal adaptation and anticipation**. *Brain Res* 2015, **1626**:66-87.

69. • Bose A, Byrne A, Rinzel J: **A neuromechanistic model for rhythmic beat generation**. *PLoS Comput Biol* 2018, **15**:e1006450 http://dx.doi.org/10.1371/journal.pcbi.1006450.
A plausible neural implementation of beat generation producing continuation behavior (Figure 2a) with robustness to tempo changes and noise in event onset times. Interval timing differences between inputs and beat generation are computed via cycle counts of gamma oscillations as per [44].

70. Naud R, Houtman DB, Rose GJ, Longtin A: **Counting on dis-inhibition: a circuit motif for interval counting and selectivity in the anuran auditory system**. *Am J Physiol: Heart Circ Physiol* 2015, **114**:2804-2815 http://dx.doi.org/10.1152/jn.00138.2015 PMID: 26334004.

71. Simen P, Vlasov K, Papadakis S: **Scale (in)variance in a unified diffusion model of decision making and timing**. *Psychol Rev* 2016, **123**:151.

72. Merchant H, Averbeck BB: **The computational and neural basis of rhythmic timing in medial premotor cortex**. *J Neurosci* 2017. 0367-17.