Hierarchical Representations in the Auditory Cortex

Tatyana O. Sharpee[1], **Craig A. Atencio**[2,3], and **Christoph E. Schreiner**[2,3]

[1]The Salk Institute for Biological Studies, La Jolla CA 92037 and the Center for Theoretical Biological Physics, University of California, San Diego, La Jolla, CA 92093

[2]W.M. Keck Foundation Center for Integrative Neuroscience, University of California, San Francisco CA 94143-0732

[3]Coleman Memorial Laboratory, Department of Otolaryngology-HNS, University of California, San Francisco CA 94143-0732

Abstract

Understanding the neural mechanisms of invariant object recognition remains one of the major unsolved problems in neuroscience. A common solution that is thought to be employed by diverse sensory systems is to create hierarchical representations of increasing complexity and tolerance. However, in the mammalian auditory system many aspects of this hierarchical organization remain undiscovered, including the prominent classes of high-level representations (that would be analogous to face selectivity in the visual system or selectivity to bird's own song in the bird) and the dominant types of invariant transformations. Here we review the recent progress that begins to probe the hierarchy of auditory representations, and the computational approaches that can be helpful in achieving this feat.

Introduction

Although object recognition seems effortless, it is a challenging computational problem [1]. The difficulty arises because of the dual need to be able to discriminate stimuli based on potentially subtle yet important clues [2,3], such as discriminating between different syllables, regardless of how fast they are spoken or the speaker's voice. This suggests that peripheral representations might be recoded in a form that is tolerant, or "invariant" to "identity-preserving" transformations [1]. Computer science studies show that increasing the number of processing layers can broaden the range of recognition tasks that the circuit can handle, while simultaneously improving performance, learning time, and exponentially reducing the number of neurons (reviewed in [4**]). However, finding the right hierarchical structure for broad recognition systems that can match human performance regarding object recognition in natural environments remains an open problem.

Characterizing feature selectivity

How can we systematically characterize the preferred stimulus features along the auditory hierarchy? In the visual system, progress was made possible by Hubel and Weisel's discovery that bars and edges represent a close-to-optimal stimulus feature for many cells in the primary visual cortex (V1), and by the discoveries of other researchers that preferred stimulus features further along the hierarchy include curved contours [5], followed by hands and faces [6]. In the auditory system, one can use neuroethology to help guess the optimal stimulus [7]. For example, selectivity for a bird's own song was demonstrated within areas of avian auditory systems that are homologous to mammalian secondary auditory cortices. However, analogous stimuli for mammalian auditory neurons have been difficult to identify. For example, when vocalizations are played in forward and reverse directions, the

differences between the responses are less robust in auditory cortical areas [8-11] compared to those found in the avian song selective areas. Based on the presumed complexity of representations at the level of the primary auditory cortex (A1), it has been argued that A1 is less similar to primary visual cortex and more similar to the final stages of visual processing located in the inferotemporal cortex [12]. This viewpoint is further supported by observation that neural responses are less redundant in auditory cortex and thalamus than they are in inferior colliculus [13**]. By comparison, neurons in the output layers of V1 are more correlated than in the input V1 layers [14], suggesting the redundancy among V1 neurons is higher than at the earlier visual stations.

One-dimensional models

Without an ethologically guided guess of what the best stimulus might be, there are two types of statistical approaches for characterizing auditory feature selectivity. One family of approaches relies on adaptive search procedures whereby the stimulus is generated according to the responses to past stimuli in an effort to increase the strength of the response [15-17]. Theoretical work continues on improving methods for adaptive stimulus design [18], making it a promising research direction for the characterization of neurons with complex feature selectivity.

The second family of statistical methods consists in recording the responses of neurons to large numbers of sounds. After the experiment, one correlates which stimuli elicited a spike and which did not. In its simplest form, this procedure involves computing the spike-triggered average (STA) – the difference between the average of all segments of stimuli that elicited a spike and the average of all presented stimuli [19-23]. Pioneered by de Boer [19], this approach has been generalized to characterize spectrotemporal filtering properties of neurons from the auditory periphery to A1 [20,22,23], as well as in other sensory modalities (for a review focused on the visual system see [24]). Applying this approach in A1, one study found that neurons of awake primates were particularly selective for frequency modulations [25**]. This was later confirmed using synthetic stimuli termed "drifting dynamic ripples" [26*]. Other studies in A1 found simpler features that often consisted of a single excitatory region surrounded by one or more inhibitory regions [27,28]. Additionally, when relevant spectrotemporal features have been obtained with linear models, a range of context-dependent phenomena have been observed: in some cases the features were very similar across different stimulus ensembles [29], in other cases essential nonlinearities were observed [30**, 31-32], and in some there was a reduced fraction of explained variance [33,34].

How can we reconcile these findings? It turns out that the complexity of relevant stimulus features and the degree to which neural responses can be described by a linear model varies systematically across the cortical column in A1 (Figure 1). Granular layers have preferred stimulus features that are more separable in frequency/time space than those in supra- or infra-granular layers. Furthermore, responses in granular layers were more "linear", i.e. they were better described by a model with one relevant stimulus feature than neurons in output layers [35*,36]. These findings can help explain the observed range of complexity in feature selectivity of A1 neurons and how well linear models account for their responses (see also [34,37]).

One of the advantages of spike-triggered methods is that they provide not only the estimate of the preferred stimulus feature (spectrotemporal receptive field, STRF) but also the nonlinear gain function (nonlinearity) that describes how spiking probability changes as a function of the similarity between the presented stimuli and the optimal stimulus feature. The nonlinearity can capture some inherent neural response properties, such as rectification

(firing rate cannot be negative) and saturation (firing rate is limited by the refractory period). These effects become increasingly more pronounced along the auditory neuroaxis (Figure 1). Incorporation of the nonlinearity into models can improve the accuracy of neural response predictions [38]. Crucially, it can help reconcile the observed fast rise times and short response durations in neural responses to frequency modulated (FM) tones with the relatively slow time course of STRFs. Although in many cases the presence of the nonlinearity does not affect STRF estimation [19,39-42], (but see [43-45] for exceptions to this rule; dimensionality reduction methods are summarized in Figure 2), taking the nonlinearity into account can sharpen the predicted tuning. This effect alone could account for differences in the dynamics of V1 responses and their relevant spatiotemporal features [46,47] (reviewed in [48**]). It would be exciting to see if similarly simple calculations that take the nonlinearity into account can resolve the analogous controversies in the auditory research.

Multidimensional models

However successful, models based on one stimulus feature are problematic from the standpoint of invariant sound identification. For example, any 1D model will confound responses to a suboptimal feature presented loudly with an optimal feature presented softly. Similarly, a 1D model cannot implement invariance to cadence, because the stretching in time will alter the match with any given STRF. There are two types of general strategies for solving the context-independent sound identification problem. One is to expand the stimulus space, treating the same stimuli as different depending on the value of a contextual variable. This approach is comprehensive, but can be done only for a few well defined context variables, such as mean sound level. Multilinear models that work in the three dimensional stimulus space defined by time lag, frequency, and sound level provide good descriptions of A1 responses [49*]. Mean sound level holds special importance in auditory perception, and could thus justify the expansion of stimulus dimensionality. The second strategy for solving the object recognition problem is not to explicitly expand the stimulus space, but to use combinations of different features. For example, phenomena such as two-tone masking (one tone affects the response to another simultaneously present tone) and forward suppression (a preceding tone suppresses the response to a following tone) [50-53] can be seen as building blocks of invariant sound identification [54]. These classic auditory phenomena can be accounted for by multidimensional LN models with a nonlinearity that depends on stimulus components along two (or more) relevant stimulus features. The form of nonlinearity is not restricted, and can characterize both cooperative and suppressive effects between the relevant components. A variety of computations with respect to two features have been observed in A1 [55,56*] (Figure 1).

The presence of multiple relevant stimulus features is consistent with other nonlinear and context-dependent effects observed throughout the hierarchy of auditory representations. For example, multidimensional descriptions of neural coding may provide a complementary way to account for the dynamic effects of first spike latency [57]. While 1D models with a fixed threshold applied to low-pass filtered sound amplitude could account for much of the latency data [57**], the introduction of a time-dependent threshold was necessary for a full description. A two-dimensional model with a fixed threshold that is a function of both the low-pass filtered amplitude and its first time derivative may provide another way of accounting for this phenomenon. Another example concerns the hypersensitivity of auditory cortical neurons to small perturbations of their acoustic input [31], as well as the large effect that naturalistic background noise can have on the responses [30]. These effects are difficult to explain based on a single relevant stimulus feature, but they can be explained with a multidimensional LN model [55,56*]. Curiously, the hypersensitivity of auditory neurons as manifested by strong suppression of the responses from the addition of a subthreshold tone

is not observed in the inferior colliculus; it first appears in the medial geniculate body and is present in A1 [58]. This suggests that multidimensional feature selectivity – an essential property for performing object recognition – appears as a result of hierarchical auditory processing.

The emergence of 'multiplexed' temporal coding

The temporal dimension is central to the sense of hearing, and is reflected in the sensitivity to different time-scales of information in auditory signals, perhaps as the spatial dimensions are to the sense of vision [2,3]. In both modalities, neurons are able to integrate information over long scales while also remaining sensitive to fine details. In the visual system, invariance to image translation is one of the prominent characteristics of high-level neurons. Face-selective neurons in the inferotemporal cortex integrate visual signals across large regions of the visual space while maintaining fine spatial sensitivity (which is necessary to distinguish between different faces). Analogously, acoustic stimuli can also be characterized on multiple temporal scales such as the "envelope" (the contour of amplitude modulations (AM) of the spectral components) and the "fine-structure" (the cycle-by-cycle variations of the spectral components that contribute to the time waveform). Along the auditory neuroaxis, which extends from the peripheral stations to A1, there is a progressively increasing emphasis on encoding at coarser temporal resolution, as well as increased tolerance to other stimulus parameters, such as sound pressure level or modulation depth [59]. However, neurons maintain the ability for precise spike timing. Spike timing with single millisecond precision has been observed for sound onset detection [60,61] and with precision of a few milliseconds for a variety of other auditory stimuli, including tones [62,63], AM sounds [64], their combinations [65**], as well as natural stimuli [11,66] (although the utilized temporal precision is coarser than found in peripheral or subcortical auditory structures [66]). Thus, a picture is emerging A1 neurons use a relatively wide range of temporal cues, in a manner that is reminiscent of the wide range of spatial scales that characterize responses of high-level visual neurons.

Though there may be a shift from a temporal code to an average rate and place code along the neuroaxis, both types of information may still be utilized. On one hand, the rate coding of AM sounds in the auditory cortex carries significant information in A1, and it has been suggested that rate information alone is sufficient to explain the behavioral performance of monkeys in a low-AM discrimination task [67] (but see [68]). On the other hand, much of the amplitude modulation information is reflected in the timing of cortical responses (e.g., [69]). For example, spike-doublets with <15ms interspike intervals (ISIs) were shown to convey more event information than long-ISI spikes. Pairs of short-ISI spikes express over three times as much information as long-ISI spikes, well over what would be expected from summing two independent information sources [70]. Therefore, short-ISI spikes appear to be particularly important in A1 stimulus encoding and have the potential to provide low-noise, robust, and efficient representations of sound features. It is also possible that different populations of A1 neurons use distinct encoding schemes, either synchronized (temporal coding) or non-synchronized (rate-coding) [71-73]. A recent study expanded on this issue by demonstrating mixed schemes with synchronized responses at some modulation frequencies and non-synchronized responses at others [74]. Accordingly, cortical neurons are capable of carrying multiple signals via different codes with regard to AM. The information conveyed by these different codes (rate and time) is likely non-redundant, in that a joint code of rate and timing parameters provides more information than either code alone, as demonstrated for high modulation frequencies [75]. A similar observation has been made for low modulations frequencies (<60Hz) that dominate temporally encoded A1 activity [76]. Here, repetition–rate information carried jointly by firing rate and inter-stimulus intervals

exceeded that of either code alone, thus indicating non-redundant contributions of the two codes.

Concurrently employed codes may also provide complementary information, as demonstrated for LFP and spiking signals for natural sounds [77**]. Further analysis showed that the angular phase of the LFP at the time of spike generation adds significant extra information, beyond that contained in the firing rate alone [78]. These findings provide further credence to the notion of 'multiplexed' coding at different timescales [79], with each code carrying complementary information.

While the impact of hierarchical and parallel schemes of information processing beyond A1 on 'multiplexed' coding is still unresolved, there is evidence of hierarchical processing within A1, namely across the different cortical layers. Here, granular layers phase-lock to the highest pure-tone frequencies [62]. Additionally, granular layers may also follow faster amplitude modulations, and outside of the main thalamic input zone the following rates generally decrease [80].

Outlook

The emerging picture is that auditory processing becomes increasingly multidimensional. This is expected on computational grounds because the invariant representation of auditory objects requires that neural responses be tuned to conjunctions of features. For example, David Marr argued that to detect an edge [81] one should measure both the presence of an oriented element and the absence of an oriented element of the perpendicular orientation. Mechanisms of forward- and two-tone masking may serve as examples with a similar computational purpose. Additionally, since there is a high degree of tolerance and selectivity at the level of A1, neural responses may be simultaneously sensitive to a large number of stimuli. Thus, the full understanding of auditory representations will likely require the development of new statistical methods that can recover large numbers of relevant stimulus features from responses to naturalistic sounds. The development of these methods may be guided by the construction of Bayesian methods for model selection [49*,82], especially for building minimal nonlinear models [83]. In addition, to achieve a full understanding of temporal processing, temporal aspects of neural coding for the same stimulus need to be separated from stimulus dynamics. This requires that the information content of the stimulus be quantified, and then compared to the content in the neural response. In this way the maximally achievable stimulus information may be compared to the encoding that is actually provided by the neuron.

Acknowledgments

References

1. DiCarlo JJ, Cox DD. Untangling invariant object recognition. Trends Cogn Sci. 2007; 11:333–341. [PubMed: 17631409]

2. Moore BC. The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. J Assoc Res Otolaryngol. 2008; 9:399–406. [PubMed: 18855069]

3. Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BC. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc Natl Acad Sci U S A. 2006; 103:18866–18869. [PubMed: 17116863]

*4. Bengio, Y.; LeCun, Y. Scaling learning algorithms to {AI}. In: Bottou, L.; Chapelle, O.; DeCoste, D.; Weston, J., editors. Large_scale Kernel Machines. MIT Press; 2007. p. 321-360.This article lays out, in an intuitive form, the computational arguments in favor of deep hierarchical organization in object recognition circuits

5. Connor CE, Brincat SL, Pasupathy A. Transformation of shape information in the ventral pathway. Curr Opin Neurobiol. 2007; 17:140–147. [PubMed: 17369035]

6. Desimone R, Albright TD, Gross CG, Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci. 1984; 4:2051–2062. [PubMed: 6470767]

7. Garcia-Lazaro JA, Ahmed B, Schnupp JW. Tuning to natural stimulus dynamics in primary auditory cortex. Curr Biol. 2006; 16:264–271. [PubMed: 16461279]

8. Recanzone GH. Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey. J Neurosci. 2008; 28:13184–13193. [PubMed: 19052209]

9. Wang X, Kadia SC. Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. J Neurophysiol. 2001; 86:2616–2620. [PubMed: 11698547]

10. Qin L, Wang JY, Sato Y. Representations of cat meows and human vowels in the primary auditory cortex of awake cats. J Neurophysiol. 2008; 99:2305–2319. [PubMed: 18305087]

11. Schnupp JW, Hall TM, Kokelaar RF, Ahmed B. Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. J Neurosci. 2006; 26:4785–4795. [PubMed: 16672651]

12. King AJ, Nelken I. Unraveling the principles of auditory cortical processing: can we learn from the visual system? Nat Neurosci. 2009; 12:698–701. [PubMed: 19471268]

**13. Chechik G, Anderson MJ, Bar-Yosef O, Young ED, Tishby N, Nelken I. Reduction of information redundancy in the ascending auditory pathway. Neuron. 2006; 51:359–368. This articles demonstrates a greater redundancy among neural responses in inferior colliculus compared to auditory thalamus and cortex. This findings strengthens the viewpoint that auditory cortex is more analogous to higher, extrastriate visual areas than is to the primary visual cortex where correlations between neurons increase from input to output layers. [PubMed: 16880130]

14. Bryan H, Chelaru M, Dragoi V. Correlated variability in malinar cortical circuits. Computational and Systems Neuroscience. 2011

15. O'Connor KN, Petkov CI, Sutter ML. Adaptive stimulus optimization for auditory cortical neurons. J Neurophysiol. 2005; 94:4051–4067. [PubMed: 16135553]

16. Nelken I, Prut Y, Vaadia E, Abeles M. In search of the best stimulus: an optimization procedure for finding efficient stimuli in the cat auditory cortex. Hear Res. 1994; 72:237–253. [PubMed: 8150740]

17. Carlson ET, Rasquinha RJ, Zhang K, Connor CE. A sparse object coding scheme in area V4. Curr Biol. 2011; 21:288–293. [PubMed: 21315595]

18. Lewi J, Schneider DM, Woolley SM, Paninski L. Automating the design of informative sequences of sensory stimuli. J Comput Neurosci. 2010; 30:181–200. [PubMed: 20556641]

19. de Boer E, Kuyper P. Triggered Correlation. IEEE Trans Biomed Eng. 1968; 15:169–179. [PubMed: 5667803]

20. Aertsen AM, Johannesma PI. The spectro-temporal receptive field A functional characteristic of auditory neurons. Biol Cybern. 1981; 42:133–143. [PubMed: 7326288]

21. Johannesma P, Aertsen A. Statistical and dimensional analysis of the neural representation of the acoustic biotope of the frog. J Med Syst. 1982; 6:399–421. [PubMed: 6982953]

22. Eggermont JJ, Johannesma PM, Aertsen AM. Reverse-correlation methods in auditory research. Q Rev Biophys. 1983; 16:341–414. [PubMed: 6366861]

23. Eggermont JJ, Aertsen AM, Hermes DJ, Johannesma PI. Spectro-temporal characterization of auditory neurons: redundant or necessary. Hear Res. 1981; 5:109–121. [PubMed: 6976342]

24. Schwartz O, Pillow JW, Rust NC, Simoncelli EP. Spike-triggered neural characterization. J Vis. 2006; 6:484–507. [PubMed: 16889482]

**25. deCharms RC, Blake DT, Merzenich MM. Optimizing sound features for cortical neurons. Science. 1998; 280:1439–1443. The first demonstration of complex and inseparable STRFs in the auditory system. [PubMed: 9603734]

*26. Depireux DA, Simon JZ, Klein DJ, Shamma SA. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J Neurophysiol. 2001; 85:1220–1234. The authors demonstrate that many A1 neurons show differences in responses to upward vs. downward drifting ripple stimuli. [PubMed: 11247991]

27. Schnupp JW, Mrsic-Flogel TD, King AJ. Linear processing of spatial cues in primary auditory cortex. Nature. 2001; 414:200–204. [PubMed: 11700557]

28. Fritz JB, Elhilali M, Shamma SA. Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. J Neurosci. 2005; 25:7623–7635. [PubMed: 16107649]

29. Klein DJ, Simon JZ, Depireux DA, Shamma SA. Stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex. J Comput Neurosci. 2006; 20:111–136. [PubMed: 16518572]

**30. Bar-Yosef O, Nelken I. The effects of background noise on the neural responses to natural sounds in cat primary auditory cortex. Front Comput Neurosci. 2007; 1:3. This study demonstrates highly nonlinear auditory responses that are specifically tuned to the acoustic natural background, suggesting that A1 already participates in auditory source segregation. [PubMed: 18946525]

31. Bar-Yosef O, Rotman Y, Nelken I. Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. J Neurosci. 2002; 22:8619–8632. [PubMed: 12351736]

32. Nelken I, Rotman Y, Bar Yosef O. Responses of auditory-cortex neurons to structural features of natural sounds. Nature. 1999; 397:154–157. [PubMed: 9923676]

33. Machens CK, Wehr MS, Zador AM. Linearity of cortical receptive fields measured with natural sounds. J Neurosci. 2004; 24:1089–1100. [PubMed: 14762127]

34. Sahani, M.; Linden, JF. How Linear are Auditory Cortical Responses?. In: Becker, S.; Thrun, S.; Obermayer, K., editors. Advances in Neural Information Processing Systems 15. MIT Press; 2003. p. 109-116.

*35. Atencio CA, Sharpee TO, Schreiner CE. Hierarchical computation in the canonical auditory cortical circuit. Proc Natl Acad Sci U S A. 2009 This article demonstrates that STRFs are more separable and nonlinearities are less pronounced in granular layers compared to either the supra- or infra-granular layers of auditory cortex. This resolves a long-standing controversy as to why some studies have reported mostly simple and separable STRFs in A1, whereas others have found complex and non-separable STRFs in spectrotemporal space.

36. Atencio CA, Schreiner CE. Columnar connectivity and laminar processing in cat primary auditory cortex. PLoS One. 2010; 5:e9521. [PubMed: 20209092]

37. Ahmed B, Garcia-Lazaro JA, Schnupp JW. Response linearity in primary auditory cortex of the ferret. J Physiol. 2006; 572:763–773. [PubMed: 16497713]

38. Lesica NA, Grothe B. Dynamic spectrotemporal feature selectivity in the auditory midbrain. J Neurosci. 2008; 28:5412–5421. [PubMed: 18495875]

39. Meister M, Berry MJ. The neural code of the retina. Neuron. 1999; 22:435–450. [PubMed: 10197525]

40. Victor J, Shapley R. A method of nonlinear analysis in the frequency domain. Biophys J. 1980; 29:459–483. [PubMed: 7295867]

41. Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network. 2001; 12:289–316. [PubMed: 11563531]

42. Theunissen FE, Sen K, Doupe AJ. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci. 2000; 20:2315–2331. [PubMed: 10704507]

43. Sharpee T, Rust NC, Bialek W. Analyzing neural responses to natural signals: maximally informative dimensions. Neural Comput. 2004; 16:223–250. [PubMed: 15006095]

44. Christianson GB, Sahani M, Linden JF. The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. J Neurosci. 2008; 28:446–455. [PubMed: 18184787]

45. Sharpee TO, Miller KD, Stryker MP. On the importance of static nonlinearity in estimating spatiotemporal neural filters with natural stimuli. J Neurophysiol. 2008; 99:2496–2509. [PubMed: 18353910]

46. Gardner JL, Anzai A, Ohzawa I, Freeman RD. Linear and nonlinear contributions to orientation tuning of simple cells in the cat's striate cortex. Vis Neurosci. 1999; 16:1115–1121. [PubMed: 10614591]

47. Jones JP, Palmer LA. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. J Neurophysiol. 1987; 58:1187–1211. [PubMed: 3437330]

**48. Priebe NJ, Ferster D. Inhibition, spike threshold, and stimulus selectivity in primary visual cortex. Neuron. 2008; 57:482–497. This study demonstrates how a static nonlinearity can sharpen the predicted tuning of a neuron. The resulting values match experimental measurements using gratings. Similar analyses can perhaps be done in the auditory system to resolve differences in tuning predictions based on STRF profiles. [PubMed: 18304479]

*49. Ahrens MB, Linden JF, Sahani M. Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. J Neurosci. 2008; 28:1929–1942. This article introduces a novel methodology for building three dimensional receptive fields in the space of spectrotemporal variations and sound pressure levels. [PubMed: 18287509]

50. Calford MB, Semple MN. Monaural inhibition in cat auditory cortex. J Neurophysiol. 1995; 73:1876–1891. [PubMed: 7623087]

51. Brosch M, Schreiner CE. Time course of forward masking tuning curves in cat primary auditory cortex. J Neurophysiol. 1997; 77:923–943. [PubMed: 9065859]

52. Bartlett EL, Wang X. Long-lasting modulation by stimulus context in primate auditory cortex. J Neurophysiol. 2005; 94:83–104. [PubMed: 15772236]

53. Nelken I, Versnel H. Responses to linear and logarithmic frequency-modulated sweeps in ferret primary auditory cortex. Eur J Neurosci. 2000; 12:549–562. [PubMed: 10712634]

54. Nelken I, Fishbach A, Las L, Ulanovsky N, Farkas D. Primary auditory cortex of cats: feature detection or something else? Biol Cybern. 2003; 89:397–406. [PubMed: 14669020]

55. Atencio CA, Sharpee TO, Schreiner CE. Cooperative nonlinearities in auditory cortical neurons. Neuron. 2008; 58:956–966. [PubMed: 18579084]

*56. Pienkowski M, Eggermont JJ. Nonlinear cross-frequency interactions in primary auditory cortex spectrotemporal receptive fields: a Wiener-Volterra analysis. J Comput Neurosci. 2010; 28:285–303. The authors demonstrate the persistence of selectivity to combinations of frequencies using temporally dense stimuli. [PubMed: 20072806]

**57. Heil P, Neubauer H. A unifying basis of auditory thresholds based on temporal summation. Proc Natl Acad Sci U S A. 2003; 100:6151–6156. This study provides evidence for precise spike timing throughout the auditory system. [PubMed: 12724527]

58. Las L, Stern EA, Nelken I. Representation of tone in fluctuating maskers in the ascending auditory system. J Neurosci. 2005; 25:1503–1513. [PubMed: 15703404]

59. Joris PX, Schreiner CE, Rees A. Neural processing of amplitude-modulated sounds. Physiol Rev. 2004; 84:541–577. [PubMed: 15044682]

60. Heil P. Auditory cortical onset responses revisited. I. First-spike timing. J Neurophysiol. 1997; 77:2616–2641. [PubMed: 9163380]

61. Phillips DP. Effect of tone-pulse rise time on rate-level functions of cat auditory cortex neurons: excitatory and inhibitory processes shaping responses to tone onset. J Neurophysiol. 1988; 59:1524–1539. [PubMed: 3385471]

62. Wallace MN, Coomber B, Sumner CJ, Grimsley JM, Shackleton TM, Palmer AR. Location of cells giving phase-locked responses to pure tones in the primary auditory cortex. Hear Res. 2010; 274:142–151. [PubMed: 20630479]

63. De Ribaupierre F, Goldstein MH Jr, Yeni-Komshian G. Cortical coding of repetitive acoustic pulses. Brain Res. 1972; 48:205–225. [PubMed: 4645205]

64. Schreiner CE, Urbas JV. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. Hear Res. 1988; 32:49–63. [PubMed: 3350774]

**65. Elhilali M, Fritz JB, Klein DJ, Simon JZ, Shamma SA. Dynamics of precise spike timing in primary auditory cortex. J Neurosci. 2004; 24:1159–1172. This study demonstrated that slow STRF dynamics are often compatible with precise spike timing in the same neurons. [PubMed: 14762134]

66. Kayser C, Logothetis NK, Panzeri S. Millisecond encoding precision of auditory cortex neurons. Proc Natl Acad Sci U S A. 2010; 107:16976–16981. [PubMed: 20837521]

67. Lemus L, Hernandez A, Romo R. Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. Proc Natl Acad Sci U S A. 2009; 106:9471–9476. [PubMed: 19458263]

68. Liu Y, Qin L, Zhang X, Dong C, Sato Y. Neural correlates of auditory temporal-interval discrimination in cats. Behav Brain Res. 2011; 215:28–38. [PubMed: 20599565]

69. Scott BH, Malone BJ, Semple MN. Transformation of temporal processing across auditory cortex of awake macaques. J Neurophysiol. 2011; 105:712–730. [PubMed: 21106896]

70. Shih JY, Atencio CA, Schreiner CE. Improved stimulus representation by short interspike intervals in primary auditory cortex. J Neurophysiol. 2011; 105:1908–1917. [PubMed: 21307320]

71. Bartlett EL, Wang X. Neural representations of temporally modulated signals in the auditory thalamus of awake primates. J Neurophysiol. 2007; 97:1005–1017. [PubMed: 17050830]

72. Liang L, Lu T, Wang X. Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. J Neurophysiol. 2002; 87:2237–2261. [PubMed: 11976364]

73. Lu T, Liang L, Wang X. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat Neurosci. 2001; 4:1131–1138. [PubMed: 11593234]

74. Yin P, Johnson JS, O'Connor KN, Sutter ML. Coding of amplitude modulation in primary auditory cortex. J Neurophysiol. 2011; 105:582–600. [PubMed: 21148093]

75. Bizley JK, Walker KM, King AJ, Schnupp JW. Neural ensemble codes for stimulus periodicity in auditory cortex. J Neurosci. 2010; 30:5078–5091. [PubMed: 20371828]

76. Imaizumi K, Priebe NJ, Sharpee TO, Cheung SW, Schreiner CE. Encoding of temporal information by timing, rate, and place in cat auditory cortex. PLoS One. 2010; 5:e11531. [PubMed: 20657832]

**77. Kayser C, Montemurro MA, Logothetis NK, Panzeri S. Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. Neuron. 2009; 61:597–608. Using information theoretic techniques, the authors demonstrated that multiple temporal codes operate concurrently in the auditory cortex. They found that nested codes where spike-train patterns were measured together with the phase of ongoing local field potentials not only were the most informative, but were also robust to noise. [PubMed: 19249279]

78. Samengo I, Montemurro MA. Conversion of phase information into a spike-count code by bursting neurons. PLoS One. 2010; 5:e9669. [PubMed: 20300632]

79. Panzeri S, Brunel N, Logothetis NK, Kayser C. Sensory neural codes using multiplexed temporal scales. Trends Neurosci. 2010; 33:111–120. [PubMed: 20045201]

80. Atencio CA, Schreiner CE. Laminar diversity of dynamic sound processing in cat primary auditory cortex. J Neurophysiol. 2010; 103:192–205. [PubMed: 19864440]

81. Marr D, Hildreth E. Theory of edge detection. Proc R Soc Lond. 1980; b207:187–217. [PubMed: 6102765]

82. Sahani, M.; Linden, JF. Evidence Optimization Techniques for Estimating Stimulus-Response Functions. In: Becker, S.; Thrun, S.; Obermayer, K., editors. Advances in Neural Information Processing Systems 15. MIT Press; 2003. p. 301-308.

83. Fitzgerald JD, Sincich LC, Sharpee TO. Minimal models of multidimensional computations. PLoS Comput Biol. 2011; 7:e1001111. [PubMed: 21455284]

84. Rowekamp RJ, Sharpee TO. Analyzing multicomponent receptive fields from neural responses to natural stimuli. Network: Comp in Neural Systems. 2011 in press.
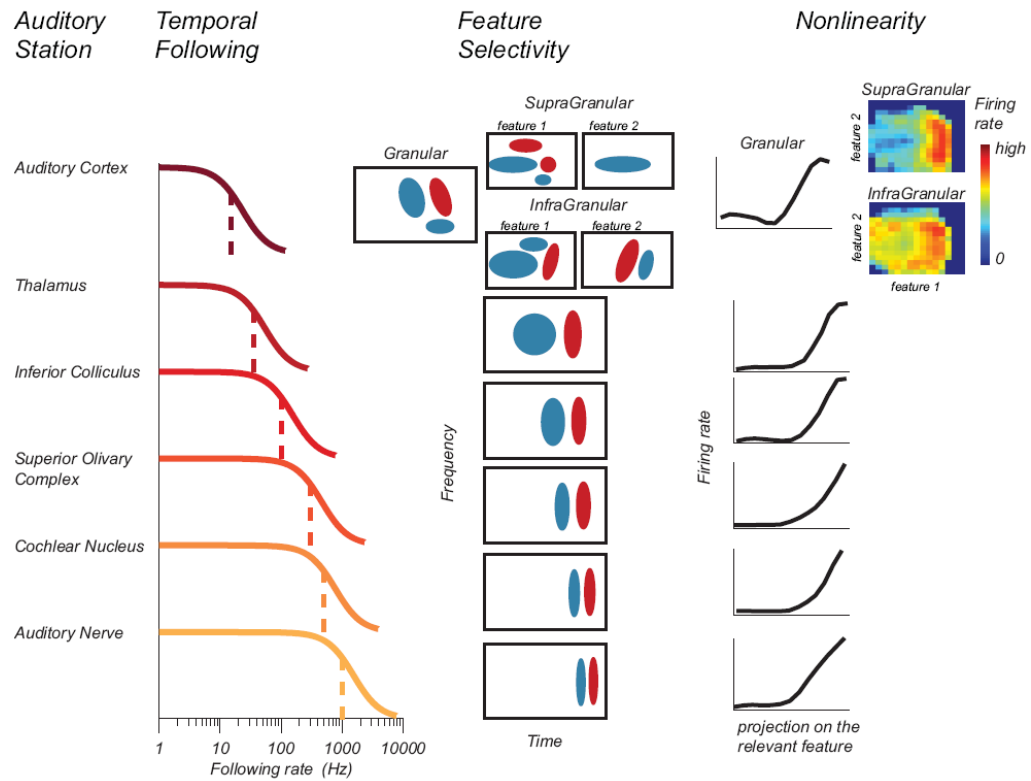
Figure 1. A schematic of the hierarchical transformation of feature selectivity in the auditory system

Columns from left to right are: name of the auditory station, the range over which neurons can follow the rate of temporal changes in auditory stimuli, types of relevant auditory features (positive and negative values are denoted with red and blue), and nonlinearities that describe the neural firing rate as a function of stimulus projections on the relevant stimulus features. In supra- and infra-granular layers of A1, at least two features become relevant (heat map shows nonlinearity with respect to two relevant components). In addition, relevant features become more spectrotemporally inseparable with the synaptic distance from the input A1 layer.
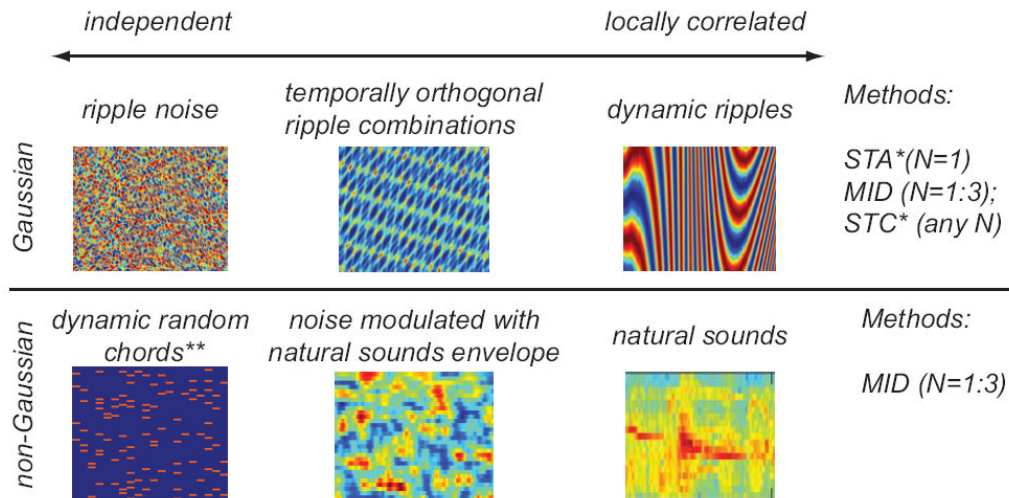
Figure 2. A summary of stimuli often used to characterize feature selectivity of auditory neurons, the stimulus statistical properties, and methods currently available

For independent Gaussian stimuli, relevant stimulus features can be computed as spike-triggered average (STA) or by diagonalizing the spike-triggered covariance (STC) matrix. STA yields only one of the relevant features, whereas STC yields all of N them. *If stimuli are Gaussian but globally correlated, then both methods can still be used, provided the features are normalized by the second-order stimulus covariance matrix [41,42]. If stimuli are non-Gaussian (or equivalently exhibit correlations between more than two points in the spectrotemporal space, as in the case of natural sounds), then relevant features can be found as those that account for maximal amount of information in the neural response. The process of searching for maximally informative dimensions (MID) corrects for stimulus correlations of any order, works with flexible nonlinearities, can be extended to multiple stimulus features (currently up to N=3 features can be estimated [84]). **Although dynamic random chords stimuli are often non-Gaussian, these effects often become negligible as a result filtering transformations within the auditory system following the rules of the central limit theorem in the absence of correlations. This allows the use of STA and STC with random chord stimuli.