Published in final edited form as:

Curr Opin Neurobiol. 2014 April ; 25: 164–168. doi:10.1016/j.conb.2014.01.011.

Adaptive Auditory Computations

Shihab Shamma^{1,2} and Jonathan Fritz¹

¹Department of Electrical and Computer Engineering Institute for Systems Research University of Maryland College park

²Department of Cognitive Studies Ecole Normale Supérieure, Paris

Abstract

The auditory system analyses acoustic signals, extracting their perceptual attributes, and exploiting them to navigate complex auditory environments. While many of the basic transformations that give rise to the early auditory representations are well studied and understood, little is known about the latter cognitive functions that bind, organize, and give meaning to them. They include the ability to attend to, segregate, and track one of many sound sources, to learn its identity, commit it to memory, robustly recognize it, and utilize it to make decisions. This review hints at the profound adaptive influences and contextual effects induced by cognitive functions during these behaviors, and the need for robust tractable mathematical models to understand them.

Introduction

Auditory processing of acoustic stimuli shares much with other sensory modalities in the way it analyses and transforms signals, extracts features, and maps them into neural responses for further utilization in processing leading to perception. Understanding the mechanisms underlying these processes requires an appreciation of how the basic perceptual attributes of hearing are represented throughout the auditory pathway. But equally important are the cognitive aspects of hearing. That is, how these acoustic attributes are exploited to navigate complex auditory scenes in which it is necessary to consider a wide range of contextual information, to form robust integrated percepts of individual sources (incorporating current and prior knowledge), and to actively interact with them.

Mathematical modeling of auditory processing has long dealt with the representation of some of the basic perceptual attributes of pitch, loudness, timbre, and spatial location, and there are many excellent detailed reviews of such models listed later. Instead, our emphasis here is on reviewing modeling efforts inspired by recent psychoacoustic and behavioral physiological experimentation. They describe the effects on auditory responses of long-term changes (e.g., developmental and hearing impairments), of short-term adaptive mechanisms (e.g., as in contextual cues), and of cognitive functions (e.g., attention, learning, memory,

© 2014 Elsevier Ltd. All rights reserved

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

behavioral task rules and objectives) that strongly influence the representation of sound in the auditory system.

Peripheral and Early Auditory Processing

Despite decades of mathematical modeling and experimentation, there is still no unanimity on the basic details of the analyses of sound in the cochlea and subsequent midbrain and thalamic auditory nuclei. For instance, it is still debated whether the phase-locked responses (time-fine-structure, or TFS) in the auditory nerve are essential for speech intelligibility in quiet and noise or if the envelopes (average firing rates) suffice [1,2], how pitch is extracted and represented, how sound loudness is encoded [3], how inter-aural disparities and pinna cues are measured and mapped to spatial locations [4], or how spectral shape and dynamics (timbre) are extracted and maintained up to the cortex [5].

While some cochlear processes appear to be susceptible to feedback control [55], it is largely the case that top-down influences due to cognitive functions (such as attention or engagement in a behavioral task) only modestly alter midbrain auditory processing [6]. However, these lower-level representations do exhibit longer-term plasticity that reflects developmental changes, as well as learning and experience [7]. Given the extended timescales of these changes, most models have not attempted to capture the dynamics but rather the end results of these slow changes [8,9]. Direct projections from the auditory cortex to the midbrain and thalamus are substantial providing a substrate for significant feedback and response modulation. Nevertheless, the functional role of the massive descending corticocollicular projections remains uncertain, except in relearning where it has been shown that the integrity of the descending pathway to the inferior colliculus is critical for learning-induced localization plasticity [10]. By contrast, recordings and mathematical models of thalamic function have been extensive, postulating a critical role as the attentionally regulated gateway of sensory information to the cortex [11], or as the key participant in the generation of various thalamocortical oscillatory rhythms [12].

There are a host of interesting adaptive mechanisms that are rapid (hundreds of milliseconds) and automatic, i.e., can occur in the absence of cognitive control, as they have been shown in anesthetized animals at various stages of the auditory pathway. A hallmark of these mechanisms is the interdependency they create among responses to sound sequences, and which can be viewed as “contextual effects”. Two such well-characterized mechanisms are the *adaptive coding and gain control* [13,14], and *stimulus-specific adaptation* or SSA [15,16]. Both arise as early as the inferior colliculus, are inherited and elaborated in higher stages and/or independently computed at higher levels, and appear to serve complementary functions. Models of adaptive coding have convincingly demonstrated how neurons can relatively quickly adapt their response sensitivity so as to reflect the dynamic range of the input stimuli, which may in turn be due to a host of stimulus parameters, such as its level, inter-aural delays, or spectrotemporal properties [13]. The second mechanism, SSA, exhibits the opposite effects: it is defined as a reduction in the response to a common sound that does not generalize, or only partially generalizes, to other, rare stimuli, whose response is enhanced. Thus, SSA in effect enhances the responses to *novel or deviant*, rather than to ongoing stimuli. It has a more complex nature than simple adaptation in that responses of a

neuron do not generally habituate to all inputs, but rather only to a particular stimulus as if the habituation is taking place at specific input terminals and not at others [15].

Primary Auditory Cortical Processing and Representations

Over the last decade, much has been learned about the representation of complex sounds in the primary auditory cortex, partly through the use of linear and nonlinear models of spectrotemporal receptive fields [17–19], and partly through more ethological approaches [20–22]. These findings have inspired cortical models for the analysis of temporal modulations typical of speech and other natural sounds [5]. More elaborate models have also been proposed that combine the temporal analysis with a joint analysis of the spectral modulations of the sound [23,24]. This joint spectrotemporal analysis effectively performs a double wavelet transform of the auditory spectrogram revealing and capturing efficiently the distinctive patterns of modulations of speech, music, and other complex environmental sounds. For example, “cortical features” have been shown to be effective in numerous applications ranging from speech intelligibility assessment [26], to hearing impairments and hearing aids, to the detection of speech, speaker identification, speech recognition [27,28], and musical timbre classification [29]. Most enlightening, however, have been the mathematical analyses of closely related constructs employing ‘cascaded wavelet transforms’ that have explained the reasons underlying the effectiveness of these cortical features [30]. For instance, unlike Fourier modes, wavelet coefficients exhibit small modifications to correspondingly small deformations of the underlying signal, which is a key property for classification (i.e., such a representation is “Lipschitz stable” to time and frequency deformations of the audio waveforms). Furthermore, the cascade of wavelet transforms and non-linearities builds a sufficiently rich locally invariant representation which captures stable high order signal statistics needed to discriminate non-Gaussian signals with intermittent behavior as in audio. One can prove that the whole signal energy is then transformed into stable and locally invariant descriptors of multiple orders [30].

Sensory cortical processing acquires its characteristic effectiveness from a host of unique mechanisms that have themselves been the subject of many mathematical treatments to explain and abstract their function. For instance, a physiological property of thalamic inputs into the cortex is “synaptic depression”, a rapid suppression of input gain that severely limits the ability of cortical responses to track input modulation rates beyond about 10–20 Hz [31]. Mathematical models have revealed important functional advantages of this synaptic mechanism, ranging from “divisive normalization” in image processing [32] to suppression of noise and reverberation in auditory responses [33], to multiscale temporal modulation analysis [5], to the inter-dependency of responses across successive stimuli [34].

Rapid Cortical plasticity

Unlike earlier stages, auditory cortical responses are highly adaptable to the behavioral context of the animal. Cognitive functions such as deciding to approach a reward or avoid danger upon hearing specific learned sounds (a combination of learned association, discrimination and decision-making), attending to one of many acoustic sources (selective attention and scene analysis), recognizing a learned acoustic pattern (long-term memory), or

maintaining it in short-term memory buffer in the presence of distractors (working memory), can profoundly and quite rapidly influence cortical responses and hence its representation of sound. For example, a variety of observed adaptations in neural responses and receptive fields can be understood and modeled as serving to maximize a performance objective in an auditory task [35,36]. In some cases, the effects are evident as rapid receptive field plasticity that maximizes the difference between responses to a warning sound versus those of the safe sounds [37,38]. In tasks involving prediction, response changes facilitate detection of an expected sound [39]. Finally, when an animal discriminated the timbre and pitch of learned artificial speech stimuli, cortical response changes reflected more the *decision* of the animal rather than its perception of the timbre or pitch of the task stimuli [40]. These task-related effects of behavior are often greater in secondary auditory areas than in primary auditory areas [57].

These rapid adaptive effects may well be related to some of the classic “contextual effects”. In speech perception, they are often described as “coarticulatory effects” in which preceding sounds influence strongly the perception of current syllables and phonemes. Other dramatic examples of contextual effects that induce their effects over relatively long periods of time (necessitating a model of short-term memory) include the Shepherd tone percepts and contrast enhancement [41]. A final example is the active engagement of infants in perceiving their linguistic environment and adapting to the statistics of the phonemes of their mother tongue [42]. **Mechanisms underlying the persistence of these “priors” in the nervous system range from the volatile (e.g., synaptic depression at the thalamocortical synapse [xx]), to long-term retention exceeding seconds, minutes, and more. Elegant models of these processes have considered the contradictory demands of flexible rapid memory acquisition versus maintaining its stability against further activity and experience, necessitating a cascade of mechanisms with multiple timescales and states** [43].

Auditory Streaming and segregation

The critical role of rapid plasticity of cortical responses comes into full view when one considers real-life listening in which animals have to disentangle and interpret signals emanating from multiple simultaneous sources, and then attend and react to one. Psychoacoustic studies of these feats inspired numerous computational models to explain and replicate this ability, originally known as the *Cocktail Party Problem* and later as *auditory scene analysis* [44]. More recently, physiological recordings coupled with behavioral contexts and computational analyses in humans and animals have suggested mechanisms and constraints on what constitutes a biologically plausible model of auditory segregation or streaming. For instance, when humans attend to a speaker, they seem to bring to the fore the physiological signals of the target while strongly suppressing those of the background [45–48]. They are able to do this in a fraction of a second, require no training on the target voices or their words, and can effortlessly integrate multimodal cues to accomplish this task. A recent computational model of these capabilities relies on the concept of temporal coherence to explain how disparate features of a single source could be both segregated from those of other sources, while simultaneously integrated (or bound) to other features of the same source [49,50]. *Temporal coherence* posits that highly correlated

responses of neuronal populations representing different features (e.g., pitch, timbre, location) are bound together as one perceptual ‘object’. Otherwise temporally *incoherent* responses are perceived as emanating from segregated sources. Physiological and psychoacoustic evidence for this principle is accumulating and may help guide the development of new biologically plausible, yet still effective models of auditory scene analysis [51–53].

Finally, another relevant phenomenon to the understanding of real-world sensory processing is that of *bistable perception* of ambiguous stimuli. Natural environments are inherently noisy and full of competing complex sounds and images with numerous features. It is therefore expected that such complex scenes often lead to multiple conflicted perceptual resolutions, often referred to as “illusions”. A classic auditory case in point is the bistable percept of two alternating tones as one or two melodies or “streams” [54]. Mathematical models that describe these ambiguous sensory percepts have shed considerable light on the possible role of attention in perception, on the mechanisms and dynamics underlying auditory scene analysis [55], on how top-down modulation can disambiguate speech depending upon context, and on the neural substrates involved in all these processes.

Into the future

Research in neuroscience in recent years has lead to many new discoveries and insights into the neural bases of cognitive functions such as attention, learning, memory, and decision-making, functions that are essential in auditory perception. One area of cognitive control is the prefrontal cortex, a region hypothesized to be the source of top-down influences that sculpt sensory responses in accordance with the demands and future predictions of ongoing behavioral tasks [56]. Computational models of auditory processing must therefore consider how these cortical functions shape perception by adapting online sensory receptive fields and auditory representations to satisfy behavioral objectives and maintain robust comprehension. Computational models of auditory perception, for example, should be able to continuously adapt to changing acoustic environments, from the highly reverberant environment of a hall, to the noisy street scene, to the interfering signal of a simultaneous speaker. On longer time-scales, they must also “learn” to anticipate incoming acoustic stimuli and maintain a robust performance in these dynamic environments, essentially exhibiting the hallmarks of “expert” listening, much like the heightened sensitivities of a musician improvising in a jazz ensemble or a bird alertly listening for the sounds of mates, offspring, competitors or predators in a complex and changing forest soundscape.

Acknowledgment

This work has been primarily funded by an NIH grant **R01 DC007657** and an Advanced **ERC 295603** (to SS)

References

1. Moore, Brian CJ. The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*. 2008; 9(4):399–406. [PubMed: 18855069]

2. Shamma S, Christian L. On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. *The Journal of the Acoustical Society of America*. 2013; 133:2818. [PubMed: 23654388]
3. Marozeau, Jeremy. Models of Loudness. In: Florentine, M.; Popper, A.; Fay, R., editors. *Loudness*. Springer; New York: 2011. p. 261-284.
4. McAlpine D, Grothe B. Sound localization and delay lines – do mammals fit the model? *Trends in neurosciences*. 2003; 26(7):347–350. [PubMed: 12850430]
5. Chi, Taishih; Ru, Powen; Shamma, Shihab A. Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am*. 2005; 118:887. [PubMed: 16158645]
6. Guinan, John J, Jr. Cochlear efferent innervation and function. *Current opinion in otolaryngology & head and neck surgery*. 2010; 18(5):447. [PubMed: 20717032]
7. Suga, Nobuo. *The Auditory Cortex*. Springer; US: 2011. Corticofugal Modulation and Plasticity for Auditory Signal Processing; p. 513-533.
8. Woolley, Sarah; Hauber, Mark E.; Theunissen, Frederic E. Developmental experience alters information coding in auditory midbrain and forebrain neurons. *Developmental neurobiology*. 2010; 70(4):235–252. [PubMed: 20039264]
9. Slee, Sean J.; Young, Eric D. Linear processing of interaural level difference underlies spatial tuning in the nucleus of the brachium of the inferior colliculus. *The Journal of Neuroscience*. 2013; 33(9): 3891–3904. [PubMed: 23447600]
10. Bajo VM, Nodal FR, Moore DR, King AJ. The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nature Neuroscience*. 2010; 13:253–260.
11. Baluch, Farhan; Itti, Laurent. Mechanisms of top-down attention. *Trends in neurosciences*. 2011; 34(4):210–224. [PubMed: 21439656] * This is a review of recent findings concerning the mechanisms that underly attention. They include the anatomical pathways and brain structures most directly involved as well as their cortical and subcortical targets, and a summary of the effects of microstimulation and ablation experiments that simulate attention or its absence. Most valuable is the computational models presented in the context of the physiological and psychophysical data which help draw connections among these three approaches and help point the way to future investigations.
12. Wang, Xiao-Jing. Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological reviews*. 2010; 90(3):1195–1268. [PubMed: 20664082]
13. Rabinowitz, Neil C.; Willmore, Ben DB.; Schnupp, Jan WH.; King, Andrew J. Contrast gain control in auditory cortex. *Neuron*. 2011; 70(6):1178–1191. [PubMed: 21689603] ** This paper offers extensive data and analyses of a remarkable example of rapid adaptive gain control of cortical spectrotemporal selectivity, driven primarily by local and global contrast cues. This change is thought to enhance the efficiency of cortical encoding of sounds rich in spectrotemporal structure such as natural vocalizations.
14. Robinson, Benjamin Louis; McAlpine, David. Gain control mechanisms in the auditory pathway. *Current opinion in neurobiology*. 2009; 19(4):402–407. [PubMed: 19665367]
15. Nelken, Israel. Biological Cybernetics. Stimulus-specific adaptation: experiments and models. in press** A lucid account is presented of the fundamental properties of *stimulus specific adaptation*, as well as of the subtle qualities that distinguishes this phenomenon from closely related effects such as adaptation of neural responses strictly due to activation or due to adaptive gain control mechanisms.
16. Thomas J, Morse C, Kishline L, O'Brien-Lambert A, Simonton A, Miller K, Covey E. Stimulus-specific adaptation in specialized neurons in the inferior colliculus of the big brown bat, *Eptesicus fuscus*. *Hearing research*. 2012; 291:34–40. [PubMed: 22743044]
17. Christianson G, Björn Maneesh Sahani, Linden Jennifer F. The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. *The Journal of Neuroscience*. 2008; 28(2):446–455. [PubMed: 18184787]
18. Theunissen, Frédéric E.; Elie, Julie E. Population Code, Noise Correlations, and Memory. *Neuron*. 2013; 78(2):209–210. [PubMed: 23622058]
19. Elhilali, Mounya; Simon, Jonathan Z.; Fritz, Jonathan B.; Shamma, Shihab. A Linear Systems View to the Concept of STRFs in *Modern methods in AI*. Depireux, Didier, editor. Nova

Publishers; * An overall review of linear models and methods used in estimating the spectrotemporal response fields in auditory cortex. The application of these response measures in different tasks is discussed, as well as their use in exploring the functional organization of the auditory cortex.

20. Wang, Xiaojin; Walker, Kerry MM. Neural mechanisms for the abstraction and use of pitch information in auditory cortex. *The Journal of Neuroscience*. 2012; 32(39):13339–13342. [PubMed: 23015423]
21. Theunissen, Frédéric E.; Mouterde, Solveig; Mathevon, Nicolas. A neuroethological analysis of the information in propagated communication calls. *Proceedings of Meetings on Acoustics*. 2013; 19
22. Centanni, Tracy Michelle; Engineer, Crystal T.; Kilgard, Michael P. Cortical speech-evoked response patterns in multiple auditory fields are correlated with behavioral discrimination ability. *Journal of neurophysiology*. 2013
23. David, Stephen V., et al. Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli. *The Journal of Neuroscience*. 2009; 29(11):3374–3386. [PubMed: 19295144]
24. Mesgarani N, David SV, Fritz JB, Shamma SA. Phoneme representation and classification in primary auditory cortex. *The Journal of the Acoustical Society of America*. 2008; 123:899. [PubMed: 18247893]
25. Atencio, Craig A.; Schreiner, Christoph E. Spectrotemporal Processing in Spectral Tuning Modules of Cat Primary Auditory Cortex. *PloS one*. 2012; 7(2):e31537. [PubMed: 22384036]
26. Elhilali, Mounya; Chi, Taishih; Shamma, Shihab A. A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. *Speech communication*. 2003; 41(2):331–348.
27. Sivaram GSVS, Nemala SK, Mesgarani N, Hermansky H. Data-driven and feedback based spectro-temporal features for speech recognition. *Signal Processing Letters, IEEE*. 2010; 17(11): 957–960.
28. Mesgarani, Nima; Slaney, Malcolm; Shamma, Shihab A. Discrimination of speech from nonspeech based on multiscale spectro-temporal modulations. *Audio, Speech, and Language Processing, IEEE Transactions on*. 2006; 14(3):920–930.
29. Patil K, Pressnitzer D, Shamma S, Elhilali M. Music in our ears: The biological bases of musical timbre perception. *PLoS computational biology*. 2012; 8(11):e1002759. [PubMed: 23133363]
30. Andén, Joakim; Mallat, Stéphane. IEEE transactions on Signal Proc. Deep Scattering Spectrum. in press** An interesting mathematical formulation of the question of why spectrotemporal modulations in acoustic signals are useful and how they can be represented. The paper offers careful analyses and a unique perspective on what the various signal transformations contribute to its robustness and enhancement, especially when used subsequently in speech recognition and separation applications.
31. Fortune ES, Rose GJ. Short-term synaptic plasticity as a temporal filter. *Trends in neurosciences*. 2001; 24(7):381–385. [PubMed: 11410267]
32. Carandini M, Heeger DJ. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*. 2011; 13(1):51–62.** This is a clear and forceful account in support of the view that the brain relies on a set of canonical neural computations, repeating them across brain regions and modalities to apply similar operations to different problems. While the authors rely primarily on insights and analyses from responses in visual cortex, the same formulations and arguments hold well for the auditory modality
33. Moore R, Channing Tyler Lee, Theunissen Frédéric E. Noise-invariant Neurons in the Avian Auditory Cortex: Hearing the Song in Noise. *PLoS computational biology*. 2013; 9(3):e1002942. [PubMed: 23505354]
34. Klampfl S, David SV, Yin P, Shamma SA, Maass W. A quantitative analysis of information about past and present stimuli encoded by spikes of A1 neurons. *Journal of Neurophysiology*. 2012; 108(5):1366–1380. [PubMed: 22696538]
35. Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature neuroscience*. 2003; 6(11):1216–1223.
36. Scheich, H.; Brosch, M. Task related activation of auditory cortex. In: Cohen, Yale E.; Popper, Arthur N.; Fay, Richard R., editors. *Neural correlates of auditory cognition*. Vol. Vol. 45.

Springer; 2013. * A detailed account of how elements of different tasks such as rewards, punishment, environment, conditioned, and unconditioned responses can fundamentally influence auditory responses. This is a valuable report for anybody contemplating behavioral physiological experiments as it demonstrates the importance of taking into consideration myriad factors related to more to the decision actions and contextual information of the task, and less to the sensory stimuli themselves.

37. Mesgarani N, Fritz J, Shamma S. A computational model of rapid task-related plasticity of auditory cortical receptive fields. *Journal of computational neuroscience*. 2010; 28(1):19–27. [PubMed: 19711179]
38. David, Stephen V.; Fritz, Jonathan B.; Shamma, Shihab A. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proceedings of the National Academy of Sciences*. 2012; 109(6):2144–2149.* Examples are presented of how cortical responses and receptive fields of even the simplest tonal stimulus can be modulated up or down purely according to their meaning to the animal (as a warning or a safe sound).
39. Jaramillo, Santiago; Zador, Anthony M. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nature neuroscience*. 2010; 14(2):246–251.
40. Bizley JK, Walker KM, Nodal FR, King AJ, Schnupp JW. Auditory Cortex Represents Both Pitch Judgments and the Corresponding Acoustic Cues. *Current Biology*. 2013
41. Englitz, B.; Akram, S.; David, SV.; Chambers, C.; Pressnitzer, D.; Depireux, D.; Fritz, JB.; Shamma, SA. Basic Aspects of Hearing. Springer; New York: 2013. Putting the Tritone Paradox into Context: Insights from Neural Population Decoding and Human Psychophysics; p. 157–164.
42. Ramus F, Peperkamp S, Christophe A, Jacquemot C, Kouider S, Dupoux E. A psycholinguistic perspective on the acquisition of phonology. *Laboratory phonology*. 2010; 10:311–340.
43. Fusi S, Drew P, Abbott L. Cascade models of synaptically stored memories. *Neuron*. 2005; 45(4): 599–611. [PubMed: 15721245]
44. Bregman, Albert S. Auditory scene analysis: The perceptual organization of sound. MIT press; 1994.
45. Ding, Nai; Simon, Jonathan Z. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*. 2012; 109(29):11854–11859.** A striking illustration of speech segregation in humans is evidenced by responses from noninvasive MEG imaging. The data show that MEG responses to a speech mixture become rapidly dominated by the envelope of the speech of the attended speaker. Furthermore, the representation of this extracted envelope is stable and robust in the presence of noise, lending credence to the notion that attention effectively induces the percept of an “auditory object”.
46. Pasley, Brian N., et al. Reconstructing speech from human auditory cortex. *PLoS biology*. 2012; 10(1):e1001251. [PubMed: 22303281]
47. Mesgarani, Nima; Chang, Edward F. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*. 2012; 485(7397):233–236. [PubMed: 22522927] ** This report offers a fascinating look at ECoG auditory responses as human subjects switch attention to different target speakers in the same speech mixture. The results demonstrate the remarkable rapidity with which neuronal selectivity in the cortex changes to reflect the objectives of the task, and offers hints at the underlying mechanisms involved.
48. Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Schroeder CE. Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a “Cocktail Party”. *Neuron*. 2013; 77(5):980–991. [PubMed: 23473326]
49. Shamma, S.; Elhilali, M.; Ma, L.; Micheyl, C.; Oxenham, AJ.; Pressnitzer, D.; Xu, Y. Basic Aspects of Hearing. Springer; New York: 2013. Temporal Coherence and the Streaming of Complex Sounds; p. 535–543.
50. Shamma, Shihab A.; Elhilali, Mounya; Micheyl, Christophe. Temporal coherence and attention in auditory scene analysis. *Trends in neurosciences*. 2011; 34(3):114–123. [PubMed: 21196054] ** An article that advocates the opinion that animals segregate sounds emanating from different sources by exploiting the temporal incoherence among their features. The idea also explains the role of attention and memory in sound segregation, and the way feature-binding naturally arises during perception of complex and multimodal stimuli.

51. Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron*. 2009; 61(2):317–329. [PubMed: 19186172]
52. Teki S, Chait M, Kumar S, Shamma S, Griffiths TD. Segregation of complex acoustic scenes based on temporal coherence. *eLife*. 2013; 2
53. Micheyl C, Hanson C, Demany L, Shamma S, Oxenham AJ. Auditory Stream Segregation for Alternating and Synchronous Tones. *Journal of Experimental Psychology: Human Perception and Performance*. 2013 2013.
54. Schwartz JL, Grimault N, Hupé JM, Moore BC, Pressnitzer D. Multistability in perception: binding sensory modalities, an overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2012; 367(1591):896–905.* This review contrasts and critiques competing models of the interesting phenomenon of multistable perception in audition. The article adopts the point of view that bistability originates in the competition for features that makeup opposing interpretations of a scene. As such, it is argued that this phenomenon can be used effectively as an indicator of the binding process, and of the memory of different scenes.
55. Winkler, I.; Bohm, TM.; Mill, RW.; Bendixen, A.; Denham, SL. Modeling auditory stream segregation by predictive processes. *Cognitive Infocommunications (CogInfoCom)*, 2012 IEEE 3rd International Conference on; IEEE; Dec. 2012 p. 479-483.
56. Alexander WH, Brown JW. Medial prefrontal cortex as an action-outcome predictor. *Nature neuroscience*. 2011; 14(10):1338–1344.
57. Diamond, David M.; Weinberger, Norman M. Classical conditioning rapidly induces specific changes in frequency receptive fields of single neurons in secondary and ventral ectosylvian auditory cortical fields. *Brain research*. 1986; 372(2):357–360. [PubMed: 3708366]

Highlights

- Models of early auditory processing explain how stable perceptual attributes emerge
- Pre-attentive *Adaptive Coding* and *Stimulus Specific Adaptation* enable dynamic representations
- Rapid plasticity enhances auditory processing in accordance with task objectives
- Perceptual segregation of simultaneous sources is essential in complex auditory scenes
- Auditory models are powerful in speech intelligibility, recognition, and detection