Neuronal responses in auditory cortex show a fascinating mixture of characteristics that span the range from almost perfect copies of physical aspects of the stimuli to extremely complex context-dependent responses. Fast, highly stimulus-specific adaptation and slower plastic mechanisms work together to constantly adjust neuronal response properties to the statistics of the auditory scene. Evidence with converging implications suggests that the neuronal activity in primary auditory cortex represents sounds in terms of auditory objects rather than in terms of invariant acoustic features.

Paradoxical response properties in auditory cortex

Precise and imprecise temporal coding

One of the complexities in auditory cortex is the interplay among multiple time scales that determine the neural responses. For example, cortical neurons respond to some auditory events with stereotypical response bursts at a fixed latency (‘locking’). The variance of the latency of such bursts might be similar to that of peripheral neurons. However, the same neurons may show sluggish responses to other features of the sounds.

Temporal coding is usually tested using repetitive stimuli, such as amplitude-modulated best-frequency tones or click trains. The ability of cortical neurons to phase-lock to repetitive stimuli is usually limited to sub-pitch rates (20-30 Hz) in both anesthetized and awake animals, although in many studies, a minority of the cortical neurons can follow fast repetitive stimuli up to much higher rates. Lu et al. demonstrated the presence of a separate population of neurons that was sensitive to the rate of faster click trains without locking to the clicks.

The most recent evidence for the sluggishness of cortical neurons is based on the spectro-temporal receptive field (STRF) [12–15]. The STRF can be interpreted as the time-frequency distribution of the most efficient stimulus for the neuron, but also as the average response following a short tone burst as a function of the burst frequency and time after the stimulus. Initially, it was hoped that STRFs would uncover relatively complex response properties. However, several recent analyses of large sets of mammalian STRFs concluded that for the majority of
neurons, STRF shapes are rather simple [13,14]. Furthermore, STRFs are often sluggish: the temporal component of the STRF (the average of the STRF at all frequencies, as a function of the delay) is often relatively slow, and the modulation transfer function of cortical neurons, derived as the Fourier transform of the temporal component of the STRF, peaks at 16 Hz in the cat [16]. Most STRFs have the form of differentiators in time and frequency, therefore representing a relatively simple operation, with time constants in the order of 30 ms.

Such slow dynamics of firing would seem to imply relatively slow dynamics for the membrane potential, and therefore, large variance in spike timings [17]. However, spike timing in auditory cortex can be remarkably precise. First spike latencies of cortical neurons show as little variability as they do in the auditory nerve [18]. The same precision of spike timing is seen in the responses to frequency-modulated (FM) sounds covering a large frequency band. In most A1 neurons, FM sounds produce a short burst of spikes when the frequency trajectory crosses the edge of the tuning curve of the neuron. This burst is locked with millisecond precision to the time at which the frequency trajectory reaches the trigger frequency, independent of the velocity of the sweep [17,19,20].

It could be argued that in the above examples, the precise temporal locking occurs only at the onset of the responses. However, Elhilali et al. [21] demonstrated such precision in the steady-state responses to dynamic ripple stimuli. These are stimuli that are composed of a sum of a large number of sine waves densely distributed over a large frequency band that covers the whole response area of the neuron. The amplitudes of these sine waves are modulated by envelopes that vary slowly in both time and frequency. Cortical neurons sometimes locked to the fine structure of the carrier even at rates that were substantially higher (up to 200 Hz) than the 20–30 Hz above which locking to click trains and other repetitive stimuli was essentially absent in cortex [9]. The precise locking to the carrier structure was present although the locking to the slow envelope often disappeared at much slower rates [21].

The conservation of highly precise timing in cortical responses is apparent even in EEG and its magnetic analog, the MEG. Rupp et al. [22] studied the mid-latency components of the MEG responses to short FM sounds. These components are believed to represent the onset responses in primary auditory cortex. The measured mid-latency potentials were modeled as sums of unit responses, presumably evoked by each individual frequency as it occurs in the FM sound, whose temporal sequence was determined by the cochlear activation due to the FM sounds. The cochlear activation depends on the direction of the FM sound and on the directionality of the cochlear traveling wave, which always travels from high to low frequencies [23]. The MEG responses had a directional sensitivity that was reproduced by this simple model. Thus, some of the temporal structure of the cochlear traveling wave is represented all the way to the auditory cortex.

Fishbach et al. [17,24] presented a simple model that accounts quantitatively for tone responses, 2-tone responses, and responses to FM sounds in the auditory cortex. Their model is essentially a differentiator of the temporal envelope in different frequency bands, followed by integration over frequency with frequency-dependent delays. To fit the data, the model required short time constants, in the order of 1–10 ms. Thus, whereas both the precise and the sluggish responses of cortical neurons can be described as differentiations, they operate on the incoming sound at different time scales [25]: the precise responses are sensitive to fast changes (1–10 ms), whereas the sluggish responses are sensitive to slow changes (10–100 ms).

The use of intracellular recordings in vivo in auditory cortex uncovered some of the mechanisms underlying the highly precise cortical responses. Wehr and Zador [33] and Zhang et al. [43] demonstrated the presence of highly balanced excitatory and inhibitory inputs to cortical neurons. The inhibition turns out to be slightly delayed with respect to the excitation, which opens a short window during which neurons can respond, thus enhancing the precision of their firing. Such results are consistent with the Fishbach model [17]. Elhilali et al. [21] suggested the possible involvement of synaptic depression for achieving the seemingly contradictory goals of sluggish yet precise firing.

Linear and non-linear responses in auditory cortex
Several studies have recently suggested a rather linear representation of sounds by neurons in auditory cortex. Most of these studies are based on characterization of auditory neurons by STRFs. Under the assumption that the system integrates signal energy linearly, STRFs can be used to predict the responses of neurons to general stimuli by convolving them with the spectrogram of the stimulus [26,27]. For example, using random chord stimuli, first introduced by deCharms et al. [15], Schnupp et al. [28] estimated STRFs in the ferret auditory cortex and successfully used them to predict the responses to virtual space stimuli. These findings demonstrated that most of the spatial sensitivity of cortical neurons could be attributed to their linear spectral sensitivity. Mrsic-Flogel et al. [29] further strengthened these results by demonstrating that at least some of the maturation of the spatial responses in A1 during development of ferrets can be attributed to the growth of their heads and ears, which transforms juvenile into adult acoustics.
However, under many other conditions, linear models, and in particular STRFs, cannot easily explain cortical responses to complex sounds. Barbour and Wang [30] described a class of neurons that respond maximally for low-contrast stimuli in the auditory cortex of awake marmosets. As the responses of linear systems should increase monotonically with the contrast, such non-monotonic responses cannot be accounted for by linear systems.

Bar-Yosef et al. [31] presented relatively simple natural sounds, tonal bird songs over natural backgrounds, to neurons in AI of halothane-anesthetized cats. They found that almost any modification of these sounds, such as extracting them from the temporal context or cleaning them from their noise background, resulted in large changes in the neural responses. Such effects are qualitatively inconsistent with linear, STRF-based descriptions of cortical neurons, although Bar-Yosef et al. [31] did not perform a quantitative comparison with predictions based on STRFs.

Going a step beyond the study of Bar-Yosef et al. [31], Machens et al. [27] fitted STRFs to the membrane potential trajectories of cortical neurons responding to natural sounds. They found that for some sounds, such models could satisfactorily account for the responses, but generally their predictive power was low.

**Feature detection or something else?**

It seems that depending upon the circumstances, a cortical neuron can choose to be sluggish or precise, linear or non-linear. Thus, the feature sensitivity of a neuron, as determined, for example, by its STRF, cannot be used as an invariant essential characterization of its responses. The multiple time scales at which cortical neurons process sounds provide another argument against a pure role in feature-detection for auditory cortex neurons [25]. Feature detectors are expected to be sensitive to the time scale of the features they represent. However, cortical neurons show sensitivity to sounds on multiple time scales.

Taking this argument to the extreme, it can be hypothesized that cortical neurons are not feature detectors, but something else. In fact, most features whose extraction has been assigned to cortical neurons are already represented subcortically. Thus, using intracellular recordings of the responses to FM sounds, Zhang et al. [4] demonstrated that the excitatory inputs to cortical neurons, presumably arriving from the auditory thalamus, are already selective to the direction of the frequency modulation. Similarly, mechanisms underlying pitch sensitivity can already be demonstrated at the level of the cochlear nucleus [32], and the extraction of periodicity, which could play a part in encoding pitch, has been hypothesized to be complete as early in the pathway as the inferior colliculus (IC), although definitive evidence is still lacking [9,33,34]. The substantial slowing in the phase-locking capability of auditory cortex neurons, that has been argued to be a specialization for analyzing invariant features of animal vocalizations [35], is already found in the auditory thalamus [9]. It is possible to add other important features to this list, such as interaural time and level differences (ITDs and ILDs), which are important for binaural hearing. For example, Shackleton et al. [36] demonstrated that just noticeable differences for interaural time disparities, computed for single neurons in IC, are already consistent with behavioral thresholds.

So what can auditory cortex neurons do, beyond feature detection? Clues to their putative role in auditory perception may arise from studies of adaptation and plasticity, and from the use of complex, natural, and naturalistic auditory scenes containing multiple sound objects.

**Adaptation and plasticity**

The plastic capabilities of auditory cortex have been studied in several preparations on many time scales. Significant changes in electrical and magnetic brain potentials (EEG and MEG) occur during training for the performance of tasks such as the perception of virtual pitch [5] and fine pitch discrimination [37]. Even simple exposure to different auditory environments can substantially change auditory cortical organization and responses: thus, raising rats in an enriched environment increases many measures of cortical responsiveness [38], whereas raising them in constant noise disrupts the tonotopic structure of auditory cortex [39]. In humans, comparisons of musicians and non-musicians show significant changes in brain structure and in responses to sounds, although such studies are weakened by the fact that correlation does not imply causation. Musicians have greater volume of gray matter in auditory areas, motor areas, and in areas related to visuo-spatial processing relative to non-musicians, and these increases were correlated with practice intensity [40]. Schneider et al. [6] have reported some of the most impressive results of this kind. They demonstrated correlations among the amplitude of mid-latency components of the MEG, the gray matter volume of the presumed primary auditory cortex on Heschl’s gyrus, and musical aptitude in groups of professional musicians, amateur musicians, and non-musicians.

Plasticity can be evoked in controlled experimental conditions by classical conditioning and can be mimicked in anesthetized animals by nucleus basalis stimulation [41]. Good reviews of these results by Weinberger and by Suga and Ma have recently appeared [41,42].

Fritz et al. [43] have recently demonstrated some of the most rapid plastic changes in auditory cortex observed in the literature so far. They showed changes in the
At the single neuron level, Malone et al. [48] demonstrated strong adaptation that resulted in context-sensitive responses to time-varying interaural phase differences in the auditory cortex of awake macaques. Such results, already observed in the IC [48], were substantially more pronounced in auditory cortex. Adaptation in brain potentials, in which the N1 component and its magnetic analog, N1m, tend to decrease with repetitive stimulation, has been known for a long time. When such a repetitive stimulus sequence is interrupted by a rare sound, a novel response component, named mismatch negativity (MMN), appears in the evoked responses [44]. This process has been linked to sensory memory and to ‘primitive intelligence’ [44] because it can be evoked when an abstract rule is violated by a rare sound [45]. MMN generators have been localized in or at the vicinity of auditory cortex [46].

At the single neuron level, Malon et al. [47] demonstrated strong adaptation that resulted in context-sensitivity of the responses to time-varying interaural phase differences in the auditory cortex of awake macaques. Such results, already observed in the IC [48], were substantially more pronounced in auditory cortex. Ulanovsky et al. [49**] demonstrated stimulus-specific adaptation in an oddball paradigm, similar to the one used to evoke MMN, in which a stimulus sequence consists of a common stimulus (‘standard’) and a rare stimulus (‘deviant’). Stimulus-specific responses developed within 2–3 presentations of the standard, which represents highly sensitive adaptation with a time scale of a few seconds. Such stimulus-specific adaptation in auditory cortex could underlie the generation of MMN.

Thus, the context-dependence of auditory cortex responses has been observed at time scales of years, months, days, hours, minutes, and seconds, which is approaching the time scale of responses to individual stimulus presentations.

Auditory scene analysis in auditory cortex

Several recent studies, using a variety of techniques, suggest a role for auditory cortex in segregation and grouping of sound components. For example, at the brain potential level, Dyson and Alain [50] reported that the amplitude of the mid-latency potentials increased when a harmonic was mistuned, potentially creating two auditory objects instead of one. Furthermore, the enhanced amplitude was correlated with an increased likelihood of reporting two concurrent auditory objects. Krumholz et al. [51] reported that when a noise burst suddenly acquired pitch without energy transition, a pitch onset potential occurred with a latency that was monotonically related to the pitch period. The source of this potential was localized to primary auditory cortex. Such a potential could also be interpreted as indicating the onset of a new auditory object.

A different aspect of auditory scene analysis was studied by Fishbach et al. [52] using multi-unit responses and local field potentials. They suggested the involvement of auditory cortex in stream segregation. They recorded responses to an alternative sequence of tones, in which the A tone was at best frequency and the B tone was away from best frequency. At slow repetition rates, both tones evoked responses, but at higher repetition rates, the responses to the B tone were selectively suppressed, which resulted in a sequence of responses at half the stimulation rate. This result is similar to the perceptual segregation of such an alternating sequence into two ‘streams’ [53].

At the single neuron level, Fishbach et al. [17,24] argued that onset responses in cortex represent ‘auditory edges’, and as such represent the onset of new objects. They demonstrated that a simple model, essentially a differentiator in time and frequency, can account for a large number of psychophysical and electrophysiological results.

If phasic responses in auditory cortex represent the onset of new objects, then the lower limit of pitch, about 40 Hz [54], could be related to the disappearance of phase locking to periodic sounds by the majority of the neurons in auditory cortex at about that rate, as discussed above. Indeed, the disappearance of such time-locked responses above about 40 Hz would create a perceptual continuity, consistent with the perceptual transition from individual events at each period to the sensation of pitch [54]. Under this interpretation, the inability of cortical neurons to follow fast modulation rates is a feature, rather than a bug.

Using a different approach, Nelken et al. [55] showed that cortical neurons sometimes respond to the combination of a bird chirp and the weak natural background as if they heard only the weak background. In the same vein, Nelken et al. [56,57] found that cortical neurons respond to the combination of a weak tone and strong fluctuating noise with patterns that are similar to their responses to weak tones presented alone. Both examples are interpreted as the results of auditory scene analysis, in which cortical neurons complete the difficult part of the problem by segregating the two components in the mixture and responding to the weak component.

Speculative synthesis and conclusions

Most of the interesting auditory features might already be extracted from the incoming sounds by the level of the IC, which should therefore be considered as the auditory analog of the primary visual cortex (V1). The role of auditory cortex is to organize these features into auditory
objects (Figure 1). To do that, auditory cortex has to use temporal and spectral context at several time scales. The large adaptive and plastic capacity of auditory cortex is used to tune the neural circuits to the statistical regularities of the environment. In fact, it might be difficult or impossible to separate the time scales at which context sensitivity becomes adaptation and plasticity, as also demonstrated in other neural systems [58,59].

As a result, neurons in auditory cortex respond to auditory objects, representing the relevant distinctive features of the objects to which they respond. When only single objects are present, as is the case in most current experimental paradigms, the neurons encode the sounds, often in a simple way, by representing changes in the sounds at multiple time scales. However, when the same sounds are presented in mixtures, the exact nature of the mixture can determine the features to which the neurons respond. This context sensitivity can account for some of the contradictions in the descriptions of the responses of cortical neurons to complex sounds.

Although this review considered responses in primary auditory cortex almost exclusively, the speculations formulated here are consistent with the recent suggestion that there are two streams of processing in the auditory system beyond primary auditory cortex, a ‘what’ pathway and a ‘where’ pathway [60,61]. It is conceivable that primary auditory cortex constructs auditory objects, to which higher auditory areas assign properties such as spatial position (in the ‘where’ pathway) and, for example, phonemic identity (in the ‘what’ pathway) [62,63].

To test these speculations it will be necessary to test auditory cortex with stimuli that will engage mechanisms of auditory scene analysis. In particular, it is the use of mixtures of sounds, in which more than one object is present, which will probably prove crucial to further our understanding of the role of auditory cortex in hearing.

Acknowledgements
Supported by grants from the Israeli Science Foundation (ISF), the German-Israeli Foundation (GIF) and the Volkswagenstiftung.

References and recommended reading
Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


   Please see annotation to Zhang et al. [*].


The authors demonstrate large correlations among the size of human A1, evoked potentials, and musical aptitude.


The authors present a simple model, essentially a differentiator in time and frequency, that is capable of accounting for responses of A1 neurons to a large set of sounds. The model parameters show smoother distribution on the cortical surface than standard experimentally derived parameters, which suggests that the model incorporates some crucial aspects of the structure of such sounds in auditory cortex.


This study demonstrates explicitly that there are at least two separate time scales at which cortical neurons are operating simultaneously in the primary auditory cortex. Cortical neurons are shown to lock to the fine structure of dynamic ripple stimuli while responding sluggishly to the slow temporal envelope.


This study explicitly addresses the issue of linearity of cortical neurons by fitting STRFs to membrane potential responses. Generally, such linear models cannot explain the results satisfactorily.


The authors manipulate the contrast of complex spectral shapes in this study. Neurons with both monotonic and non-monotonic contrast-dependence are demonstrated in the auditory cortex of awake marmosets.


In a tone detection task performed by ferrets, the structure of STRFs is shown to change within minutes in a task-related manner.


