Primary auditory cortex of cats: feature detection or something else?

Israel Nelken, Alon Fishbach, Liora Las, Nachum Ulanovsky, Dina Farkas

Department of Physiology, Hebrew University – Hadassah Medical School, and the Interdisciplinary Center for Neural Computation, The Hebrew University, P.O.Box 12272, Jerusalem 91120, Israel

Abstract. Neurons in sensory cortices are often assumed to be "feature detectors", computing simple and then successively more complex features out of the incoming sensory stream. These features are somehow integrated into percepts. Despite many years of research, a convincing candidate for such a feature in primary auditory cortex has not been found. We argue that feature detection is actually a secondary issue in understanding the role of primary auditory cortex. Instead, the major contribution of primary auditory cortex to auditory perception is in processing previously derived features on a number of different timescales. We hypothesize that, as a result, neurons in primary auditory cortex represent sounds in terms of auditory objects rather than in terms of feature maps. According to this hypothesis, primary auditory cortex has a pivotal role in the auditory system in that it generates the representation of auditory objects to which higher auditory centers assign properties such as spatial location, source identity, and meaning.

Abbreviations A1, primary auditory cortex; MGB, medial geniculate body; IC, inferior colliculus; STRF, spectrotemporal receptive field

1 Feature detection in auditory cortex

Despite a research history going as far back as that of any other sensory cortex (Woolsey and Walzl 1942), auditory cortex still poses a much greater mystery than visual or somatosensory cortices. To understand the problems currently facing research in auditory cortex, it is useful to first delineate the reasons underlying the success in understanding the visual cortex. Four important findings shape our understanding of primary visual cortex (Wurtz and Kandel 2000). First, many single neurons are selective for computationally derived features, e.g., orientation, a selectivity that is absent in the visual thalamus. Second, orientation selectivity is roughly uniform within a cortical column. Third, orientation selectivity varies in a continuous way over the cortical surface, except for a discrete set of discontinuities that have electrophysiological, imaging, and anatomical correlates. Fourth, other visual features, such as direction selectivity, binocular disparity, color, and so on, are also distributed over columns, although not necessarily in as tight an order as orientation selectivity. These findings gave rise to the idea of the cortical hypercolumn, in which all possible features in a small patch of the visual scene are represented (e.g., Bartfeld and Grinvald 1992).

The subcortical auditory system is organized very differently from the subcortical visual system. In the retina, two synapses separate the photoreceptors from the ganglion cells. Similarly, in the auditory system, two synapses separate the hair cells and the neurons of the cochlear nucleus. The analogy between the ganglion cells of the retina and the neurons of the cochlear nucleus is not as farfetched as may initially seem. For example, in both structures, multiple parallel pathways are generated by neurons of different functional classes. In the retina, there are separate functional and anatomical streams arising from X, Y, and W cells (Tessier-Lavigne 2000), whereas in the cochlear nucleus such streams arise from, e.g., primary-like neurons, choppers, and the type IV neurons of the dorsal cochlear nucleus (Rhode and Greenberg 1992). However, from this point on the analogies between the two sensory systems become at best tenuous. Whereas the ganglion cells of the retina project directly to the visual thalamus, in the auditory system many of the parallel processing streams have further stations in which additional acoustic features are extracted (e.g., aural disparities, which are analyzed in the superior olive, Yin 2002). All the parallel processing streams of the auditory system converge at the level of the inferior colliculus (IC), an obligatory nucleus of the auditory system with no analogs in other sensory system (Casseday et al. 2002). The neurons of the IC project to

*Correspondence to*: I. Nelken
(e-mail: Israel@md.huji.ac.il)

the auditory thalamus, which serves in turn as the input to auditory cortex. In terms of synaptic distance, the IC is as far from the hair cells as the visual thalamus, or even the visual cortex, is far from the photoreceptors. There is, therefore, at least one additional synapse, and possibly even two or three synapses, in the auditory system between the periphery and the cortex. The high level of integration in the IC, combined with the exquisite sensitivity of its neurons to various physical features of sounds, makes the IC a compelling analog to visual cortex. On the other hand, the auditory information arriving at the auditory cortex is much more highly processed than the visual information arriving at the visual cortex.

Despite these differences, research on auditory cortex was influenced very strongly by the successful elucidation of the organization of the visual cortex. In particular, it is often assumed that the main contribution of primary auditory cortex (A1) to auditory perception is the computation of some nontrivial parameter from the incoming sounds and its representation in terms of topographical location on the cortical surface (Middlebrooks et al. 1980; Mendelson et al. 1993; Shamma et al. 1993; Versnel et al. 1995). Since A1 has a prominent gradient of frequency sensitivity, the main interest in such studies is in mapping neuronal feature sensitivity along the orthogonal, iso-frequency direction. A number of response properties, in addition to frequency selectivity, have been shown to be roughly topographically mapped across A1 (Read et al. 2002). However, none of these properties really fits the bill of an interesting, computationally derived parameter with the same standing as, e.g., orientation selectivity in visual cortex. The search for such a parameter therefore failed to a large extent. We would like to argue that this failure is not accidental and that the issue of feature extraction in A1 is actually secondary to understanding its role in sound processing.

As specific examples for these statements, we will discuss here the coding of frequency-modulated tones (FM tones), where directional selectivity has been hypothesized to be an important computationally derived property of cortical neurons (Mendelson and Cynader 1985) and the use of spectrotemporal receptive fields (STRFs) as a nonparametric approach to the discovery of features analyzed by cortical neurons (Aertsen and Johannesma 1981a).

1.1 The case of the frequency-modulated tones

Studies of FM tones in A1 usually use unidirectional frequency trajectories spanning a very wide frequency band. Figure 1 shows the main characteristics of this type of FM tones and the responses they elicit in A1 neurons. Neurons respond to FM tones with a short burst of action potentials when the instantaneous frequency approaches the best frequency of the neuron and the "trigger frequency" at which the neuron fires occurs before the instantaneous frequency reaches the
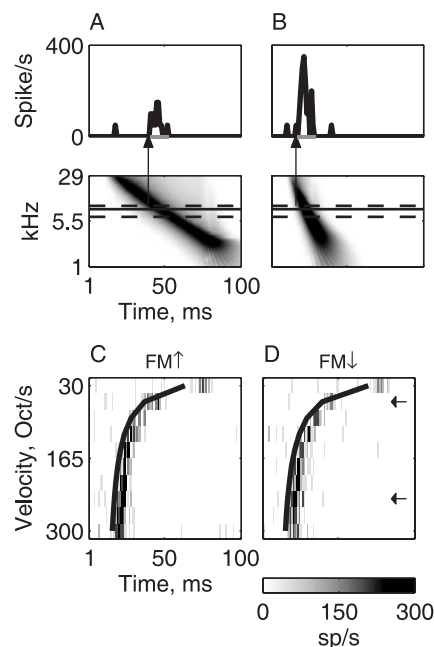
Fig. 1a–d. Responses of AI multiunit cluster to FM tones. Data from Nelken and Versnel (2000). **a** The response of the cluster to a downward FM tone with a velocity of 60 Oct/s. **b** The response of the cluster to a downward FM tone with a velocity of 240 Oct/s. In **a** and **b**, the stimulus is represented below the response. This spectrotemporal representation is the neural activity pattern predicted from a model of the auditory periphery (Auditory Image Model, Bleeck and Patterson, http://www.mrc-cbu.cam.ac.uk/cnbh/aimmanual), in which the basilar membrane motion was modeled by a gamma-tone filter bank, followed by half-wave rectification, compression, and lowpass filtering. The *continuous line* drawn on the neural activity pattern represents the best frequency of the neuron (8.5 kHz), and the *dashed lines* are the trigger frequencies for upward and downward FM tones. The *arrows* indicate the moment in time at which the instantaneous frequency crossed the trigger frequency, corrected by a constant delay. Clearly, the spike burst was evoked at that moment for both velocities. The *gray line* indicates the duration of the burst, 13 ms. **c** Responses to multiple velocities, upward FM tones. **d** Responses to multiple velocities, downward FM tones. In **c** and **d**, the *line* represents the predictions of fixed trigger frequency model fitted to the first spike latencies at each velocity. The actual trigger frequencies are drawn on the spectrotemporal display in **a** and **b** in *dashed lines*. The *arrows* in **d** mark the responses drawn as *line graphs* in **a** and **b**

best frequency (Heil et al. 1992b, c; Nelken and Versnel 2000). The timing of the burst is very stereotypical and can be precise up to a few milliseconds (Fig. 1). The number of spikes elicited may be different for upward and for downward sweeps, resulting in a preference of one direction of frequency change relative to the other one (although the neuron in Fig. 1 responded approximately equally well to both FM directions). This parameter has often been compared with the directional preference of neurons in visual cortices (Mendelson and Cynader 1985; Tian and Rauschecker 1994, 1998).

Mapping studies of directional selectivity in A1 gave variable results. Some studies (Mendelson et al. 1993; Shamma et al. 1993) reported a consistent map of directional selectivity orthogonal to the frequency gradient. Other studies (Heil et al. 1992a), using objective criteria for clustering, reported no special clustering of

directional sensitivity along an isofrequency contour. Nelken and Versnel (2000) reported mixed results, with some maps showing significant clustering of directional selectivity and others not.

More recently, a number of studies have suggested that the whole issue of maps of directional sensitivity along the isofrequency gradient is not well defined. In order for the directional sensitivity to be a reasonable computationally derived property of a neuron, it must be invariant to "nuisance parameters", such as the exact details of the frequency trajectory or the species that is studied. For example, one possible difference between the studies that reported clustering of directional selectivity along isofrequency contours (Mendelson et al. 1993; Shamma et al. 1993) and the one that did not (Heil et al. 1992b) was the fact that in the first group, exponential frequency trajectories (covering equal frequency ratios in equal times, with velocity measured in octaves/second) were used, whereas in the study of Heil et al., linear frequency trajectories (covering equal frequency differences in equal times, with velocity measured in kHz/second) were used. Nelken and Versnel (2000) tried to resolve this issue by comparing, in the same animals, a large number of different frequency trajectories. They have shown that varying such nuisance parameters may dramatically change the directional preference of neuronal clusters. To further murk the issue, Tian and Rauschecker (1994) found that in cats, directional selectivity is much more pronounced for slow than for fast FM tones, whereas Zhang et al. (2003) showed the reverse in rats: directional preference is much stronger for fast than for slow chirps in the rat. Even more worrying, Heil et al. (1992c) showed that in the chick directional preference is mapped, not along an isofrequency contour, but along the frequency gradient, with a strong correlation between best frequency and directional selectivity. Zhang et al. (2003) have demonstrated a similar organization in the auditory cortex of a mammal, the rat.

In addition to the inconclusive results reviewed above, the functional significance of such maps for the processing of other sounds, in particular the much more acoustically complex natural sounds, is undecided. The stimuli used in the majority of the studies reviewed above are FM tones with a large frequency extent. These stimuli are very different from the much shorter chirp components in animal vocalizations. For example, Bar-Yosef et al. (2002) have shown that in a large sample of natural vocalizations dominated by FM tones, the typical velocity is less than 80 kHz/s and the typical frequency extent is less than 3 kHz. In cats, however, most neurons prefer FM tones with velocities exceeding 1024 kHz/s (Heil et al. 1992b), more than an order of magnitude faster than the typical natural FM component. In the ferret, the situation is less unbalanced (Nelken and Versnel 2000), but there are still a large number of clusters preferring very high velocities.

The complexity of the picture regarding the coding of a rather simple feature, directional selectivity to a frequency-modulated chirp, suggests that this is not "the" computationally derived parameter that is the analog of orientation selectivity or even directional selectivity in primary visual cortex. In fact, the results presented above suggest that directional selectivity is not really a feature that is processed by the cortex explicitly, but rather a by-product of a computation that is doing something else.

Indeed, fairly good models are available for describing these responses. For example, Fishbach et al. (2001) presented a model whose basic operation is taking the derivative of the envelope of the sound, where the derivative is computed using time constants of 1–10 ms. They showed that their model can account for essentially all the data in the literature regarding electrophysiological and psychophysical effects of manipulations of the onset of sounds. The model suggested by Fishbach et al. (2001) is a single-channel model and therefore could not account for the responses to FM tones. An elaboration of the same model (Fishbach et al. 2003), operating on multiple frequency channels, is capable of accounting for the responses to FM tones and two-tone complexes. Surprisingly, the parameters of the spectrotemporal model of Fishbach et al. (2003), estimated from responses collected in mapping studies of ferret A1 (Nelken and Versnel 2000), are in fact more topographically clustered on the cortical surface than the feature selectivity indices derived directly from the experimental data, such as the directional selectivity to FM tones. Thus, directional selectivity might be a side product of a basic differentiation operation carried in each frequency band separately with relatively short time constants and then summed up across frequency.

1.2 The spectrotemporal receptive field

An almost diametrically opposite approach for studying feature detection in auditory cortex is to assume nothing about the nature of the preferred feature. Instead, the experiment consists of recording responses to a rich set of sounds and then asking what the parameters in the sound were that elicited responses. This approach has been codified in the reverse-correlation (or revcor) methodology (Eggermont et al. 1983). In this approach, the set of sounds that precedes spikes is extracted from the full set of sounds used for characterizing the neuron. The revcor function is the average of this set (averaging the sound waveform or, more commonly, a time-frequency representation). For general sound ensembles, the revcor function is contaminated by the correlation structure of the set of sounds used for testing the neuron. Deconvolving the revcor function by the correlation function of the set of sounds results in the kernel function, often named the spectrotemporal receptive field (STRF), of the neuron. Efficient and statistically sound methods have been developed recently for this last step, making it practicable to use very general sets of sounds (Theunissen et al. 2001; Linden et al. 2003).

The revcor approach has had its ups and downs in the history of auditory research. Initially proposed by deBoer for auditory nerve studies, it was used extensively for

studying other parts of the auditory system in the early 1980s (Aertsen and Johannesma 1980; Johannesma and Aertsen 1982; Eggermont et al. 1983). However, except for scattered studies in the auditory nerve and the cochlear nucleus (Carney and Yin 1988; Clopton and Backoff 1991; Kim and Young 1994; Nelken et al. 1997), the method was to a large extent abandoned for over 10 years, resurfacing again in the late 1990s as a major tool for characterizing neurons in higher auditory centers (deCharms et al. 1998; Klein et al. 2000; Schnupp et al. 2001; Theunissen et al. 2001; Miller et al. 2002; Linden et al. 2003). The idea of characterizing a neuron by finding regularities in the sets of sounds that precede spikes is a very strong one. The pioneers of this method defined these preevent sound ensembles as the basic statistical structure that emerges from this type of analysis (Johannesma and Aertsen 1982). From their point of view, the revcor function was only one way of characterizing these sound ensembles. Recent developments support this view, showing that other ways of characterizing the preevent sound ensembles may be more powerful than simply computing the revcor function or the STRF (Brenner et al. 2000; Paninski 2003).

Initial reports in the most recent wave of publications based on STRFs emphasized neurons with complex STRF shapes (deCharms et al. 1998), but later reports of a number of extensive databases of STRFs demonstrate that their structure is most commonly rather simple (Depireux et al. 2001; Miller et al. 2002; Linden et al. 2003). Most STRFs are separable in time and frequency in the sense that they are well approximated by a product of a function of frequency alone and a function of time alone. When they are not separable, they are still quadrant separable, in the sense that their 2-d Fourier transforms are separable in each quadrant. Thus, complex STRF shapes are rare.

The typical STRF in mammalian cortex has therefore a simple structure – it is usually composed of an excitatory patch around the best frequency of the neuron, with possibly inhibitory subfields surrounding the excitatory patch in all directions. For example, the STRF in Fig. 2a is composed of an excitatory patch at 8.5 kHz, the best frequency of the neuron as determined by tones, followed by a long suppressory region, but with very little if any inhibitory sidebands. This STRF is clearly separable in time and frequency. Furthermore, the time scales related to the STRF are relatively slow. In the cat, the typical best modulation frequency (defined by the peak of the Fourier transform of the temporal component) is about 16 Hz (Miller et al. 2002). In the mouse, the peak latency of the STRF is typically 30 ms, and the typical duration of the STRF (including both the excitatory peak and the typical late suppressory phase) is over 100 ms (Linden et al. 2003), suggesting similar slow time constants. In Fig. 2a, the Fourier transform of the temporal component of the STRF has a lowpass-filter shape, and the corner frequency (-10 dB) is 48 Hz – a rather high value for cortical neurons.

Even such relatively fast STRFs cannot explain many aspects of the responses of A1 neurons to other sounds.
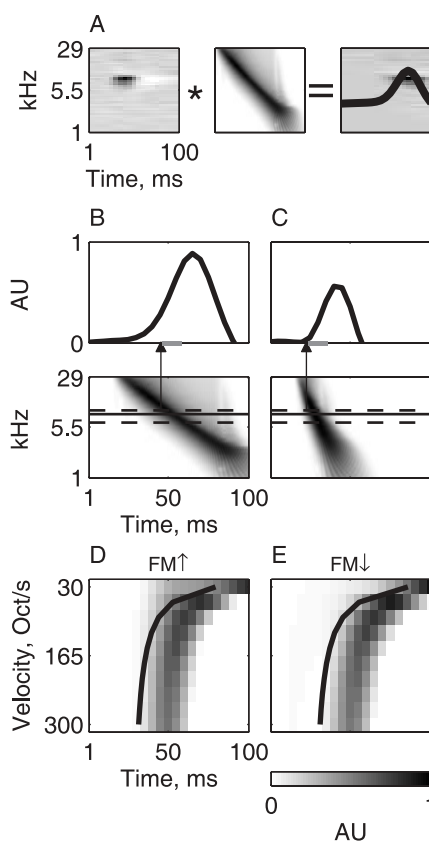
Fig. 2. a Prediction of the responses to FM tones using STRF. The *left panel* is an STRF measured for a neuron recorded in ferret A1 (data courtesy of J. Schnupp, University Laboratory of Physiology, Oxford University). The *middle panel* is a neural activity pattern, computed as in Fig. 1. Convolving the two at each frequency channel results in the data shown in the *right panel*. The predicted response is the sum of the activities in each channel as a function of time and is superimposed as a *thick black line*. **b** Predicted response for a downward FM tone with a velocity of 60 Oct/s. **c** Predicted response for a downward FM tone with a velocity of 240 Oct/s. The display follows the same conventions as in Fig. 1a and b. The responses were shifted so that the time at which the instantaneous frequency crossed the trigger frequency in **c** corresponds to the onset of the predicted response to that FM tone. **d, e** Predictions of the responses to upward and downward FM tones at multiple velocities. The *lines* are the same *lines* as in Fig. 1c and D, except for a constant shift to fit the latency of the fastest response

In particular, it cannot explain the sensitivity of A1 neurons to fast FM tones (Fig. 2; see also Fishbach et al. 2003). Whereas the neuron in Fig. 1 preferred fast FM tones (the response in Fig. 1a, to a velocity of 60 Oct/s, is smaller than the response in Fig. 1b, to a velocity of 240 Oct/s), the responses in Figs. 2b and 2c show the reverse preference. The preference of STRF predictions to slow FM tones is a direct consequence of their rather slow time course. In addition, the duration of the predicted responses is much longer than the observed responses (the gray lines below the responses in Figs. 2b and 2c mark the duration of the response bursts in Fig. 1, 13 ms). Even for the fastest FM tones, the predicted responses are longer and have a much slower rise time than the typical measured responses, illustrating the mismatch between the fast, measured and the sluggish

predicted responses. Finally, the timing of the responses is not predicted well (compare the fit of the lines of fixed trigger frequency in Fig. 1c and d to the responses with the fit of the same lines in Fig. 2d and e). Although all of these problems could in principle be fixed by a post-processing stage that would, e.g., compress and differentiate the STRF predictions, to some extent the whole point of the STRF calculation is lost. For example, differentiating the STRF predictions is identical to predicting the responses using the differentiated STRF. But the temporal properties of the STRF, and of its differentiated version, are very different from each other, raising again the question of why the "correct" kernel did not emerge from the estimation procedure to start with.

Furthermore, STRFs probably cannot capture well the responses of cortical neurons to natural sounds, as has been argued in Bar-Yosef et al. (2002), although without quantitative analysis. The reason for this is the high sensitivity of cortical neurons to small perturbations of natural sounds such as removal of background noise. Such sensitivity cannot be accounted for by STRFs of the published types since for such linear filters small perturbations of this kind necessarily result in small changes in their responses.

1.3 Complex features, simple features, or no features?

Clearly, all the characterizations described above in terms of feature detection are seriously lacking in generalizability. For example, as Nelken and Versnel (2000) have shown in the context of artificial FM tones, and as Bar-Yosef et al. (2002) have shown in the context of natural bird chirps, the nuisance parameters may influence the responses of neurons more strongly than the parameter being studied. In other words, neurons in A1 are extremely sensitive to small changes in their stimuli, changes that the experimenters may consider as irrelevant. It seems that the main problem with A1 neurons is that they are "promiscuous": they respond to too many different sounds that are too different from each other to usefully extract a single feature that is responsible for all these responses. Thus, the features that neurons seem to encode change with the set of sounds used. This conclusion had already been reached many years ago, during the first "wave" of STRF studies (Aertsen and Johannesma 1981b; Eggermont et al. 1983) and has been reproduced a number of times since, even in subcortical stations (Nelken et al. 1997; Theunissen et al. 2000).

In more concrete terms, responses to transient tonal sounds such as FM tones suggest that neurons have fast time constants. On the other hand, responses to the continuous sounds such as those used to estimate STRFs suggest that neurons have slow time constants. Remarkably, in both cases the operation that is performed is a derivative in the spectrotemporal plane. The difference is not in the operation but in the time constants over which it is computed. Computing a derivative is not a complex operation per se – derivatives at various time constants are computed all over the nervous system, from the periphery of all sensory systems on up. The complexity of the A1 neurons is to some extent due to the fact that the same neuron can presumably act both fast and sluggish.

The fact that both descriptions of cortical neurons, one based on responses to FM tones and one based on the STRF, end up with rather simple operations raises the question of where these computations take place. The subcortical auditory pathway is extremely rich, much richer than that of any other sensory system. Already in the brainstem, rather complex computations take place, both in spectrum (e.g., Nelken and Young 1996) and in time (e.g., Joris and Smith 1998). Sensitivity to transients is further enhanced in the IC (Frisina 2001; Sinex et al. 2002). In fact, the excitatory component of the subthreshold responses to FM tones in A1 is already directional selective (Zhang et al. 2003), suggesting that direction selectivity to FM tones is already computed subcortically. Thus, much of the feature extraction discussed above could already be accomplished subcortically.

This description suggests that the issue of time constants is crucial for understanding processing in A1. This is because no single physical feature of sounds was shown to be specifically extracted in the cortex – subcortical centers can do all of this kind of work. The special properties of cortical neurons have to do with the multiple interacting time constants expressed in their responses.

2 Time constants: neurons in primary auditory cortex are both sluggish and fast

We therefore claim that the major new feature of the neuronal activity in A1 is not so much the computation of complex features, but rather the introduction of multiple time constants into auditory processing. There is nothing new in the claim that long time constants appear in A1: auditory cortex is known to be more sluggish than subcortical stations on a number of parameters. Thus, the ability of neurons in A1 to follow repetitive sounds is reduced relative to subcortical stations (e.g., Eggermont 1991). In the context of natural sounds, cortical neurons follow the energy pattern of vocalizations with less fidelity than MGB neurons (Creutzfeldt et al. 1980). Such findings abound in the literature. These findings suggest that, whereas effective time constants in the IC are around 10 ms or less, in A1 the effective processing time constants are on the order of 100 ms.

These longer time constants are annoying – they cause degradation in the ability of cortical neurons to code effectively the physical structure of sounds. We would like, however, to argue that there are other time constants in play, both much faster than the 100-ms time constant (on the order of 10 ms or less) and much longer than the 100-ms time constant (on the order of 1 s or more). All of these time constants coexist in cortical neurons and are responsible to a large extent for their complex behavior.

2.1 Coexistence of fast and slow time constants

In the naïve view of the slow time constants in A1, successive stages of lowpass filtering cause membrane potential fluctuations to slow down considerably in A1 relative to lower auditory stations. The sluggishness of the spikes is a consequence of the slowing down of the membrane potential fluctuations. This view is implicit in the use of the STRF as a characterization of A1 neurons since the lowpass filtering operation is inherent in the sluggish modulation transfer functions derived from the STRF in A1 (Fig. 2). However, this view is inconsistent with a large corpus of data showing an exquisite sensitivity of A1 neurons to fast transients.

In fact, the short latency (10–14 ms) and the very low jitter of the first spike latency (often less than 1 ms) of cortical neurons (Phillips and Hall 1990) strongly argue for a fast rise time of the membrane potential at the onset of sounds. Such fast rise time of the membrane potential is to a large extent inconsistent with a model in which the membrane potential is the result of a severe lowpass filtering operation on the envelope of the stimulus. In fact, Heil and Neubauer (2003) showed that behavioral thresholds, first-spike latencies in the auditory nerve, and first-spike latencies in auditory cortex of cats follow the same rules. They interpret these rules as the result of a perfect integrator of peak pressure (rather than acoustic intensity, the square of the pressure) followed by a spike when the integrator output reaches a constant level and suggest that this integrator resides in the first synapse of the auditory system, between the hair cells and the auditory nerve fibers. These findings suggest the presence in the cortex of a faithful copy of the activity evoked by sound onsets in the auditory nerve. The presence of such a copy is essentially impossible if the membrane potential of the cortical neurons is highly smoothed by lowpass filtering.

Other recent data strengthen the case for fast membrane potential rise time in A1. For example, the data that led Heil and his collaborators to their integrator model (Heil 1997a, b; as modeled by Heil and Neubauer 2003) show that some parameters of the first spike, such as its latency and its probability, are strongly modulated by the shape of the onset ramp of a sound. Response properties of cortical neurons show large changes when the duration of the onset ramp changes between 1 ms and 10 ms. However, a lowpass filter with a corner frequency of 10 Hz cannot distinguish between ramps of these durations – both are too fast (Fig. 2). In fact, the models of Fishbach et al. (2001, 2003), which account for a large number of physiological and psychoacoustical phenomena related to sound onsets and FM chirps, required time constants of 1–10 ms to account for those, even for data from A1.

The most direct evidence for fast dynamics of subthreshold events in A1 comes from intracellular recordings. A number of recent studies (Ojima and Murakami 2002; Zhang et al. 2003) published such data for simple stimuli (pure tones and FM chirps). The rise time of the membrane potential in the study of Ojima and Murakami (2002), for example, is at least as fast as the rise time of their stimuli (10 ms).

The fast onset dynamics could, however, be followed by sluggish sustained dynamics. For many years most reports of neuronal responses in A1 concentrated on the onset response, which is usually the only response component under deep barbiturate anesthesia. However, late response components are now widely reported, even under anesthesia. Neurons in A1 respond continuously to random chord stimuli (deCharms et al. 1998; Schnupp et al. 2001) to and ripple stimuli (Klein et al. 2000; Depireux et al. 2001). Late response components are widely present in response to natural bird chirps (Bar-Yosef et al. 2002) and even in response to pure tones (Ulanovsky et al. 2003). These postonset responses, which dominate, for example, the estimation of the STRF, could in principle be much more sluggish than the onset responses.

That this is often not the case is also widely known. The late sluggishness of cortical neurons is often measured by the loss of synchrony to repetitive stimuli. When neurons lock to the envelope of repetitive stimuli, a rise in repetition rate would often result in a decrease in the probability of spikes evoked at later periods, but not in a much greater variability in their generation time when they do occur (Phillips 1989). Elhilali et al. (2003) presented direct evidence for the high level of temporal precision of cortical spikes, both in anesthetized and in awake animals, during sustained stimulation (by temporally orthogonal ripple combinations, TORCs). In their data, responses to the same stimulus may contain spikes that are precise up to 1 ms or so, within the rather sluggish responses locked to the slow envelope modulations of the TORCs. In the same vein, intracellular recordings of neurons in A1 in response to repetitive wideband bursts show that the slope of the membrane potential trajectory often remains as fast for the later bursts as it is for the first burst (Las et al. unpublished results). Cortical sluggishness has often been attributed to anesthesia effects. Recent reports, however, suggest that neurons in A1 of awake animals are not much less sluggish than those under antesthesia (Lu et al. 2001; Elhilali et al. 2003). Thus, fast and slow time constants coexist beyond the onset response.

2.2 Very long time constants in auditory cortex

The presence of time constants in auditory cortex that are much longer than the 10–100 ms range discussed above was also known for a long time. Neurons in auditory cortex are often termed "labile" because they tend to adapt very rapidly to consecutive presentations of the same stimulus. This lability requires long time constants in order for previous stimulus presentations to influence the response of a neuron to the next presentation. These time constants must be on the order of at least 1 s to account for adaptation occurring for interstimulus intervals of that duration.

Lability was considered as a serious liability, but recent data suggest that it has useful properties. Malone

et al. (2002), extending previous results in subcortical stations (Spitzer and Semple 1993, 1998) have shown that stimulus-specific adaptation, which exists under some conditions already in the IC, acts very strongly in A1.

Ulanovsky et al. (2003) extended the scope of such adaptation studies by using an oddball paradigm to test the responses of A1 neurons to rare sounds. They used pairs of frequencies, where one frequency appeared often ("standard") and the other one was rare ("deviant"). Responses to the same frequency generally were dependent on whether that frequency was playing the role of the standard or of the deviant. For most neurons, the responses to the same frequency, when deviant, were stronger than when it was standard. This was true even when the frequency difference between standard and deviant was 4%, at least an order of magnitude below the width of the tuning curves of these neurons at the sound level used. Ulanovsky et al. further showed that the relevant time constant is about 1–2 s; when the interstimulus interval was longer, the adaptation was no longer apparent. Most significantly for the argument of this section, this phenomenon was not observed in recordings of thalamic neurons under the same stimulation conditions.

2.3 Summary: an interplay between at least three time constants in auditory cortex

We documented at least three time constants that are relevant for sound processing in A1. The shortest, about 10 ms or less, is manifested in the well-locked onset responses and in the very low jitter of response components to specific ongoing acoustic events. The middle time constant, of about 100 ms, is manifested in the decrease of firing probability to late periods of repetitive sounds at rates higher than 20–30 Hz and in the general sluggish timing of STRFs in the mammalian cortex. Finally, very slow time constants, of 1 s and longer, are manifested in the lability of neurons in A1 and may be highly stimulus dependent. These are not the only time constants operating in A1; for example, forward masking (Calford and Semple 1995; Brosch and Schreiner 1997) seems to operate according to yet another time constant, probably between 100 ms and 1 s.

As discussed above, all three time constants may be expressed in the responses to the same stimulus. For example, Nelken and Versnel (2000) recorded responses to FM tones with a "trapezoid" contour in which the tone first sweeps up from low to high frequency, stays at the high frequency for about 300–400 ms, and then sweeps back to the low frequency. The lowest and highest frequencies of these contours were well outside the tuning curves of the neurons. They also used reverse trapezoids in which the frequency started at the high frequency, swept down to the low frequency, stayed for a while at the low frequency, and then swept up again. The responses were well timed for both upward and downward sweeps and for both the trapezoid and reverse trapezoid trajectories. However, the responses to the upward sweep when it appeared first (in the trapezoid trajectory) were stronger than the responses to it when it appeared second (in the reverse trapezoid trajectory). This finding was attributed to adaptation due either to the close temporal proximity of the upward and downward sweeps or to the presence of a continuous tone (although it was outside the tuning curve of the neuron). Thus, the neurons manifested both fast and slow time constants.

We believe that the complexity of cortical processing is related in an essential way to the multiplicity of time constants in auditory neurons. Neurons in A1 are both fast and sluggish. Under some conditions, an extremely well-timed spike may occur; under other, roughly similar conditions, this spike will not occur. For example, we have shown (Nelken et al. 1999, 2001) that adding a low-level tone to a strong fluctuating noise masker causes a suppression of the envelope locking. This may occur under conditions in which the tone by itself does not give rise to any response. The signal-to-noise ratios at which this suppression may occur are −30, −40, or even −50 dB. The tone presents, under these circumstances, an extremely small perturbation, but nevertheless it determines whether or not the neuron will fire spikes. Similar findings have been shown in Bar-Yosef et al. (2002), where the removal of the background noise from recordings of bird songs could shift or delete well-timed spikes. We believe, therefore, that the fast time constants are somehow under the control of the slower time constants. This is most clearly seen in the case of the longest time constants, which express themselves in a multiplicative gain factor that is extremely sensitive to the parameters of the stimuli (Ulanovsky et al. 2003).

3 Long time constants are crucial for solving the "hard problems" of auditory perception

What are the longer time constants good for? We would like to argue that some of the features of the "hard problems" of auditory perception require the presence of long time constants. The nontrivial character of the longer time constants in A1 could supply the required substrate for solving these problems.

There are many examples of hard problems in auditory perception. For example, speech perception is hard – speech is highly redundant and is perceptually resistant to noise, suggesting that it should be easy to recognize. Furthermore, the activity of neurons in the auditory nerve and the cochlear nucleus can be shown to contain enough information for supporting speech perception (Sachs 1984; Shamma 1985; Palmer et al. 1986). However, physiological accounts of speech perception are very weak – for example, there are no good candidates for physiological correlates of categorical perception, with one exception (Steinschneider et al. 1994; Eggermont 1995).

Two other, less obvious, examples are pitch and space. Pitch is based on the extraction of regularities in spectrum and in time that are strongly manifested in the firing patterns of neurons in the auditory nerve and

cochlear nucleus (Cariani and Delgutte 1996a, b; Winter et al. 2001). While it is clear that pitch processing is performed centrally, it is unclear how pitch is represented above the cochlear nucleus. Finally, space perception is based on the processing of binaural cues, which are first extracted in the brainstem. However, space perception is more than just binaural cues, involving integration of cues across time and frequency (Clifton 1987; Trahiotis and Stern 1989, 1994; Hafter and Buell 1990; Freyman et al. 1991). The integration of these cues in complex spatial scenes is very poorly understood.

These three problems share some common features. They all require integration across a wide frequency band, and they all generalize highly across the physical structure of sounds in the sense that many different sounds, with very different peripheral representations, give rise to the same percept. More importantly for the argument presented here, all three depend on context. For example, it is possible to change the identity of a vowel or the pitch of a sound by capturing some of its components into a separate auditory stream (Darwin et al. 1989, 1995), and it is possible to manipulate the detection threshold of a sound by adapting the precedence effect (Freyman et al. 1991).

We believe that the properties of neurons in A1 are insufficient for fully solving speech, pitch, and space perception, at least in the sense that there are no neurons in A1 that will respond to all sounds of the same pitch independent of their spectral content or that there are no neurons in A1 that will respond to all /a/ sounds, independent of their physical structure. However, the longer time constants expressed in A1 can serve for building the auditory objects to which later processes can assign speech sound identity, pitch value, or spatial location. Building auditory objects requires long time constants because of the contextual effects described above. The coexistence of short and long time constants in A1 may enable cortical neurons to both extract the auditory objects and code for their features at the same time.

Based on these considerations, we suggest the following model for auditory processing. Feature extraction is actually done below the level of the cortex, and we hypothesize that the most detailed physical representation of sounds in terms of their spectrotemporal structure is actually complete by the level of the IC. Thalamic and cortical processing operate on this representation to generate auditory objects. As suggested by de Cheveigne (2001), we hypothesize that the most important operation performed by cortical neurons is that of splitting sound in one frequency channel into multiple objects when necessary. It is this operation that is subserved by the multiple time constants in A1 processing. These auditory objects are then operated upon by higher auditory centers, producing the perception of phonemic quality, pitch value, spatial location, and all other auditory qualities.

In this model, A1 has a pivotal role, not in terms of sophisticated feature extraction, but rather in terms of the integration processes that occur in it. We believe that studying the interplay between the different time constants will lead us to a better understanding of the operations performed by A1 and therefore to a more precise formulation of its role in the auditory pathway.

References

Aertsen AMHJ, Johannesma PIM (1980) Spectro-temporal receptive fields of auditory neurons in the grassfrog. I. Characterization of tonal and natural stimuli. Biol Cybern 38: 223–234

Aertsen AM, Johannesma PI (1981a) The spectro-temporal receptive field. A functional characteristic of auditory neurons. Biol Cybern 42: 133–143

Aertsen AM, Johannesma PI (1981b) A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. Biol Cybern 42: 145–156

Bartfeld E, Grinvald A (1992) Relationships between orientation-preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex. Proc Natl Acad Sci USA 89: 11905–11909

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

Brenner N, Bialek W, de Ruyter van Steveninck R (2000) Adaptive rescaling maximizes information transmission. Neuron 26: 695–702

Brosch M, Schreiner CE (1997) Time course of forward masking tuning curves in cat primary auditory cortex. J Neurophysiol 77: 923–943

Calford MB, Semple MN (1995) Monaural inhibition in cat auditory cortex. J Neurophysiol 73: 1876–1891

Cariani PA, Delgutte B (1996a) Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. J Neurophysiol 76: 1717–1734

Cariani PA, Delgutte B (1996b) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J Neurophysiol 76: 1698–1716

Carney LH, Yin TCT (1988) Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model. J Neurophysiol 60: 1653–1677

Casseday JH, Fremouw T, Covey E (2002) The inferior Colliculus: A hub for the central auditory system. In: Oertel D, Fay RR, Popper AN (eds) Integrative functions in the mamalian auditory pathway. Springer, Berlin Heidelberg New York, pp 238–318

Clifton RK (1987) Breakdown of echo suppression in the precedence effect. J Acoust Soc Am 82: 1834–1835

Clopton BM, Backoff PM (1991) Spectrotemporal receptive fields of neurons in cochlear nucleus of guinea pig. Hear Res 52: 329–344

Creutzfeldt O, Hellweg FC, Schreiner C (1980) Thalamocortical transformation of responses to complex auditory stimuli. Exp Brain Res 39: 87–104

Darwin CJ, Pattison H, Gardner RB (1989) Vowel quality changes produced by surrounding tone sequences. Percept Psychophys 45: 333–342

Darwin CJ, Hukin RW, al-Khatib BY (1995) Grouping in pitch perception: evidence for sequential constraints. J Acoust Soc Am 98: 880–885

de Cheveigne A (2001) The auditory system as a "separation machine". In: Breebart DJ, Houtsma AJM, Kohlrausch A, Prijs VF, Schoonhoven R (eds) Physiological and psychophysical bases of auditory function. Shaker, Maastricht pp 453–460

deCharms RC, Blake DT, Merzenich MM (1998) Optimizing sound features for cortical neurons. Science 280: 1439–1443

Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J Neurophysiol 85: 1220–1234

Eggermont JJ (1991) Rate and synchronization measures of periodicity coding in cat primary auditory cortex. Hear Res 56: 153–167

Eggermont JJ (1995) Representation of a voice onset time continuum in primary auditory cortex of the cat. J Acoust Soc Am 98: 911–920

Eggermont JJ, Johannesma PM, Aertsen AM (1983) Reverse-correlation methods in auditory research. Q Rev Biophys 16: 341–414

Elhilali M, Klein D, Fritz J, Simon JZ, Shamma S (2003) The engima of cortical responses: slow yet precise. In: Pressnitzer D, de Cheveigne A, McAdams S, Collet L (eds) Auditory signal processing: physiology, psychoacoustics, and models. Springer, Berlin Heidelberg New York (in press)

Fishbach A, Nelken I, Yeshurun Y (2001) Auditory edge detection: a neural model for physiological and psychoacoustical responses to amplitude transients. J Neurophysiol 85: 2303–2323

Fishbach A, Yeshurun Y, Nelken I (2003) A neural model for physiological responses to frequency and amplitude transitions uncovers topographical order in the auditory cortex. J Neurophys (in press)

Freyman RL, Clifton RK, Litovsky RY (1991) Dynamic processes in the precedence effect. J Acoust Soc Am 90: 874–884

Frisina RD (2001) Subcortical neural coding mechanisms for auditory temporal processing. Hear Res 158: 1–27

Hafter ER, Buell TN (1990) Restarting the adapted binaural system. J Acoust Soc Am 88: 806–812

Heil P (1997a) Auditory cortical onset responses revisited. II. Response strength. J Neurophysiol 77: 2642–2660

Heil P (1997b) Auditory cortical onset responses revisited. I. First-spike timing. J Neurophysiol 77: 2616–2641

Heil P, Neubauer H (2003) A unifying basis of auditory thresholds based on temporal summation. Proc Natl Acad Sci USA 100: 6151–6156

Heil P, Rajan R, Irvine DR (1992a) Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. II. Organization of response properties along the 'isofrequency' dimension. Hear Res 63: 135–156

Heil P, Rajan R, Irvine DR (1992b) Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. I. Effects of variation of stimulus parameters. Hear Res 63: 108–134

Heil P, Langner G, Scheich H (1992c) Processing of frequency-modulated stimuli in the chick auditory cortex analogue: evidence for topographic representations and possible mechanisms of rate and directional sensitivity. J Comp Physiol A 171: 583–600

Johannesma P, Aertsen A (1982) Statistical and dimensional analysis of the neural representation of the acoustic biotope of the frog. J Med Syst 6: 399–421

Joris PX, Smith PH (1998) Temporal and binaural properties in dorsal cochlear nucleus and its output tract. J Neurosci 18: 10157–10170

Kim PJ, Young ED (1994) Comparative analysis of spectro-temporal receptive fields, reverse correlation functions, and frequency tuning curves of auditory-nerve fibers. J Acoust Soc Am 95: 410–422

Klein DJ, Depireux DA, Simon JZ, Shamma SA (2000) Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. J Comput Neurosci 9: 85–111

Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM (2003) Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. J Neurophysiol (in press)

Lu T, Liang L, Wang X (2001) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat Neurosci 4: 1131–1138

Malone BJ, Scott BH, Semple MN (2002) Context-dependent adaptive coding of interaural phase disparity in the auditory cortex of awake macaques. J Neurosci 22: 4625–4638

Mendelson JR, Cynader MS (1985) Sensitivity of cat primary auditory cortex (AI) neurons to the direction and rate of frequency modulation. Brain Res 327: 331–335

Mendelson JR, Schreiner CE, Sutter ML, Grasse KL (1993) Functional topography of cat primary auditory cortex: responses to frequency-modulated sweeps. Exp Brain Res 94: 65–87

Middlebrooks JC, Dykes RW, Merzenich MM (1980) Binaural response-specific bands in primary auditory cortex (AI) of the cat: topographical organization orthogonal to isofrequency contours. Brain Res 181: 31–48

Miller LM, Escabi MA, Read HL, Schreiner CE (2002) Spectro-temporal receptive fields in the lemniscal auditory thalamus and cortex. J Neurophysiol 87: 516–527

Nelken I, Young ED (1996) Why do cats need a dorsal cochlear nucleus? J Basic Clin Physiol Pharmacol 7: 199–220

Nelken I, Versnel H (2000) Responses to linear and logarithmic frequency-modulated sweeps in ferret primary auditory cortex. Eur J Neurosci 12: 549–562

Nelken I, Kim PJ, Young ED (1997) Linear and nonlinear spectral integration in type IV neurons of the dorsal cochlear nucleus. II. Predicting responses with the use of nonlinear models. J Neurophys 78: 800–811

Nelken I, Rotman Y, Bar Yosef O (1999) Responses of auditory-cortex neurons to structural features of natural sounds. Nature 397: 154–157

Nelken I, Jacobson G, Ahdut L, Ulanovsky N (2001) Neural correlates of comodulation masking release in auditory cortex of cats. In: Houtsma A, Kohlrausch A, Prijs V, Schoonhoven R (eds) Physiological and psychophysical bases of auditory function. Shaker, Maastricht

Ojima H, Murakami K (2002) Intracellular characterization of suppressive responses in supragranular pyramidal neurons of cat primary auditory cortex in vivo. Cereb Cortex 12: 1079–1091

Palmer AR, Winter IM, Darwin CJ (1986) The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons. J Acoust Soc Am 79: 100–113

Paninski L (2003) Convergence properties of three spike-triggered analysis techniques. Network 14: 437–464

Phillips DP (1989) Timing of spike discharges in cat auditory cortex neurons: implications for encoding of stimulus periodicity. Hear Res 40: 137–146

Phillips DP, Hall SE (1990) Response timing constraints on the cortical representation of sound time structure. J Acoust Soc Am 88: 1403–1411

Read HL, Winer JA, Schreiner CE (2002) Functional architecture of auditory cortex. Curr Opin Neurobiol 12: 433–440

Rhode WS, Greenberg S (1992) Physiology of the cochlear nucleus. In: Popper AN, Fay RR (eds) The mammalian auditory pathway: neurophysiology. Springer, Berlin Heidelberg New York, pp 94–152

Sachs MB (1984) Speech encoding in the auditory nerve. In: Berlin CI (ed) Hearing science, recent advances. College-Hill, San Diego, pp 263–307

Schnupp JW, Mrsic–Flogel TD, King AJ (2001) Linear processing of spatial cues in primary auditory cortex. Nature 414: 200–204

Shamma SA (1985) Speech processing in the auditory system. I. The representation of speech sounds in the responses of the auditory nerve. J Acoust Soc Am 78: 1612–1621

Shamma SA, Fleshman JW, Wiser PR, Versnel H (1993) Organization of response areas in ferret primary auditory cortex. J Neurophysiol 69: 367–383

Sinex DG, Henderson J, Li H, Chen GD (2002) Responses of chinchilla inferior colliculus neurons to amplitude-modulated tones with different envelopes. J Assoc Res Otolaryngol 3: 390–402

Spitzer MW, Semple MN (1993) Responses of inferior colliculus neurons to time-varying interaural phase disparity: effects of shifting the locus of virtual motion. J Neurophysiol 69: 1245–1263

Spitzer MW, Semple MN (1998) Transformation of binaural response properties in the ascending auditory pathway: influence of time-varying interaural phase disparity. J Neurophysiol 80: 3062–3076

Steinschneider M, Schroeder CE, Arezzo JC, Vaughan HG, Jr. (1994) Speech-evoked activity in primary auditory cortex: effects of voice onset time. Electroencephalogr Clin Neurophysiol 92: 30–43

Tessier-Lavigne M (2000) Visual processing by the retina. In: Kandel ER, Schwartz JH, Jessell TM (eds) Principles of neural science, 4th edn. McGraw-Hill, New York, pp 507–522

Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci 20: 2315–2331

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

Tian B, Rauschecker JP (1994) Processing of frequency-modulated sounds in the cat's anterior auditory field. J Neurophysiol 71: 1959–1975

Tian B, Rauschecker JP (1998) Processing of frequency-modulated sounds in the cat's posterior auditory field. J Neurophysiol 79: 2629–2642

Trahiotis C, Stern RM (1989) Lateralization of bands of noise: effects of bandwidth and differences of interaural time and phase. J Acoust Soc Am 86: 1285–1293

Trahiotis C, Stern RM (1994) Across-frequency interaction in lateralization of complex binaural stimuli [letter]. J Acoust Soc Am 96: 3804–3806

Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat Neurosci 6: 391–398

Versnel H, Kowalski N, Shamma SA (1995) Ripple analysis in ferret primary auditory cortex. III. Topographic distribution of ripple response parameters. Auditory Neurosci 1: 271–286

Winter IM, Wiegrebe L, Patterson RD (2001) The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig. J Physiol 537: 553–566

Woolsey CN, Walzl EM (1942) Topical projection of nerve fibers from local regions of the cochlea to the cerebral cortex of the cat. Bull Johns Hopkins Hosp 71: 315–344

Wurtz RH, Kandel ER (2000) Central Visual Pathways. In: Kandel ER, Schwartz JH, Jessell TM (eds) Principles of neural science, 4th edn. McGraw-Hill, New York, pp 523–571

Yin TCT (2002) Neural mechanisms of encoding binaural localization cues in the auditory brainstem. In: Oertel D, Fay RR, Popper AN (eds) Integrative functions in the mammalian auditory pathway. Springer, Berlin Heidelberg New York, pp 99–159

Zhang LI, Tan AY, Schreiner CE, Merzenich MM (2003) Topography and synaptic shaping of direction selectivity in primary auditory cortex. Nature 424: 201–205