Chapter 6

Time varying sounds: amplitude envelope modulations

Brian Malone and Christoph E. Schreiner

6.1 Introduction

Auditory signals are necessarily extended in time, which implies that all auditory processing can be construed as a form of temporal coding. In the case of an unmodulated 'pure' tone, however, the only temporal variation in the signal occurs at onset and offset. Of course, a pure tone consists of sinusoidal variations in pressure at the tone frequency, but the parameters defining the tone, such as its frequency and amplitude, are constant while it endures. Time-varying sounds, by contrast, exhibit fluctuations in amplitude or frequency. Nearly all naturally occurring, ecologically relevant sounds exhibit ongoing changes in their temporal structure, and answers to fundamental perceptual and informational questions regarding incoming sounds, such as 'who or what or where' require the accurate discrimination of such changes.

When discussing fluctuations in frequency and amplitude, it is useful to distinguish between the 'fine structure' and the 'envelope' of a temporal waveform. The fine structure is more closely associated with the spectral content of the signal, and refers to relatively fast pressure variations, typically exceeding several hundred Hertz (Hz). Superimposed on the fine structure are relatively slow changes in the overall signal magnitude (Fig. 6.1A). These changes define the envelope of the signal (for a mathematical treatment of the envelope, see Hartmann, 1997). For pure tones, the fine structure is a sinusoidal variation in pressure, and the envelope is said to be 'flat'. By combining the fine structure of one sound with the envelope of a different sound to form 'auditory chimeras,' researchers have demonstrated that envelope cues dominate speech recognition, and fine structure dominates pitch perception and sound localization (Smith *et al.*, 2002).

In this chapter, we focus on one class of experimental stimuli that involves relatively simple modulations of the envelope: sinusoidal modulations of amplitude. We stress that these stimuli cannot capture the full complexity of auditory temporal processing, which has also been studied with a range of informative paradigms (e.g., forward masking, gap detection, and duration discrimination). The coding of time-varying sounds is really a subset of the broader issue of how context, in general, impacts the processing of acoustic signals. There is abundant evidence that the auditory system is sensitive to the context in which a particular sound occurs (Sanes *et al.*, 1998; Galazyuk *et al.*, 2000; Malone and Semple, 2001; Malone *et al.* 2002; Ulanovsky *et al.*, 2004). All enduring sounds, in a sense, provide their own context because the features of the sound in the past bear on how it will be processed in the present due to activity-dependent biophysical mechanisms throughout the auditory pathway. The encoding of time-varying signals is thus necessarily complex because these mechanisms, as well as passive filtering mechanisms operating in cells and synapses, have their own time constants, and these too inform the neural representation of the stimulus as it passes through the system.

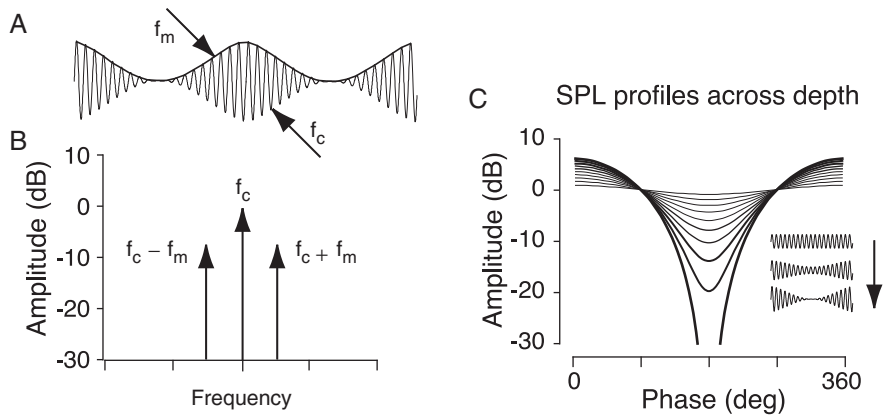

Fig. 6.1 Introduction to sinusoidal amplitude modulation. **A** Fluctuations in sound pressure that constitute a SAM stimulus. These rapid changes are referred to as the 'fine structure' of the carrier signal (f_c). Overlaid on the outline of this stimulus is the 'envelope' of the modulating stimulus (f_m), which in this case occurs at a much lower frequency ($f_m << f_c$). **B** Representation of a SAM stimulus in the spectral domain shows that the stimulus contains power at three frequencies: the carrier, f_c, and two sidebands, $f_c \pm f_m$. The relative height of the sidebands is determined by the modulation depth, m, such that a fully (100%) modulated stimulus has sidebands at –6 dB relative to the carrier. **C** Representations of SAM stimuli for a range of modulation depths with respect to an axis defined in decibels of sound pressure level (dB SPL) relative to an unmodulated pure tone. As the modulation depth increases, the asymmetry between the cyclical increases and decreases in the 'instantaneous' amplitude of SAM signals grows. For clarity, the curves have been aligned such that the amplitude minimum occurs at 180°, rather than 270°, as it does when the signals are presented in sine phase. The *icon* indicates the changes in the stimulus waveform (see **A**) as modulation depth increases.

6.2 The SAM stimulus

Because the temporal variations in natural sounds tend to be complex, auditory scientists have traditionally employed much simpler sounds to characterize the neural representation of time-varying sounds at various stages of the auditory pathway. The best-studied example of this is sinusoidal amplitude modulation (SAM), in which the amplitude of a pure tone (the carrier) is modulated by another tone at lower frequency (the modulator). Just as a pure tone is the simplest stimulus in the spectral domain because it contains only a single frequency, SAM is the simplest stimulus in the modulation domain because it contains only a single modulation frequency. A SAM stimulus is fully described by four parameters: (1) the frequency of the carrier (f_c); (2) the level of the carrier (A); (3) the frequency of the modulation (f_m); and (4) the depth of the modulation (m). The equation for a SAM stimulus can be written as:

$$s(t) = A[1 + m \sin(2\pi f_m t)] \sin(2\pi f_c t).$$

When f_c substantially exceeds f_m, the term $[1 + m \sin(2\pi f_m t)]$ describes the envelope of the stimulus, and the term $\sin(2\pi f_c t)$ describes the fine structure. By rearranging the foregoing equation using the trigonometric identity $\cos(B) \sin(A) = 1/2[\sin(A+B) + \sin(A-B)]$, we obtain a simple expression for the spectrum of a SAM signal:

$$s(t) = \sin(2\pi f_c t) + m/2\ [\sin(2\pi(f_c + f_m)t) + \sin(2\pi(f_c - f_m)t)].$$

The frequency spectrum of SAM contains three components: the primary peak at f_c, and sidebands at $f_c - f_m$ and $f_c + f_m$ (Fig. 6.1B). It is also worth noting that the average power of SAM signals increases with the modulation depth, m, which determines the magnitude of the sidebands. The time domain representation of the SAM signal indicated in Fig. 6.1A is perhaps more intuitive. The 'instantaneous' magnitude of the fully modulated SAM signal depicted here can be ascertained by considering the length of the fine structural changes within a narrow window. Although the modulator is sinusoidal in amplitude, it is important to note that the 'instantaneous' amplitude of the SAM signal, expressed in perceptually more appropriate logarithmic units of decibels of sound pressure level (dB SPL), is not. For a fully modulated stimulus, the signal will vary from +6 to $-\infty$ dB, relative to the unmodulated carrier level. Figure 6.1C indicates how the SPL of SAM signals varies as a function of modulation depth. These curves are useful in relating the amplitudes of SAM signals to measures of amplitude tuning, such as rate level functions for tone bursts, that are measured in dB SPL.

The modulation frequencies relevant for human perception span a range from ~1 to 1000 Hz. In fact, the modulation spectrum of natural sounds is highly constrained, so the 'acoustic biotopes' of many species overlap substantially, and low temporal modulation frequencies dominate communication sounds in many species (Singh and Theunissen, 2003). Different subregions of this range are associated with different perceptual and information-bearing domains. For human listeners, modulations below ~4 Hz are associated with the occurrence of essentially isolated acoustical events and each individual event or modulation cycle is perceived as a separate entity. In speech, this range corresponds to the occurrence rate of words and in music to the slower range of rhythm. Modulation rates between 4 and 20–30 Hz are perceived as fluctuations, i.e., individual cycles are still discernable but more difficult to count or to assign a precise moment of occurrence (Fastl *et al.*, 1986). In speech this range is associated with the occurrence rate of syllables and phonemes, and in music it covers faster rhythms and sequences of notes. From ~30 to ~300 Hz a sensation referred to as roughness is associated with SAMs (although that sensation is not limited to SAM signals). Each stimulus cycle is not perceived individually, but the rapid sequence of events creates a continuous, uniform sensation. Maximum roughness is usually perceived around 70-Hz modulation rate or slightly below, depending on the carrier frequency (Fastl, 1990). This range partially overlaps with that of the fundamental frequency in speech and the pitch of musical instruments. Modulation frequencies between 30 and ~800 Hz also create a percept of tonal quality, the residue or periodicity pitch. It is closely matched to the pitch produced by a pure tone with a frequency of f_m. The perceptual distinctions noted above for different ranges of the modulation spectrum underscore the notion that temporal analysis plays a critical role in identifying, segregating, and discriminating natural sounds and helps in determining the nature of the sound source, its location, and the information carried by the sound.

6.3 Analysis of neural responses to SAM: VS and MTFs

The physiological responses to SAM stimuli that have typically been measured consist of spike trains recorded from neurons at different stages of the auditory pathway. Although many methods for analyzing spike trains exist, the response measures that have been applied to SAM data almost universally are (1) the discharge rate, in spikes per second, averaged over multiple cycles of the SAM stimulus; and (2) an envelope synchronization measure that relates the time of spike occurrence to the phase of the modulating waveform. By far the most common measure of synchrony is vector strength (VS; Goldberg and Brown, 1969), which can be computed by normalizing the spectral magnitude of the peristimulus time histogram at the modulating frequency (f_m) by the average spike rate. More commonly, VS is calculated by treating each recorded spike as a

unit vector that 'points' towards the phase that the modulating waveform achieved when the spike occurred (0 to 2π). The spikes ($i = 1 \ldots n$) are summed vectorially, and normalized by the total number of spikes, n, such that:

$$VS = SQRT\ ((\sum x_i)^2 + (\sum y_i)^2) / n.$$

The angle of the resultant vector is the mean phase of the response. If all spikes occur at the same point in the modulation cycle, the VS will equal 1. If spike times are distributed uniformly in the modulation period, the VS will equal 0. However, there are many other ways for VS to equal 0, for example, spikes occurring 180° out of phase cancel each other in the vector summation. Significant synchronization is generally assessed by the Rayleigh test for circular distributions using the quantity $2nVS^2$. It is also common to report the degree of response synchronization relative to the stimulus modulation in terms of modulation gain, in decibels, as 20 log (2VS/m), where m is the modulation depth.

Because VS is integral to the vast majority of physiological studies of SAM processing, it is important to understand what aspects of the neural response VS does and does not embrace (for a more detailed discussion, see Joris *et al.*, 2004). Fundamentally, VS is a measure of how narrowly distributed spike times are within the modulation period histogram (MPH, the histogram of neural response times relative to the modulation cycle period; see Fig. 6.2). It is a measure of 'synchrony' only insofar as the spikes are understood to be synchronized to a single phase – or time point – of the modulation waveform. For example, a neuron that reliably fired spikes at two different points in the modulation waveform would yield VS values that decrease with the phase separation of those two points, reaching 0 when they are 180° out of phase. This means that higher VS values indicate 'better synchronization' only if the underlying model for the neuron's optimal encoding of SAM is a maximally peaked MPH.

Of course, there are alternatives to this underlying model. For example, the vector strength of the sinusoidal modulating waveform itself is 0.5, and that of a half-rectified sinusoid is 0.784. If the faithful representation of the modulating waveform is taken as the optimal encoding, then either of these values could be taken as 'ideal', and the shape of the MPH should be sinusoidal (with a degree of rectification set by the spontaneous rate, for example), rather than sharply peaked. Alternatively, the encoding process could be evaluated in terms of the 'instantaneous' amplitude of the SAM stimulus. For example, neurons that give sustained responses to tones might be expected to produce 'notched' MPHs in response to fully modulated SAM stimuli at low modulation frequencies, and high carrier levels; that is, the brief decrement in level occurring at 270° (for SAM modulated in sine phase) is represented faithfully by a concomitant reduction in the instantaneous probability of discharge, as shown in Fig. 6.2C, D. Despite being a faithful representation of the stimulus amplitude, VS will be very low (0.22 in the example shown in Fig. 6.2) since the cell is firing spikes at nearly all phases of the MPH. Although VS is typically used interchangeably with 'synchrony' – and we too shall uphold this convention – the reliance on VS as the sole metric for temporal encoding carries an implicit judgment that 'time-stamping' a particular phase of the modulating waveform is paramount.

A useful distinction can be drawn between the synchrony and the fidelity of the neural response. If a given SAM stimulus consistently results in an MPH with a particular shape, one would say that the given SAM stimulus is encoded with high fidelity. The foregoing 'notched' MPH is a good example of a response exhibiting high fidelity but low VS. The notion of fidelity is particularly relevant to a further distinction between the encoding of SAM generally and the encoding of modulation frequency specifically. Consider a perfect synchronizer, a neuron that always produces a VS approaching 1 for SAM stimuli. A spike-time histogram of this neuron's response

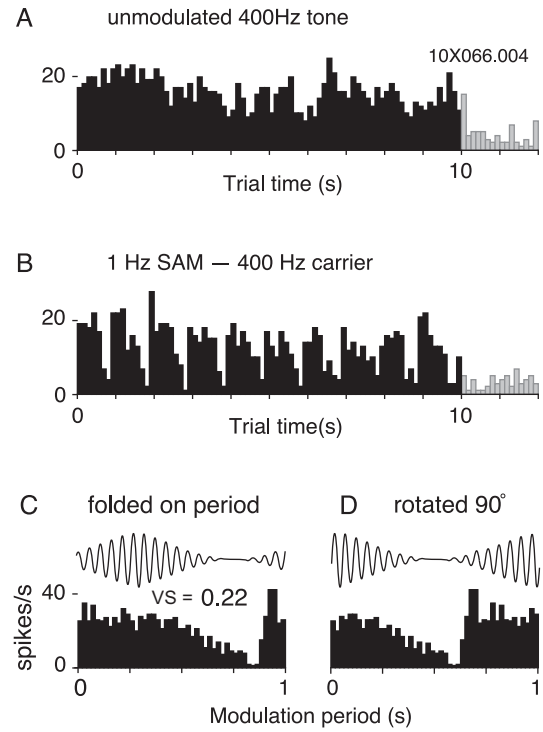

Fig. 6.2 Construction of the modulation period histogram (MPH). **A** Peristimulus time histogram (PSTH) of the responses to an unmodulated tone. Histograms of spikes occurring during the tone duration are shown in *black* and histograms of spikes occurring during the interstimulus intervals are shown in *gray*. **B** PSTH of responses to 100% modulated SAM at 1 Hz. **C** The MPH was constructed by folding the PSTH in **B** on the modulation period (1 s). The stimulus icon indicates the relative amplitude of the SAM stimulus at different phases of the modulation period. **D** The MPH shown in **C** has been rotated so that the instantaneous amplitude minimum of the stimulus is centered, and the responses at the beginning and end of the MPH in **C** are joined, as a reminder that the time axis in this representation is circular. These data were obtained from recordings of a neuron in the primary auditory cortex of an awake rhesus monkey. (With permission of Malone *et al*., 2007.)

would unambiguously indicate the modulation frequency of the SAM stimulus, regardless of what the modulation depth or carrier level happened to be. However, what if the goal were not discrimination of the modulation frequency but determination of the SAM stimulus itself, in the four-dimensional parameter space of all possible SAM stimuli? In this case, the perfect synchronizer would hinder discrimination of SAM parameters other than modulation frequency, unless those were somehow retained as differences in average firing rate, or in the mean phase of the MPH, because the shapes of the MPHs are highly constrained. As Joris *et al.* (2004: p.546) noted, VS 'does not capture the full harmonic content of the cycle histogram at f_m so that histograms with rather different shape can result in the same [VS] value'. Nevertheless, progressively higher values of VS do increasingly curtail the shapes of the MPHs that are compatible with them, which may constitute an important limit for encoding the shape of the envelope (Swarbrick and Whitfield, 1972), if not for modulation frequency coding. Thus, one cannot safely assume that

'better synchronization' is always equivalent to 'better encoding' of a time-varying stimulus – different aspects of the stimuli may require different codes.

Finally, we emphasize that VS is an analytical tool developed for a limited set of experimental contexts and requires knowledge of f_m, when one of the tasks of the brain, presumably, is estimation of f_m. More importantly, the notion of synchrony embodied by VS is inapplicable when the envelope of the acoustic signal is nonperiodic, as is the case for most natural sounds. Alternatives to VS and VS-based metrics have been developed recently (e.g., Kajikawa and Hackett, 2005; Joris *et al.*, 2006; Malone *et al.*, 2007) and we will discuss their application in a later section.

The predominant response metric for neural responses to SAM is the modulation transfer function (MTF), which describes how the response changes as a function of different modulation frequencies. Typically, responses to SAM are characterized in terms of average firing rate, VS, and mean phase; when these quantities are plotted against modulation frequency, we shall refer to them as the rate MTF (rMTF), temporal MTF (tMTF), and phase MTF (pMTF), respectively. In some cases, the average firing rate and the VS are combined, via multiplication, into what is termed the 'phase-locked rate' or 'synchronized rate'. Note that the tMTF, pMTF, and synchronizarion rate all depend on VS, and inherit its limitations.

Each MTF represents a slice through the four-dimensional SAM parameter space. MTFs are most commonly obtained for fully modulated ($m = 1$) SAM stimuli at the neuron's characteristic frequency, and either best level, or at a fixed level with respect to the neuron's response threshold. Many of the studies we review in the following sections have also reported MTFs at a range of carrier levels and modulation depths, and in a few cases carrier frequencies. A central goal of research in this vein has been characterization of the changes in temporal coding that occur as one records from neurons at successively more central nuclei in the auditory pathway.

Analysis of corresponding changes in the shape of the MTF has been central to this endeavor. This often involves categorization of MTF in terms of filter shape categories such as lowpass, bandpass, or highpass (and, in some instances, flat or band-reject, etc.), and identification of the best modulation frequencies (BMFs) for both rate and synchrony. The representativeness of these summary measures, if derived from a single MTF, will depend on the extent to which MTF shape is invariant to changes in SAM parameters other than f_m. Importantly, MTFs have nothing to say about the reproducibility of the responses, nor do they provide an explicit means for determining how well different SAM stimuli can be discriminated from one another (Wohlgemuth and Ronacher, 2007).

In the next few sections, we will examine how the nature of SAM coding varies across different nuclei in the ascending auditory pathway, with an eye towards relating changes in temporal coding to other general changes in the response properties of more central neurons. An important confound to keep in mind is the increasing effects of anesthetics on more central structures (Ter-Mikaelian *et al.*, 2007), as well as other factors such as the increasing prevalence of nonmonotonic tuning for SPL (Semple and Kitzes, 1993a, b; Phillips *et al.*, 1994; Malone *et al.*, 2007) and transient responses to tonal stimuli. Our goal is not a comprehensive review of the literature (for that see Joris *et al.*, 2004), but rather a survey of the implications for temporal coding revealed by a selection of representative studies.

6.4 Auditory nerve

In a comprehensive study of SAM coding in cat auditory nerve (AN), Joris and Yin (1992) noted, 'To a first approximation, the behavior of AN fibers to changes in SPL of an AM signal can be predicted by considering the compressive shape of its rate-level curve as an input-output curve' (see also Yates, 1987). Since AN fibers contain all the information about the auditory signal that

will ever be available to central processors, this simple model represents a baseline for evaluating central transformations in how SAM is encoded. For example, Smith and Brachman (1980) demonstrated much earlier that for some modulation frequencies (roughly 150–300 Hz) the modulation gain of AN fibers could be predicted fairly accurately by considering the slope of the rate level function in the appropriate SPL range.

The activity recorded from AN fibers in response to SAM stimuli closely resembles the actual stimulus waveform. If the carrier frequency is below a few kiloHertz, this representation includes phase locking to both the fine structure and the envelope of the modulating waveform (Joris and Yin, 1992; Fig. 6.3A). Because f_m is quite low (10 Hz, m = 0.99) in this example, the relationship with the instantaneous amplitude is particularly clear: as the carrier level is increased from 39 to 104 dB SPL, the SAM stimulus falls below the fiber's threshold for a smaller fraction of the MPH. This can be seen in the progressive narrowing of the notch near 270° (0.75 cycles), which also results in a dramatic reduction in VS. This nonmonotonic relationship between VS and stimulus level was universal in AN fibers. Note that in low CF fibers, phase locking to the carrier increased monotonically with stimulus level, indicating that the reduction in envelope VS is not a loss of temporal resolution per se. Average rate responses increased monotonically with level, as expected from the sigmoidal shape of AN rate-level functions. The mean phase of the MPH evidenced slight but measurable increases in phase lead with respect to the envelope, an effect that the authors attributed to adaptation. Finally, changes in the carrier frequency were shown to be essentially equivalent to reducing the stimulus level, since using a non-optimal f_c shifted the modulated rate and VS functions to the right on the SPL axis.

Another universal feature of AN responses to SAM is the monotonic increase in VS for increasing modulation depths. As is evident in Fig. 6.3B, MPHs obtained from AN fibers tend to be relatively isomorphic to the half-rectified SAM waveform. At low modulation depths, AN responses exhibit more modulation than the SAM envelopes (i.e., have a modulation gain > 0 dB), but all AN fibers exhibit monotonic increases in VS when the modulation depth is increased. The foregoing observations reflect the general applicability of an amplitude-based model for predicting how AN fibers will behave when SAM parameters other than f_m are varied. Historically, however, the primary motivation for most SAM studies relates to modulation frequency coding. Once again, AN fibers exhibit highly stereotyped responses to increases in modulation frequency. rMTFs are generally flat (see below), and tMTFs are lowpass, though with a shallow decline toward low f_m. As was discussed, modulation gain is highest for low modulation depths and low SPLs, so modulation gain functions (i.e., 20 $\log_{10}$ (tMTF/m)) from the same fiber but obtained at different modulation depths, or carrier levels, would be vertically displaced on a modulation frequency axis like that of Fig. 6.3C, where a single modulation gain function is shown. Because of spectral filtering, modulation gain functions from fibers that differ in CF would be horizontally displaced on the f_m axis. Recall that the spectrum of a SAM signal has sidebands at $f_c \pm f_m$ (Fig. 6.1B). As f_m increases, the sidebands become increasingly distant from f_c, and eventually fall outside of the spectral filter bandwidth of the AN fiber. Since spectral bandwidth is proportional to CF, the 3-dB corner frequency of the modulation gain function is proportional to both CF and tuning curve bandwidth, with maximal values as high as 1500 Hz. Nevertheless, there was a saturating trend that suggested that temporal factors limited synchronization at high CFs. It is also worth noting that both VS maxima and the VS cutoffs for f_c were significantly greater than those obtained for f_m, suggesting that the mechanisms limiting temporal resolution to the carrier fine structure and the modulation envelope may not be strictly equivalent (for a discussion see Joris and Yin, 1992).

The pMTFs of AN fibers were also highly stereotyped and, in this case, linear. When the cumulative phase of the response is plotted against f_m, the slope of the resulting function provides an

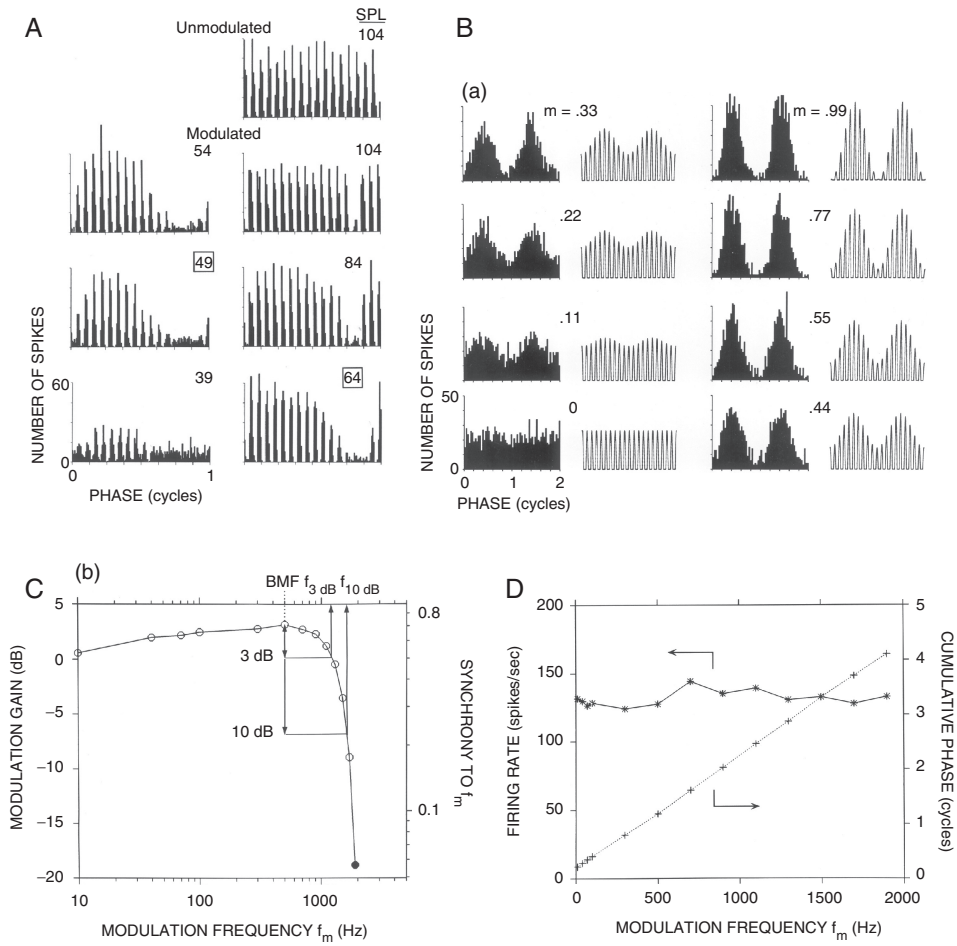

Fig. 6.3 Responses of auditory nerve (AN) fibers. **A** and **B** MPHs of the responses of an AN fiber to SAM at increasing carrier levels (indicated by the *insets*, in decibels of sound pressure level (dB SPL)). These responses exhibit phase locking to the carrier (170 Hz) and to the modulation (10 Hz). As the carrier level increases, phase locking to the carrier remains robust, and those phasic responses occupy greater portions of the modulation cycle, vanishing only during a narrow portion of the modulation period near the amplitude minimum of the stimulus (0.75 cycles, or 270°, for SAM in sine phase), as in Fig. 6.2B). MPHs comprising two modulation cycles are compared to half-wave rectified illustrations of SAM waveforms across modulation depth (f_m = 100 Hz; SPL = 49 dB; f_c = 20.2 kHz). **C** The modulation transfer function (MTF) of a typical AN fiber is shown relative to synchrony (VS; tMTF), and modulation gain (20 log (2VS/m)). The 3- and 10-dB cutoff values are indicated by *arrows*. **D** The MTF for the responses depicted in **C** is shown plotted against firing rate (rMTF) and cumulative phase (pMTF; note that a linear rather than logarithmic f_m axis is shown to indicate the linear relationship between cumulative phase and f_m). These data were obtained in the auditory nerve of anesthetized cats. (With permission of Joris and Yin, 1992.)

estimate of the cochlear group delay, which was found to vary inversely with the CF of the fibers (approximately 7 to 2 ms), as expected from the propagation of the traveling wave along the basilar membrane (Fig. 6.3D). The y-intercept of these functions converged on 0.25 cycles, which represents the instantaneous amplitude maximum for SAM signals presented in sine phase.

Thus far, we have treated AN fibers as homogenous – which, relative to central auditory neurons, they are – but AN fibers do fall into two groups, defined largely by differences in their spontaneous rates (SRs). Maximal VS values were inversely correlated with SR, as one might expect from the simple notion that spontaneous spikes distributed randomly throughout the modulation period would tend to reduce VS. In addition, low SR fibers, but not high SR fibers, typically showed a (>20%) reduction in average firing rate as f_m increased. However, Cooper *et al.* (1993) noted that the synchronized rate of high SR fibers exceeds that of low SR fibers due to their higher firing rates overall. In the absence of knowledge of how SAM is encoded – given that it is highly unlikely that VS can be computed by the brain at all – it is difficult to parse the relative contributions of high and low SR fibers to this process. Nevertheless, it does seem likely that the existence of these two fiber classes serves to extend the range of amplitude changes that can be encoded.

6.5 Cochlear nuclei

Møller's (1972, 1974, 1976) pioneering studies of the cochlear nuclei (CN) documented high fidelity representations of the stimulus envelopes, enhanced modulation gain (relative to the AN) over large ranges of carrier levels, higher tolerance for background noise, and impressive constancy of tMTF shapes for different stimulus types, such as SAM applied to both tones and noise carriers, and both noise-modulated tonal and noise carriers. Neurons of the CN also exhibit a striking degree of diversity in response to pure tones relative to the AN, and these physiological differences have been successfully mapped to several distinct morphological classes. Typically, the CN is further subdivided into anteroventral (AVCN), posteroventral (PVCN), and dorsal (DCN) nuclei. Unsurprisingly, differences in pure tone responses among the cell classes of the CN also extend to SAM stimuli.

The different response classes, their morphologies, and their distributions within the divisions of the CN have been reviewed in detail elsewhere (see Frisina, 2001). We focus on the aspects of SAM responses in the CN that diverge most powerfully from those of AN fibers. Figure 6.4 captures a number of the differences, which have been evaluated almost entirely in the context of tMTFs. In addition to AN fiber responses (A), the seven major response classes depicted here are as follows (see Blackburn and Sachs, 1989; Rhode and Greenberg, 1992; Chapter 2, this Volume): B – primary-like (spherical bushy, AVCN); C – onset (octopus, PVCN); D – primary-like with notch (globular busy, AVCN); E – onset-chopper (multipolar stellate, PVCN); F – chopper-sustained (stellate); G – chopper-transient (stellate); and H – pauser/buildup (fusiform, DCN). At a glance, the major differences between the AN (A) and the CN (B–H) are the increased prevalence of bandpass-tuned tMTFs and the reduction in the VS cutoffs. As we will see, both of these trends remain consistent at progressively more central auditory structures.

Traditionally, the relative increase in VS has received the most attention (e.g., Frisina *et al.*, 1990), and this increase is most pronounced at higher sound levels, which significantly depress VS values in the AN. In a comprehensive study of the CN of the cat, Rhode and Greenberg (1994) ordered the various CN cell classes largely in terms of their tMTFs. Obviously, primary-like neurons responded much like (high SR) AN fibers and had the highest synchrony cutoffs. Chopper, onset-L, and pauser/buildup were described as being roughly comparable to low SR fibers, while onset-chopper and pauser/buildup neurons possessed 'considerably enhanced' phase-locking to f_m relative to other CN neurons.

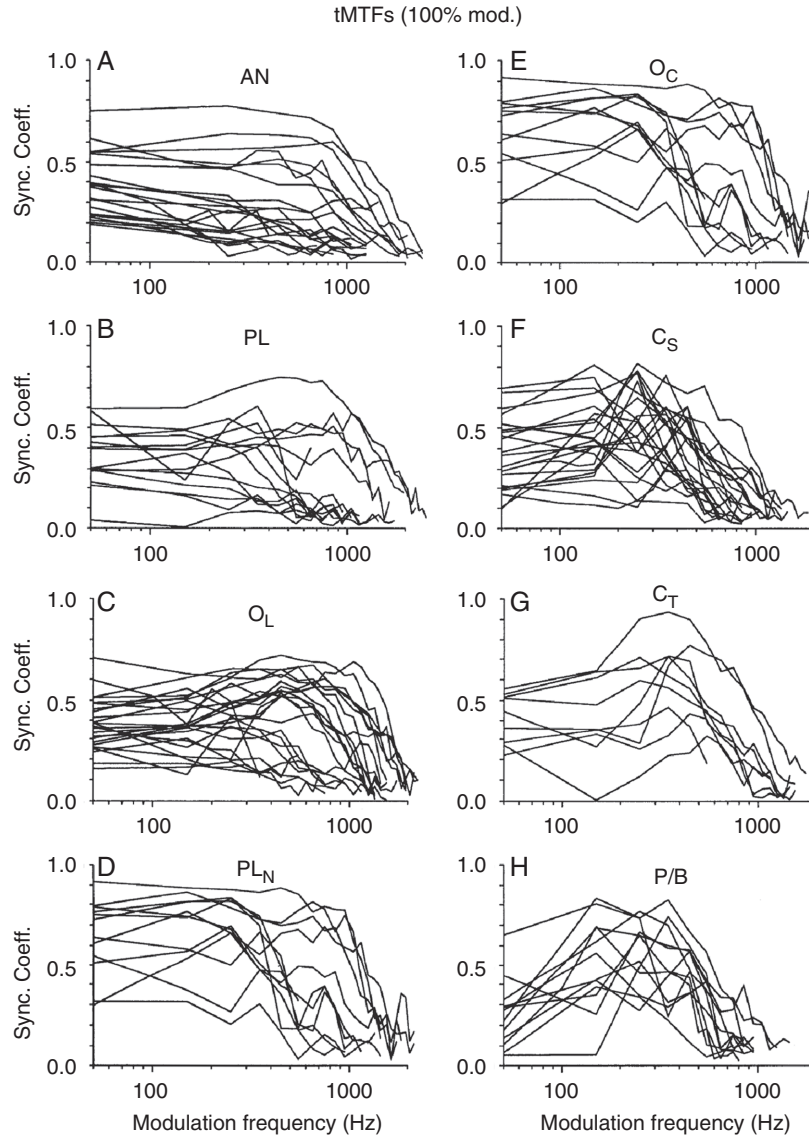

Fig. 6.4 Comparison of tMTFs obtained in the auditory nerve (AN) (**A**) and cochlear nucleus (CN) (**B–H**). The tMTFs obtained from different cell classes in the AN and CN indicate the diversity in the MTF shapes in the early auditory system. Each *curve* represents the response of a different cell, and all stimuli were 100% modulated and presented at 50 or 60 dB SPL. The cell classes were: **A** auditory nerve; **B** primary-like; **C** onset-L; **D** primary-like with notch; **E** onset-chopper; **F** chopper-sustained; **G** chopper-transient; **H** pauser/buildup. These data were obtained in the CN of anesthetized cats. (With permission of Rhode and Greenberg, 1994.)

Nevertheless, it is not entirely clear what this 'enhanced synchrony' means for temporal coding. A class of onset cells (Oi) (Rhode, 1994) shows the highest modulation gain of the CN cell classes (i.e., the amount by which modulation depth in the response exceeds the modulation depth in the stimulus). Their tMTFs also appear to be invariant to SPL, and many even exhibit sharply bandpass rMTFs with rBMFs from 300–500 Hz. Relative to AN fibers – or the primary-like neurons of the CN – Oi units appear to engage in a distinct and novel form of temporal coding. Specifically, they appear to embody the 'perfect synchronizer' discussed previously, since the invariance of their tMTFs to carrier level also implies a loss of information about carrier level. Conversely, the low VS values obtained from primary-like neurons for low depth, high level SAM signify the retention of information about 'instantaneous' SPL. The responses of onset (and onset-chopper) units, by contrast, sacrifice the encoding of amplitude in favor of encoding changes in amplitude.

The remaining cell classes in the CN fall between these extremes, and it has been proposed that this menagerie comprises a set of parallel pathways for auditory processing. It seems plausible that the different cell classes are specialized for extracting different features of acoustic signals, such as the envelope shape versus the stimulus periodicity. Before accepting the notion that higher VS values represent a true 'enhancement' of temporal coding in the CN, however, it would be instructive to perform an analysis that explicitly tests how well SAM stimuli can be identified on the basis of spike trains obtained from different cell classes in the CN, particularly for the (low) f_m ranges where low VS values for high carrier levels reflect sustained firing throughout the modulation period. It is worth noting that Rhode and Greenberg's (1994) study was restricted to f_m above 50 Hz, which is substantially higher than the modulation range (>20 Hz) most important for communication sounds, including human speech (e.g., Drullman *et al.*, 1994; Drullman, 1995).

6.6 Superior olivary complex and lateral lemniscus

Relatively few studies have examined responses to SAM in the superior olivary complex (SOC) or the nuclei of the lateral lemniscus (NLL). Kuwada and Batra (1999) recorded from monaural units in the SOC of awake rabbits. These units were recorded within the SOC, but outside the principal binaural nuclei of the SOC and the lateral and medial superior olivary nuclei. The authors observed two distinct response types with respect to pure tones. 'Sustained' neurons responded in a sustained fashion to tones, and included a number of the response types observed in the CN (chopper, pauser, etc.). 'Off' neurons ceased to fire during tone presentation, but fired a 'rebound' response at tone offset, a discharge pattern that has rarely been observed in the CN. Both of these response classes exhibited wider dynamic ranges for sound amplitude than did the neurons of the CN.

The responses of sustained and off neurons are essentially complementary. Sustained neurons respond at the envelope maximum (0.25 cycles for signals in sine phase), which suggests that their responses reflect excitation, whereas off neurons respond immediately prior to the envelope minimum (0.75 cycles), suggesting that their responses reflect a rebound from inhibition that prevailed during the period of maximum amplitude. The differences in the shape of the sustained and off MTFs are consistent with these observations. The most obvious difference – the dramatic relative increase in VS values for off neurons – reflects the fact that the envelope trough occupies a relatively small fraction of the modulation period. As a result, off neuron VS values were very high (>0.9) even at the lowest tested f_m (25 Hz), resulting in lowpass tMTFs. Sustained neurons, by contrast, exhibited bandpass tMTFs at moderate carrier levels, and had modulation gains statistically comparable to those measured in the CN. Off neuron tMTFs were uniform across modulation depth (Fig. 6.5) because of the strong constraints on when spikes could occur (i.e., during rebounds from high amplitudes).

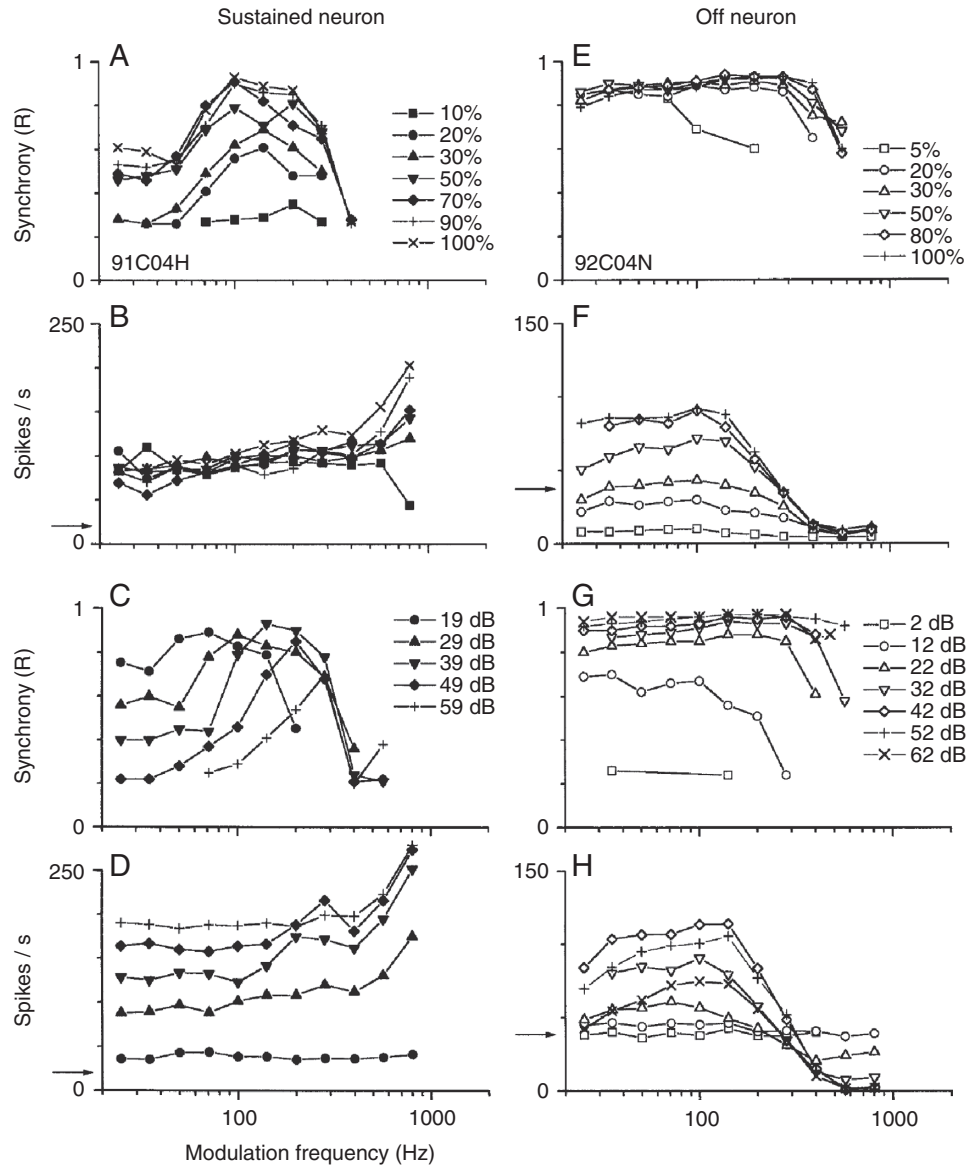

Fig. 6.5 Comparison of tMTFs (**A**, **C**, **E**, **G**) and rMTFs (**B**, **D**, **F**, **H**) obtained from a sustained (*left column*) and an off (*right column*) neuron of the superior olivary complex (SOC). The striking differences between the rMTFS and tMTFs of these two neurons indicate that the character of the response to pure tones crucially informs the responses to modulated stimuli. Modulation depth was varied as indicated in the *upper four panels* (**A**, **B**, **E**, **F**), using a stimulus intensity of 29 dB SPL for the sustained neuron and 28 dB SPL for the off neuron. Carrier level was varied as indicated in the *lower four panels* (**C**, **D**, **G**, **H**), using a modulation depth of 80%. *Horizontal arrows* indicate the spontaneous rate. These data were obtained in the SOC of awake rabbits. (With permission of Kuwada and Batra, 1999.)

Off neurons showed a strong monotonic dependence of average rate on modulation depth – at low depths the envelope troughs are rarely sufficient to release the neuron from inhibition, resulting in few spikes, but in contrast there was a nonmonotonic dependence of average rate on carrier level. Low carrier levels produce less inhibition, limiting the magnitude of a potential rebound. At sufficiently high carrier levels, envelope troughs (for m = 80%) might not be deep enough to release the neuron from inhibition (alternatively, level-related latency effects could apply, as suggested by the authors). Off-neuron tMTFs were monotonic with level. Sustained neurons, by contrast, showed monotonic increases in average firing rate with increasing carrier level, as well as the familiar lowpass to bandpass transition caused by reduction in VS at low (<100 Hz) f_ms. Interestingly, the tBMFs for sustained neurons often shifted to higher values as the level was increased. The lack of invariance of the tBMF to level changes is relevant because it is required if neurons are to function as 'modulation filterbanks' for signal envelopes.

It was only very recently that responses to SAM were recorded from the ventral nucleus of the lateral lemniscus (VNLL; Batra, 2006; unanaesthetized rabbits), a major source of monaural inhibition to the inferior colliculus. The responses there fell into two main categories: neurons with sustained responses to tones typically had flat rMTFs, and lowpass or bandpass tMTFs. Neurons with onset responses to tones, by contrast, had bandpass rMTFs and 'flat' tMTFs. Such bandpass rMTFs are reminiscent of those described for Oi neurons of the CN (weak bandpass tuning was also observed for a few onset-choppers (Rhode, 1994; Rhode and Greenberg, 1994)), and much is made of their appearance because of the notion that a conversion from a 'temporal code' to a[n] '[average] rate code' for modulation frequency occurs in the ascending auditory pathway (Langner and Schreiner, 1988; Hewitt and Meddis, 1994). Batra (2006) also cites the wide distribution of rBMFs (14–283 Hz) as being compatible with a rate-based modulation filterbank in the subcortical auditory system.

There is an obvious correspondence between tone PSTH shapes and MTF shapes, and the responses of onset neurons appear compatible with the notion that these cells fire an onset response to each cycle of the (100% modulated) SAM stimulus, as if each cycle were a separate tone. The mean response phase coincided with the rising phase of the envelope, and the population mean of the slopes of the function relating f_m to discharge rate (below the rBMF) was 1, indicating that the firing rate was roughly proportional to the number of such 'tone onsets'. Above the rBMF, however, it would appear that these pseudo-onsets occur too rapidly with regard to local integration and/or adaptation time constants to elicit spikes effectively.

In the context of this explanation, one must ask whether this represents a true 'temporal to rate' conversion for the representation of modulation frequency. Instead, one could also refer to it as an 'instantaneous' rate code for rapid amplitude increases, since the rBMF is essentially set by a temporal resolution limit for f_m. When nearly all spikes are elicited by large, rapid, excursions in amplitude, then modulation frequency tuning for SAM cannot be differentiated from sensitivity to the rise-time ('attack') of sound envelopes, which is not 'temporal' filtering as it is typically understood. More importantly, if restrictions on the carrier level or modulation depth of SAM stimuli effectively eliminate the responses of such neurons, this would complicate the argument that they encode modulation frequency, rather than simply responding to sufficiently large amplitude transients. The currently available data are not sufficient to disambiguate these competing interpretations. More fundamentally, however, it is not clear that one should describe such responses as 'encoding' the modulation envelope at high gain, since only a single feature of the envelope – its rising phase – is represented in the spike train.

6.6 Inferior colliculus

Setting aside the teleology of bandpass rMTFs, their prevalence in SAM responses obtained from neurons in the inferior colliculus (IC) has been extensively documented, at least in the central nucleus (ICc). As an obligatory relay of the primary lemniscal pathway, the IC performs an important integrative role, gathering inputs from subcortical and cortical areas. By the level of the IC, the response characteristics of the typical neuron are quite different from those observed in AN fibers. For example, the relationship between the CF and the synchrony cutoff no longer strictly prevails, as is true of many of the response classes in the CN, excepting the primary-like and onset-L neurons. Rees and Møller (1987) noted that the restriction of VS cutoffs to values below 360 Hz in the IC could not be explained by bandwidth limitations, and suggested instead that central limitations on VS for high f_ms were related to the 'number of synapses interposed between the sound input and the colliculus.' (p. 140). On average, VS cutoffs drop by about one half between the CN and ICc, and these authors reported modal tBMFs in a range from 100–120 Hz (anesthetized rat). In the anesthetized gerbil, only 15% of neurons had significant VS above 300 Hz (Krishna and Semple, 2000). This progressive reduction in VS cutoffs is the most obvious, and perhaps most important, trend in the ascending auditory pathway (see Joris *et al.*, 2004, their Fig. 9).

At low modulation frequencies and low modulation depths, however, the opposing trend of increased VS from AN to CN also continues in the IC. Relative to the CN, IC neurons generally showed higher maximum VS values and higher VS for low modulation depths (with the exception of CN onset-choppers), which compresses the metric's dynamic range and steepens the decline of modulation gain with increasing modulation depth (Krishna and Semple, 2000). As was true of the CN, a qualitatively similar transition from lowpass to bandpass tMTFs has been observed in the IC by a number of investigators in a number of species. The tMTFs of a few units did not exhibit this shift; in keeping with the role of the IC as an integrative center, Krishna and Semple (2000) recorded both onset neurons (their Fig. 5) and a single offset neuron (their Fig. 8) whose responses were strongly reminiscent of the SAM responses of VNLL onset and SOC offset neurons described above, albeit shifted to lower ranges of f_m.

The shapes of tMTFs are also sensitive to changes in carrier level and carrier frequency. Rees and Palmer (1989) demonstrated that it was possible to reverse the lowpass to bandpass shift in the tMTF shape by adding broadband noise (critically, however, neurons that were driven by continuous noise were excluded from this analysis). They also demonstrated that neurons with nonmonotonic tuning for sound level could also exhibit a return from bandpass to lowpass tMTFs at high levels. The increased heterogeneity of tuning for sound level in the IC (Semple and Kitzes, 1993a,b) may explain the increased heterogeneity of level-based tMTF changes there (Krishna and Semple, 2000). Changes in carrier frequency tended to reproduce changes in tMTF shape observed for changes in carrier level in a given neuron, such that mismatching the carrier frequency to the BF was roughly equivalent to reducing the SPL. Interestingly, the high f_m slope of tMTFs is relatively invariant to parametric changes in SAM stimuli, suggesting that the reduction in VS at high f_ms reflects genuine and inflexible limits on temporal resolution, rather than the 'artificial' reductions in VS that occur when neurons track amplitude changes throughout the modulation cycle at low f_ms.

Relative to other brainstem nuclei, the rMTFs of IC neurons are more commonly and more sharply tuned to f_m (Langner and Schreiner, 1988). In the anesthetized gerbil, where the shapes of rMTFs were examined in most detail, the maximum rBMF was 140 Hz and over half were less than 25 Hz. However, in cat IC, approximately 25% of the rBMFs were above 100 Hz and as high as several hundred Hertz (Langner and Schreiner, 1988). Such differences in the overall estimate

of temporal coding capacity may be influenced by species-specific differences and also by the uniformity of spatial sampling throughout a given structure due to the observation that the distribution of response preferences may not be spatially uniform (Schreiner and Langner, 1988).

The shape of the rMTFs of the IC are also substantially more diverse (Fig. 6.6). For example, Krishna and Semple (2000) observed not only rBMFs, but also f_m ranges where the firing rate was significantly suppressed ('worst' modulation frequencies, or WMFs) in slightly less than half of IC neurons, and occasionally (eight neurons) a secondary f_m range where the average firing rate resurged, though VS was low. As was true of the VNLL, particular rMTF shapes tended to be associated with particular tone response classes: the rMTFs of onset and onset-sustained neurons contained a suppressive region more rarely (3/21) than did sustained or pauser neurons (19/24). In general, the shapes of rMTFs were constant, but scaled in magnitude as the modulation depth increased. Experiments that varied the carrier frequencies for SAM were also used to demonstrate that the suppressive regions were not related to sideband inhibition, suggesting a temporal basis for the phenomenon.

Changes in carrier level tended to produce more complex changes in IC rMTFs, including substantial variation in the rBMF. Krishna and Semple (2000) did not find a consistent trend, but did report that the variation was substantial, exceeding 66% of the mean BMF in half of the neurons tested over an SPL range of 20 dB or more. This finding poses a difficulty for theories that view the tuning of rMTFs as evidence for neural filters tuned to particular f_ms. There is a correlation between the minimum response latency to tones and the rBMF in the IC, but it appears to be different from that observed in the AN (see above), since the correlation between response latency and CF no longer prevails in the IC. There is also a correlation between the rBMF and the shape of the tMTF, such that neurons that have a clearly defined rBMF typically exhibit maximal VS at the same f_m.

Of course, it should be noted that if there is reason to believe that firing rates of neurons with transient responses to tones increase quasi-linearly up to a fixed limit of temporal resolution – as suggested by the onset neurons of the VNLL, for example – the foregoing correlation between the rBMF and tBMF is to be expected. It would be of interest to know the extent to which correlations of this sort are predictable in terms of PSTH classifications such as onset versus sustained. Similar analyses would apply to the magnitudes of lowpass to bandpass shifts with increasing SPL at low f_ms, too.

In fact, the relationship between firing rate and VS is an important one, in part because the two show increasing interdependence in the ascending auditory pathway. In the AN, where rMTFs are flat, the relationship is absent: if there is no rBMF, it cannot predict the tMTF. The relationship is likely to reflect adaptive mechanisms sensitive to recent history of stimulation (or response).

Rees and Palmer (1989) reported a similar finding –43% of IC neurons had coincident tBMFs and rBMFs. Nevertheless, they also found that modulation gain for f_ms less than 50 Hz was inversely correlated with the discharge rate elicited by either tones or SAM at equivalent levels. Generally speaking, lower discharge rates favor higher values of VS because of how VS is calculated. Krishna and Semple (2000) noted that the suppressive regions of the rMTF where firing rates were low were also often associated with high VS values. These findings seem contradictory until one considers that the typical tBMF peaked between 50 and 100 Hz (Langner and Schreiner, 1988; Rees and Palmer, 1989). The relevant metric is the spikes per modulation cycle elicited by a SAM stimulus, because each additional spike in the same modulation cycle must occur at a different phase, reducing the VS in proportion to its difference from the mean phase. As one ascends the auditory pathway, however, there is a gradual reduction in absolute discharge rates. In the f_m range where a neuron fires less than a single spike per cycle, on average, an increase in firing rate introduced by increasing the carrier level may not significantly impact the VS because the

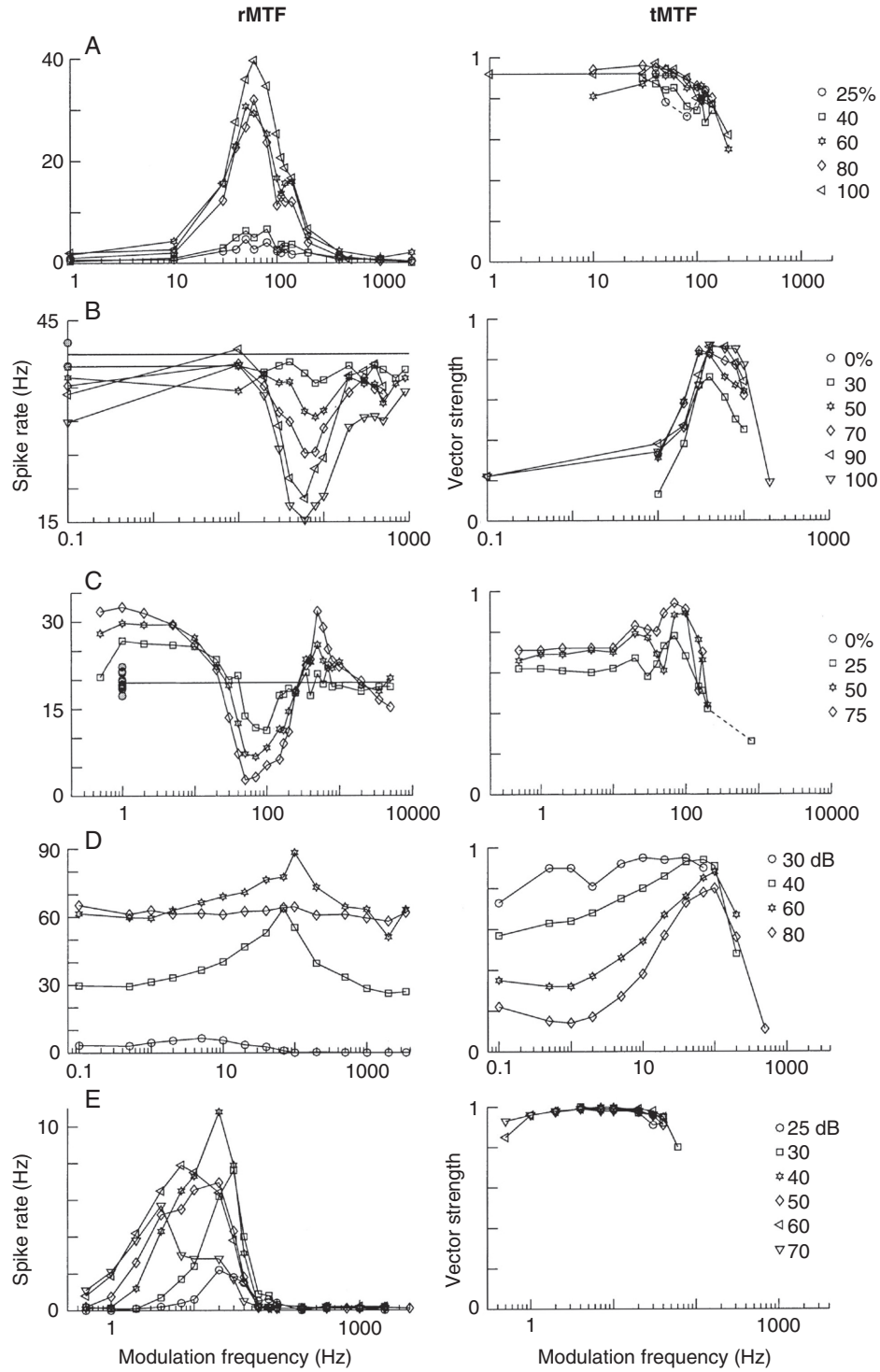

Fig. 6.6 Comparison of MTFs obtained in the inferior colliculus (IC) across modulation depth (**A**–**C**) and carrier level (**D** and **E**) for five single neurons. rMTFs and tMTFs are shown side by side, and the *symbols* used to plot each *curve* indicate the stimulus values. The stimulus intensities for **A**, **B**, and **C** were 50, 70, and 40 dB SPL, respectively. Modulation depth was 100% for **D** and **E**. **A**–**C** show cells that show bandpass enhancement and/or suppression with change in modulation depth. **D** and **E** show level dependencies in sustained neurons (**D**) and onset neurons (**E**). These data were obtained in the central nucleus of IC in anesthetized gerbils. (With permission of Krishna and Semple, 2000.)

'additional' spikes may fall in cycles where spikes would otherwise not have occurred. At very low modulation frequencies, however, IC neurons commonly fire multiple spikes per modulation cycle. At a minimum, refractory periods enforce a degree of scatter in phase, such that 'additional' spikes at higher levels are likely to reduce the VS. Although it is common to think of VS in terms of temporal resolution and filter bandwidths, it must be remembered that all VS calculations ultimately derive from spikes, and the number of spikes per cycle represents an ineradicable constraint on temporal coding.

6.7 Medial geniculate and auditory cortex

Relatively few studies have examined responses to SAM in the medial geniculate body (MGB). The bulk of more recent work has involved click trains or more complex dynamic stimuli. The most detailed of these (Preuss and Müller-Preuss, 1990), conducted in awake squirrel monkeys, was largely focused on MTF shape classification and tBMF determination, and reported a predominance of bandpass rMTFs and tMTFs, with tBMFs ranging from 2–128 Hz, with a mode at 32 Hz. On the other hand, SAM responses have been extensively studied in auditory cortex in a range of preparations and animal models. Because anesthetics affect cortical responses far more than those of subcortical nuclei (Ter-Merkaelian *et al.*, 2007), however, the results of studies conducted in awake and anesthetized animals differ (for a discussion of such differences see Liang *et al.*, 2002). For example, responses in awake animals extend to higher f_m ranges – VS cutoffs in awake rhesus monkeys could be as high as several hundred Hertz in a few cases (>600 Hz), and roughly one in ten neurons had significant VS at 200 Hz (Malone *et al.*, 2007). Comparable results were obtained for awake marmosets (Liang *et al.*, 2002). Nevertheless, both of the foregoing studies indicate that relative to the MGB and IC, auditory cortex does continue the trend of emphasizing very low (<20 Hz) modulation frequencies in terms of both average discharge rate and VS. Liang *et al.* (2002) reported that the vast majority (>80%) of tBMFs fell below 20 Hz; the distribution of rBMFs was wider, and distributed among higher values, with a mode at 16 Hz. Malone *et al.* (2007) found rBMF and tBMF modes at 5 Hz in macaque primary auditory cortex (AI). Different auditory cortical fields in the cat and squirrel monkey all follow the same general trend of low BMFs, but tend to have slightly different ranges of temporal encoding capacity, with BMFs in core or primary areas about twice that of belt or non-primary areas (Schreiner and Urbas, 1988; Bieser and Müller-Preuss, 1996).

The observation of progressive reduction in the upper limits of synchronized temporal encoding cannot solely be explained by an increase in the number of intervening synapses accompanied by increased temporal transmission jitter. Behavioral training of owl monkeys in an SAM discrimination task resulted in a doubling of their cortical BMFs and limiting frequencies (Beitel *et al.*, 2003). This suggests that the observed cortical emphasis on slower modulations is not a physiological limitation, but reflects specific processing strategies for different tasks as, for example, in auditory object formation and consideration of temporal contexts.

6.8 Codes of timing and codes for tasks

Having reported summary measures such as the rBMFs and tBMFs throughout the auditory pathway, it is necessary to ask again what such values mean for the encoding of time-varying sounds. In the context of a modulation filterbank hypothesis, BMFs capture the tuning of putative filters for different envelope periodicities. The relative degree of tuning expressed in the rMTF versus the tMTF has been thought to indicate the relative balance of 'temporal' and 'rate' codes for modulation frequency. The importance of (average) rate codes in the auditory pathway

has been emphasized recently, and it has been postulated that two separate coding populations in auditory cortex exist: a synchronized population that encodes f_m via phase-locked responses, and a non-synchronized population that encodes f_m solely in terms of average discharge rate (for a review see Wang, 2007). In other words, the temporal-to-rate conversion is complete in the neurons of the non-synchronized population. This represents a specific instance of the more general transformation from a neural representation that is isomorphic to the stimulus, to one that is not (in principle, such a transformation could also exchange one temporal code for another). Currently, the evidence for this hypothesis is based primarily on responses to periodic click trains (Lu *et al.*, 2001). Although Liang *et al.* (2002) identified a substantial fraction (30–40%) of AI neurons in the awake marmoset within the non-synchronized population, neurons that exhibited rate tuning in the absence of synchrony at any tested f_m were rare (1.7%) in the cortex of rhesus monkeys (Malone *et al.*, 2007). Nevertheless, a significant percentage (16%) of such neurons did show significant variations in average discharge rate beyond their VS cutoffs, indicating that changes in firing rate were not strictly limited to modulation frequencies that elicited synchronized responses.

The suitability of different encoding schemes can be evaluated by determining the accuracy they allow for decoding stimulus information. For example, one can test the quality of 'rate' coding of SAM explicitly by asking how effectively one can guess the modulation frequency presented on a given trial based on the neural response (Foffani and Moxon, 2004).

Briefly, the responses to several different SAM frequencies are each summed and binned to create PSTH templates. Then, the response to a 'test' trial of a given SAM frequency is matched to the most similar template (e.g., using a Euclidean distance measure), and the stimulus used to generate the matching template is guessed, resulting in a confusion matrix of the actual stimuli and the guesses. For spike train classifiers of this sort, an average rate code is simply the limit case of a single bin whose width equals the 'test' duration, because the bin width determines the temporal precision of the code. Conversely, one can also eliminate average discharge rate information by normalizing the total spike counts across the templates, while preserving the relative distribution of spikes across the chosen time interval (the 'phase-only' classifier). In effect, this analysis 'flattens' the rMTF. Figure 6.7 shows the results of this analysis applied to SAM responses recorded from the cortices of awake rhesus macaques. Performance of the classifier using the full spike train is indicated along the abscissa in each panel. For modulation frequency, for example Fig. 6.7C, average firing rate information resulted in a significant improvement over chance performance in only a minority of neurons (36%), compared to 94% for the phase-only classifier. Overall, the case for average rate coding of SAM stimuli in the cortex was surprisingly weak. These data suggest that the code employed by the cortex involves spike phases, rather than spike counts – or, more properly, that the cortex counts spikes in windows narrower than a single modulation cycle for most of the tested f_ms.

The foregoing findings cast doubt on the relevance of cortical rMTFs for SAM frequency discrimination, and call into question the primacy of the MTF itself when one considers that cortical responses carried as much information about carrier level as they did about modulation frequency. This suggests that SAM encoding captures more about the stimulus envelope than its periodicity, and, as such, is fundamentally about envelope shape discrimination, rather than modulation frequency extraction. Figure 6.8 shows that the shapes of MPHs can reveal changes in f_m, depth, and level in ways that cannot be captured by VS alone. Malone *et al.* (2007) assayed the fidelity of MPH shapes by computing the 'trial similarity' (TS), which was simply the correlation between MPHs calculated from two separate SAM trials. Unlike VS, which depends on the shape (i.e., the narrowness) of the MPH, TS depends only on the reproducibility of the MPH. TS was shown to be highly predictive of the performance of the spike train classifiers discussed above, as

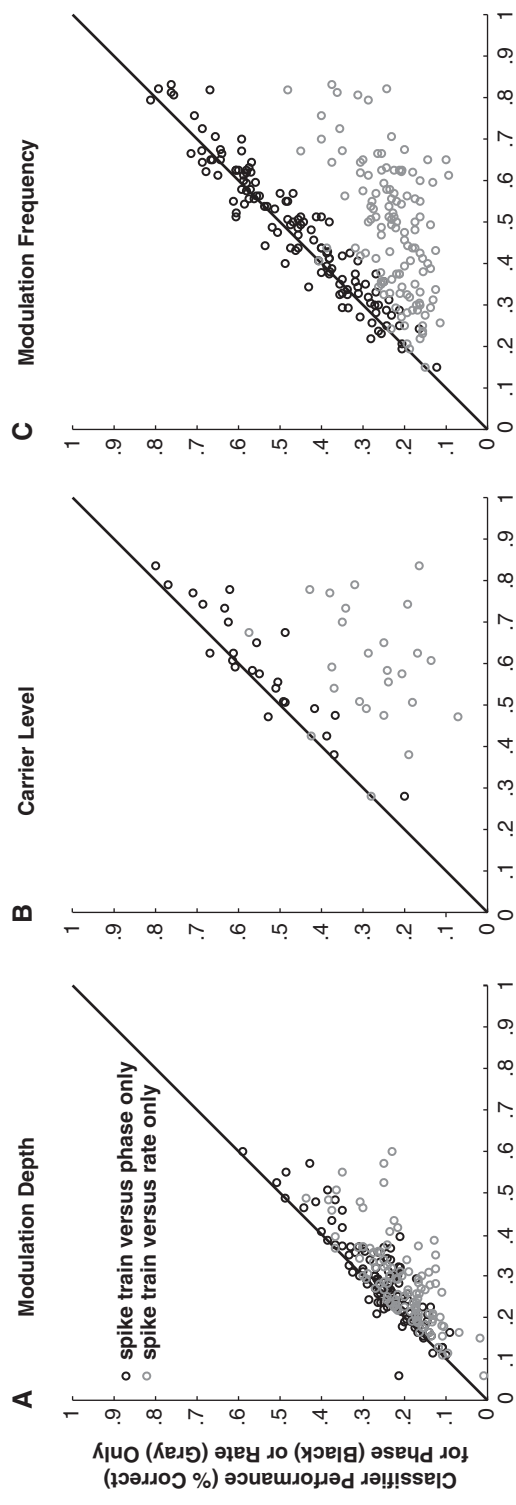

Fig. 6.7 SAM stimulus discrimination based on spike train analysis. The scatterplots compare the performance (in percent correct) of the full spike train classifier, which is plotted on the *abscissa*, to the performances of the classifiers using only information about spike phases (*black circles*) and average spike rate (*gray circles*), plotted on the *ordinate*. In **A**, the SAM stimuli varied in terms of modulation depth; in **B**, carrier level; and in **C**, modulation frequency. The *diagonal line* indicates equivalent performance. These data were obtained in the auditory cortex of awake rhesus monkeys. (With permission of Malone *et al.*, 2007.)

well as being significantly more predictive than VS. The TS curves in Fig. 6.8 suggest that at very low modulation frequencies, what VS fails to capture – 'the full harmonic content of the cycle histogram' – is likely essential to the cortical representation of SAM, and modulated signals more generally.

6.9 Conclusions

What, then, does the foregoing suggest about the transformation of the representation of SAM signals in the ascending auditory pathway? In addition to the growing systemic focus on very low

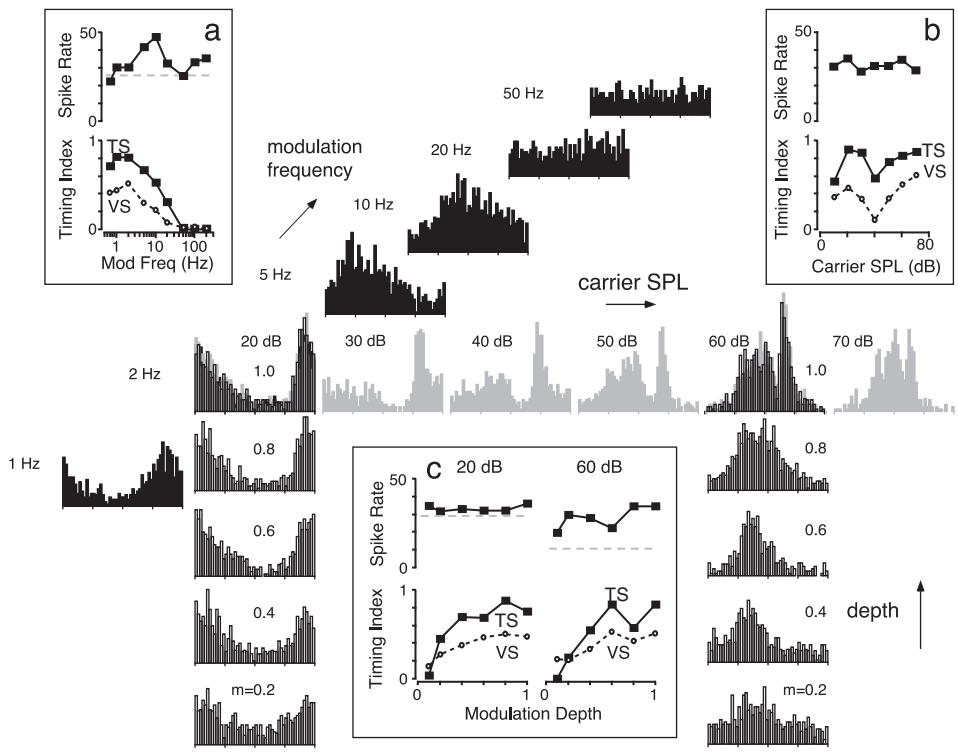

Fig. 6.8 MPH representations of SAM stimuli varying along multiple dimensions for a single neuron. The matrix of response profiles illustrates how changes in modulation frequency, modulation depth, and carrier level produce different but highly reproducible effects on the shape of the MPH. The response profiles corresponding to changes in modulation frequency for a 20-dB carrier level are aligned along the diagonal axis in *black*. The associated rMTFs and tMTFs appear in the *upper* and *lower panels* of *inset* **a**. Response profiles corresponding to the changes in carrier level for a 2-Hz modulation are aligned along the horizontal axis in *gray*. The associated curves depicting the changes in firing rate and timing indices for different carrier levels appear in the *upper* and *lower panels* of *inset* **b**. Response profiles corresponding to changes in modulation depth are shown in the *vertical columns on the left* (2 Hz at 20 dB) and *right* (2 Hz at 60 dB). The associated curves depicting the changes in firing rate and time indices appear in *inset* **c**. Because the data were obtained from separate runs, the 100% modulated, 2-Hz stimulus at 60 dB was repeated twice, and the analogous 20 dB stimulus was repeated thrice (the carrier frequency was constant at 700 Hz). The resulting MPHs are overlaid to illustrate the robustness of the MPH shapes for repeated trials. These data were obtained in the auditory cortex of an awake rhesus monkey. (With permission of Malone *et al*., 2007.)

modulation frequencies, there is also increasing heterogeneity in MPH shapes, which certainly must reflect, at least in part, the increased heterogeneity in level tuning among central auditory neurons. This in turn suggests that rather than abstracting information about modulation frequency, the auditory pathway instead diversifies the way in which amplitudes are encoded. Figure 6.8 demonstrates that the same cortical neuron fires to the instantaneous amplitude maximum, or minimum, of an SAM stimulus depending on the relationship between the carrier level and its preferred level. Consider also that as fewer spikes per modulation cycle are available, the range of MPH shapes becomes constrained; if we assume that there is an underlying distribution reflecting the underlying probability of discharge at each point in the modulation cycle, as the modulation frequency increases, it becomes harder to 'mark' the less preferred phases because of refractoriness relating to having recently fired a spike at the optimal phase. As a result, TS and VS values nearly always converge at a particular f_m and then decline in tandem. In fact, it is possible that the point where these metrics coincide identifies the 'hinge' value of f_m for tMTFs exhibiting a lowpass to bandpass transition with increasing carrier level. From this perspective, the average discharge rate does not encode the envelopes of modulated signals, but instead represents a limit on the resolution at which the features of envelopes can be encoded.

The timing of spikes, and not just the average spike count, clearly is an important aspect of the neural code for processing communication sounds, most of them containing many time-varying components. Neuronal synchronization to the timing of stimulus events and spike timing precision are most useful in performing periodicity analyses and improving the signal-to-noise ratios for signal detection and discrimination tasks based, for example, on information measures (Liu and Schreiner, 2007).

As is evident in our survey of various auditory stations, it is indisputable that the activity distributed across different neuronal populations represents different aspects of information regarding time-varying sounds (although there is also evidence that single neurons respond to stimulus features on multiple timescales (see Elhilali *et al.*, 2004)). These various coding strategies differ in their ability to convey particular features of time-varying sounds, such as their time of occurrence, frequency of occurrence, rate of change, and duration. It is not surprising that the panoply of auditory processing tasks has resulted in such a variety of coding schemes. An important task for future research will be the identification of the tasks that are best served by these different neuronal populations, and definition of their roles in establishing the psychophysical and perceptual capacities of the listener.

Acknowledgement

This work was supported by grants DC02260 (C.E.S.) and MH12993 (B.J.M.) from the NIH, the Coleman Fund, and Hearing Research Incorporated.

References

Batra R (2006) Responses of neurons in the ventral nucleus of the lateral lemniscus to sinusoidally amplitude modulated tones. *J Neurophysiol* **96**(5): 2388–98.

Beitel RE, Schreiner CE, Cheung SW, Wang X, Merzenich MM (2003) Reward-dependent plasticity in the primary auditory cortex of adult monkeys trained to discriminate temporally modulated signals. *Proc Natl Acad Sci USA* **100**(19): 11070–5.

Bieser A, Müller-Preuss P (1996) Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. *Exp Brain Res* **108**: 273–84.

Blackburn CC, Sachs MB (1989) Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. *J Neurophysiol* **62**: 1303–29.

Cooper NP, Robertson D, Yates GK (1993) Cochlear nerve fiber responses to amplitude-modulated stimuli: variations with spontaneous rate and other response characteristics. *J Neurophysiol* **70**(1): 370–86.

Drullman R (1995) Temporal envelope and fine structure cues for speech intelligibility. *J Acoust Soc Am* **97**: 585–92.

Drullman R, Festen JM, Plomp R (1994) Effect of reducing slow temporal modulations on speech reception. *J Acoust Soc Am* **95**: 2670–80.

Elhilali M, Fritz JB, Klein DJ, Simon JZ, Shamma SA (2004) Dynamics of precise spike timing in primary auditory cortex. *J Neurosci* **24**: 1159–72.

Fastl H (1990) The hearing sensation, roughness and neuronal responses to AM-tones. *Hear Res* **46**(3): 293–5.

Fastl H, Hesse A, Schorer E, Urbas J, Müller-Preuss P (1986) Searching for neural correlates of the hearing sensation fluctuation strength in the auditory cortex of squirrel monkeys. *Hear Res* **23**(2): 199–203.

Foffani G, Moxon KA (2004) PSTH-based classification of sensory stimuli using ensembles of single neurons. *J Neurosci Meth* **135**: 93–107.

Frisina RD (2001) Subcortical neural coding mechanisms for auditory temporal processing. *Hear Res* **158**(1–2): 1–27.

Frisina RD, Smith RL, Chamberlain SC (1990) Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement. *Hear Res* **44**: 99–122.

Galazyuk AV, Llano D, Feng AS (2000) Temporal dynamics of acoustic stimuli enhance amplitude tuning of inferior colliculus neurons. *J Neurophysiol* **83**(1): 128–38.

Goldberg J, Brown P (1969) Responses of binaural neurons of dog superior olivary complex to to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J Neurophysiol* **32**: 631–6.

Hartmann WH (1997) *Signals, Sound, and Sensation*. AIP Press, Woodbury.

Hewitt MJ, Meddis R (1994) A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus. *J Acoust Soc Am* **95**(4): 2145–59.

Kuwada S, Batra R (1999) Coding of sound envelopes by inhibitory rebound in neurons of the superior olivary complex in the unanesthetized rabbit. *J Neurosci* **19**(6): 2273–87.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J Neurophysiol* **60**(6): 1799–822.

Liu RC, Schreiner CE (2007) Auditory cortical detection and discrimination correlates with communicative significance. *PLoS Biol* **5**(7): e173.

Joris PX, Yin TC (1992) Responses to amplitude-modulated tones in the auditory nerve of the cat. *J Acoust Soc Am* **91**: 215–32.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. *Physiol Rev* **84**: 541–77.

Joris PX, Louage DH, Cardoen L, van der Heijden M (2006) Correlation index: a new metric to quantify temporal coding. *Hear Res* **216–217**: 19–30.

Kajikawa Y, Hackett TA (2005) Entropy analysis of neuronal spike train synchrony. *J Neurosci Meth* **149**: 90–93.

Krishna BS, Semple MN (2000) Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J Neurophysiol* **84**: 255–73.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J Neurophysiol* **60**: 1799–822.

Liang L, Lu T, Wang X (2002) Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. *J Neurophysiol* **87**: 2237–61.

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**: 1131–8.

Malone BJ, Semple MN (2001) Effects of auditory stimulus context on the representation of frequency in the gerbil inferior colliculus. *J Neurophysiol* **86**: 1113–30.

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**: 4625–38.

Malone BJ, Scott BH, Semple MN (2007) Dynamic amplitude coding in the auditory cortex of awake rhesus macaques. *J Neurophysiol* **98**(3): 1451–74.

Møller AR (1972) Coding of amplitude and frequency modulated sounds in the cochlear nucleus of the rat. *Acta Physiol Scand* **81**: 540–56.

Møller AR (1974) Responses of units in the cochlear nucleus to sinusoidally amplitude-modulated tones. *Exp Neurol* **45**: 104–17.

Møller AR (1976) Dynamic properties of primary auditory fibers compared with cells in the cochlear nucleus. *Acta Physiol Scand* **98**: 157–67.

Phillips DP, Semple MN, Calford MB, Kitzes LM (1994) Level-dependent representation of stimulus frequency in cat primary auditory cortex. *Exp Brain Res* **102**(2): 210–26.

Preuss A, Müller-Preuss P (1990) Processing of amplitude modulated sounds in the medial geniculate body of squirrel monkeys. *Exp Brain Res* **79**(1): 207–11.

Rees A, Møller AR (1987) Stimulus properties influencing the responses of inferior colliculus neurons to amplitude-modulated sounds. *Hear Res* **27**: 129–43.

Rees A, Palmer AR (1989) Neuronal responses to amplitude-modulated and pure-tone stimuli in the guinea pig inferior colliculus, and their modification by broadband noise. *J Acoust Soc Am* **85**: 1978–94.

Rhode WS (1994) Temporal coding of 200% amplitude modulated signals in the ventral cochlear nucleus of the cat. *Hear Res* **77**: 43–68.

Rhode WS, Greenberg S (1992) Physiology of the cochlear nuclei. In: The Mammalian Auditory Pathway: Neurophysiology, Vol 2 (eds Popper AN, Fay RR), pp 94–152. Springer, New York.

Rhode WS, Greenberg S (1994) Encoding of amplitude modulation in the cochlear nucleus of the cat. *J Neurophysiol* **71**(5): 1797–825.

Sanes DH, Malone BJ, Semple MN (1998) Role of synaptic inhibition in processing of dynamic binaural level stimuli. *J Neurosci* **18**: 794–803.

Schreiner, CE, Langner G (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J Neurophysiol* **60**(6): 1823–40.

Schreiner CE, Urbas JV (1988) Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear Res* **32**: 49–63.

Semple MN, Kitzes LM (1993a) Binaural processing of sound pressure level in cat primary auditory cortex: evidence for a representation based on absolute levels rather than interaural level differences. *J Neurophysiol* **69**: 449–61.

Semple MN, Kitzes LM (1993b) Focal selectivity for binaural sound pressure level in cat primary auditory cortex: two-way intensity network tuning. *J Neurophysiol* **69**: 462–73.

Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am* **114**: 3394–411.

Smith RL, Brachman ML (1980) Response modulation of auditory-nerve fibers by AM stimuli: effects of average intensity. *Hear Res* **2**(2): 123–33.

Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. *Nature* **416**: 87–90.

Swarbrick L, Whitfield IC (1972) Auditory cortical units selectively responsive to stimulus 'shape'. *J Physiol* **224**: 68P–69P.

Ter-Mikaelian M, Sanes DH, Semple MN (2007) Transformation of temporal properties between auditory midbrain and cortex in the awake Mongolian gerbil. *J Neurosci* **27**(23): 6091–102.

Ulanovsky N, Las L, Farkas D, Nelken I (2004) Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci* **24**: 10440–53.

Wang X (2007) Neural coding strategies in auditory cortex. *Hear Res* **229**(1–2): 81–93.

Wohlgemuth S, Ronacher B (2007) Auditory discrimination of amplitude modulations based on metric distances of spike trains. *J Neurophysiol* **97**: 3082–92.

Yates GK (1987) Dynamic effects in the input/output relationship of auditory nerve. *Hear Res* **27**: 221–30.