ELSEVIER

Speech Communication 41 (2003) 107–121

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

# Cortical processing of temporal modulations

Xiaoqin Wang [a,*], Thomas Lu [a], Li Liang [a,b]

[a] *Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, 720 Rutland Avenue, Ross 424, Baltimore, MD 21205, USA*

[b] *Hearing Center, Pear River Hospital of First Medical University, Guangzhou 510282, Guangdong Province, China*

**Abstract**

Temporal modulations are fundamental components of human speech and animal communication sounds. Understanding their representations in the auditory cortex is a crucial step towards our understanding of brain mechanisms underlying speech processing. While modulated signals have long been used as experimental stimuli, their cortical representations are not completely understood, particularly for rapid modulations. Known physiological data do not adequately explain psychophysical observations on the perception of rapid modulations, largely due to slow stimulus-synchronized temporal discharge patterns of cortical neurons. In this article, we summarize recent findings from our laboratory on temporal processing mechanisms in the auditory cortex. These findings show that the auditory cortex represents slow modulations *explicitly* using a temporal code and fast modulations *implicitly* by a discharge rate code. Rapidly modulated signals within a short-time window (~20–30 ms) are integrated and transformed into a discharge rate-based representation. The findings also indicate that there is a shared representation of temporal modulations by cortical neurons that encodes the temporal profile embedded in complex sounds of various spectral contents. Our results suggest that cortical processing of sound streams operates on a "segment-by-segment" basis with a temporal integration window on the order of ~20–30 ms.


*Keywords:* Auditory cortex; Temporal processing; Temporal integration; Amplitude modulation; Frequency modulation; Temporal asymmetry; Species-specific vocalization

## 1. Introduction

The neural representation of temporal modulations in the auditory cortex is of special interest to our understanding of mechanisms underlying speech processing. Temporal modulations are fundamental components of communication sounds such as human speech and animal vocalizations, as well as musical sounds. Low-frequency modulations are important for speech perception and melody recognition, while higher-frequency modulations produce other types of sensations such as pitch and roughness (Houtgast and Steeneken, 1973; Rosen, 1992). Both humans and animals are capable of perceiving the information contained in temporally modulated sounds across a wide range of time scales from millisecond to tens and hundreds of milliseconds. How the auditory cortex encodes this wide dynamic range of temporal modulations is not well understood.

The neural representation of temporal modulations begins at the auditory periphery where

* Corresponding author. Tel.: +1-410-614-4547; fax: +1-410-614-9599.

*E-mail address:* xwang@bme.jhu.edu (X. Wang).

0167-6393/02/$ - see front matter 
doi:10.1016/S0167-6393(02)00097-3

auditory-nerve fibers faithfully represent fine details of complex sounds in their temporal discharge patterns (Johnson, 1980; Joris and Yin, 1992; Palmer, 1982; Wang and Sachs, 1993). At subsequent nuclei along the ascending auditory pathway (CN—cochlear nucleus, IC—inferior colliculus and MGB—auditory thalamus), the precision of this temporal representation gradually degrades (e.g., CN: Blackburn and Sachs, 1989; Frisina et al., 1990; Wang and Sachs, 1994; IC: Langner and Schreiner, 1988; MGB: Creutzfeldt et al., 1980; de Ribaupierre et al., 1980) due to biophysical properties of neurons and temporal integration of converging inputs from one station to the next. In a modeling study of the transformation of temporal discharge patterns from the auditory nerve to the cochlear nucleus, Wang and Sachs (1995) showed that the reduction of phase-locking in stellate cells can result from three mechanisms: convergence of subthreshold inputs on the soma, inhibition, and the well-known dendritic low-pass filtering (Rall and Agmon-Snir, 1998). These basic mechanisms may also operate at successive nuclei leading to the auditory cortex, progressively reducing the temporal limit of stimulus-synchronized responses.

It has long been noticed that neurons in the auditory cortex do not faithfully follow rapidly changing stimulus components (de Ribaupierre et al., 1972; Goldstein et al., 1959; Whitfield and Evans, 1965). A number of previous studies have shown that cortical neurons can only be synchronized to temporal modulations at a rate far less than 100 Hz (Bieser and Müller-Preuss, 1996; de Ribaupierre et al., 1972; Eggermont, 1991, 1994; Gaese and Ostwald, 1995; Lu and Wang, 2000; Schreiner and Urbas, 1988), compared with a limit of ~1 kHz at the auditory-nerve (Joris and Yin, 1992; Palmer, 1982). The lack of synchronized cortical responses to rapid, but perceivable temporal modulation has been puzzling. Because most of the previous studies in the past three decades on this subject were conducted in anesthetized animals, with a few exceptions (Bieser and Müller-Preuss, 1996; Creutzfeldt et al., 1980; de Ribaupierre et al., 1972; Evans and Whitfield, 1964; Goldstein et al., 1959; Whitfield and Evans, 1965), it has been speculated that the reported low temporal response rate in the auditory cortex might be caused partially by anesthetics, which have been shown to alter the temporal response properties of the auditory cortex (Goldstein et al., 1959; Zurita et al., 1994). Neural responses obtained under unanesthetized conditions are therefore of particular importance to our understanding of cortical representations of temporally modulated signals. This article summarizes recent findings from our studies in awake primates on the issues related to cortical representations of temporal modulations.

## 2. The limit on stimulus-synchronized discharges in the auditory cortex and a two-stage mechanism

Goldstein et al. (1959) showed that click-following rates of cortical evoked potentials were higher in unanesthetized cats than in anesthetized ones. In our study, we systematically investigated responses of single neurons in the primary auditory cortex (A1) of awake marmoset monkeys to rapid sequences of clicks (Lu et al., 2001b). Both wide- and narrow-band click trains with inter-click intervals (ICIs) ranging from 3 to 100 ms were studied. Narrow-band clicks were centered at each neuron's characteristic frequency (CF). In contrast to neurons studied in A1 of anesthetized animals, which responded strongly to both wide- and narrow-band clicks (Lu and Wang, 2000), the majority of neurons examined in A1 of unanesthetized animals responded strongly to narrow-band clicks, but were only weakly driven or, more often, unresponsive to wide-band (rectangular) clicks (Lu et al., 2001b). One type of response to click trains is illustrated in Fig. 1(A). This neuron exhibited significant stimulus-synchronized responses to click stimuli at long ICIs (>25–30 ms). The discharges to click trains became non-synchronized at medium ICIs and diminished at short ICIs (<20–25 ms), apparently due to inhibition (Fig. 1(A)). The second type of neural response did not exhibit stimulus-synchronized discharges. They responded, however, to changes in ICI with monotonically changing discharge rate when the ICI was shorter than ~20–30 ms, as illustrated by the example in Fig. 1(B). We have shown that the limit on stimulus-synchronized responses is on the order of

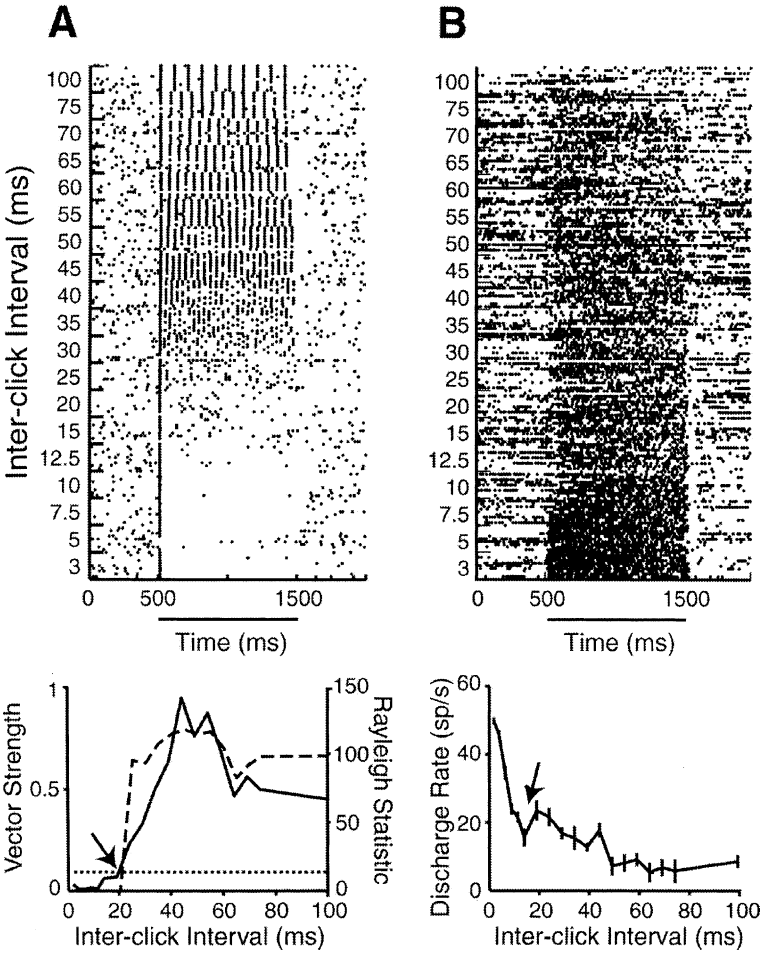


Fig. 1. Examples of stimulus-synchronized and non-synchronized responses to click trains recorded from single neurons in the primary auditory cortex (A1) of awake marmosets (Lu et al., 2001b). Click trains at various ICIs were delivered for 10 repetitions in randomized blocks. (A) (top): Responses of a representative neuron with stimulus-synchronized discharges to click train stimuli are shown in the form of dot raster plot and organized by ICI along the ordinate. ICIs ranged from 3 to 100 ms (3, 5, 7.5, 10, 12.5, 15, 20, 30, ..., 70, 75, 100 ms). Each dot indicates the occurrence of an action potential (spike). Stimulus onset was at 500 ms, and duration was 1000 ms as indicated by the horizontal bar below the time-axis. (bottom): Vector strengths, VS (dashed line) and Rayleigh statistic (solid line) of the responses to the click-trains are shown. The dotted line at the Rayleigh statistics of 13.8 indicates the threshold for statistically significant stimulus-synchronized activity ($p < 0.001$). Insignificant VS were set to zero. A synchronization boundary was calculated and is indicated by an arrow. (B) (top): Dot raster plot showing responses of a representative neuron with non-synchronized discharges to click train stimuli. (bottom): Driven discharge rate is plotted versus ICI. Spontaneous discharge rate (estimated with a time window of 0–500 ms) was subtracted from the total discharge rate. Vertical bars represent standard errors of the means (SEM). The arrow indicates calculated rate-response boundary.

20–25 ms (median value of sampled population) in A1 in the unanesthetized condition (Lu et al., 2001b). The observation that neurons are sensitive to changes of short ICIs indicates that a discharge rate-based mechanism may be in operation when ICIs are shorter than ~20–30 ms.

We have identified two populations of A1 neurons that displayed these two response types, referred to as *synchronized* and *non-synchronized* populations, respectively (Fig. 2). The two populations appeared to encode sequential stimuli in very different manners. Neurons in the

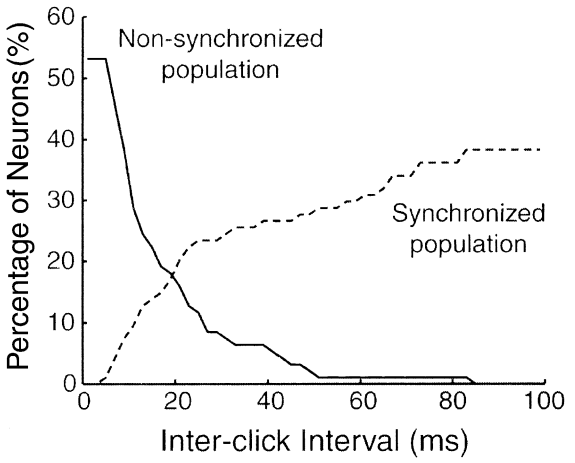


Fig. 2. A combination of temporal and rate representations can encode a wide range of ICIs. Curves are the cumulative sum of the histograms representing response boundaries of two neural populations, one with stimulus-synchronized discharges ($N = 36$) and the other with non-synchronized discharges ($N = 50$), respectively. The dashed line shows the percentage of neurons with synchronization boundaries less than or equal to a given ICI. The solid line shows the percentage of neurons with rate-response boundaries greater than or equal to a given ICI. The total number of neurons analyzed is 94, including eight neurons (not shown) with combined response characteristics (Lu et al., 2001b).

synchronized population showed stimulus-synchronized discharges at long ICIs, but few responses at short ICIs. This population of neurons can thus represent slowly occurring temporal events *explicitly* using a temporal code. The non-synchronized population of neurons did not exhibit stimulus-synchronized discharges at either long or short ICIs. This population of neurons can *implicitly* represent rapidly changing temporal intervals by their average discharge rates. On the basis of these two populations of neurons, the auditory cortex can represent a wide range of time intervals in sequential, repetitive stimuli (Fig. 2). The large number of neurons with non-synchronized and sustained discharges at short ICIs observed in A1 of awake animals (Lu et al., 2001b) was not observed in A1 of anesthetized animals (Lu and Wang, 2000).

The issues of rate and temporal coding have long been studied at the auditory periphery and brainstem regarding speech sounds (Sachs et al., 1992). Both temporal-place and rate-place representations appear to be adequate to represent a vowel's spectrum by the auditory-nerve fibers (Sachs and Young, 1979; Young and Sachs, 1979). At the cochlear nucleus, different types of neurons begin to show specificity in utilizing rate and temporal information in representing a vowel's spectrum, with the temporal-place code preserved by bushy cells but degraded by chopper cells (Blackburn and Sachs, 1990). As demonstrated by findings from our studies, rate and temporal information are further segregated among populations of neurons in the auditory cortex, a processing stage that is several synapses away from the cochlear nucleus.

## 3. Temporal integration window of the primary auditory cortex

The limited stimulus-synchronized responses observed in neural responses to click trains (Figs. 1 and 2) suggest that cortical neurons integrate sequential or continuous stimuli over a brief time window that we define operationally as the *temporal integration window*. Two sequential acoustic events falling within the temporal integration window are not distinguished as separate events at the output of a neuron. An implication of the temporal integration window is that it should result in a maximum response of a neuron when the integration of sequential or continuing stimulus events is performed over the duration of this window. We have further investigated these notions through a series of experiments in which cortical neurons were tested with a variety of temporal modulations (Liang et al., 1999, 2002). In these experiments, temporal modulations were introduced by sinusoidally modulating tone or noise carriers in amplitude or frequency. For amplitude-modulated tones (sAM), the carrier frequency remained constant while the amplitude was sinusoidally modulated. In the case of amplitude-modulated noises (nAM), the amplitude of a noise carrier (narrow-band or broad-band) was modulated by a sinusoid. For frequency-modulated tones (sFM), the amplitude remained constant

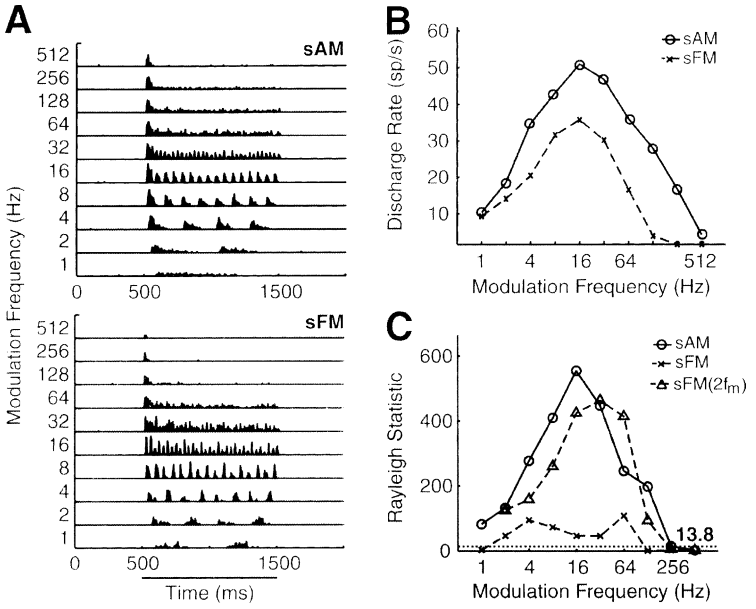


Fig. 3. Cortical responses to sinusoidally amplitude-modulated (sAM) and frequency-modulated (sFM) tones in a representative neuron (Liang et al., 2002). (A) Post-stimulus histograms (PSTHs) showing responses to sAM (upper) and sFM (lower) stimuli. The displays are arranged according to modulation frequency along the ordinate. Stimulus duration was 1000 ms (onset at 500 ms), as indicated by a horizontal bar below the time axis. Stimulus parameters were as follows: $f_c = 16.07$ kHz, sound level = 80 dB SPL, $d_{sAM} = 100\%$, $d_{sFM} = 1024$ Hz ($f_c$: carrier frequency, $d$: modulation depth). $f_c$ was set equal to the neuron's CF. (B) Discharge rMTF derived from the data shown in (A). Average discharge rates were calculated over a period including the stimulus duration and 100 ms after stimulus offset, with spontaneous discharge rates subtracted (Liang et al., 2002). rMTFs produced by sAM (solid line with circles) and sFM (dashed line with crosses) stimuli are shown. (C) Temporal modulation transfer functions (tMTF), derived from the data shown in (A), plotted in the form of Rayleigh statistics (equal to $2nVS^2$, where $n$ is the total number of spikes and VS is the vector strength). The horizontal dotted line at 13.8 indicates the threshold for statistically significant stimulus-synchronized responses ($p < 0.001$). tMTFs due to sAM (solid line with circles) and sFM (dashed line with crosses) stimuli were calculated at $f_m$. An additional tMTF (dashed line with triangles) is calculated at the frequency equal to twice the modulation frequency ($2f_m$) for sFM responses.

while the carrier frequency was sinusoidally modulated.

Fig. 3 shows responses of a representative auditory cortical neuron to sAM and sFM stimuli at various modulating frequencies. At a modulation frequency of 16 Hz, the discharge rate reached the maximum in this neuron for both sAM and sFM stimuli (Fig. 3(B)). This modulation frequency is conventionally referred to as the discharge rate-based best modulation frequency (rBMF). Data shown in Fig. 3(B) are generally referred to as the discharge rate-based modulation transfer function (rMTF). The majority of neurons in A1 of awake marmosets responded maximally to modulated tones at a particular modulation frequency with sustained firings. In general, rMTFs derived from responses of a neuron to sAM and sFM stimuli had similar shapes and closely matched rBMF (Fig. 3(B)). This similarity was also observed between cortical responses to amplitude-modulated tone or noise, as shown by an example neuron in Fig. 4(A) and (B). This neuron discharged maximally at modulation frequency of 64 Hz, regardless of whether the temporal modulation was introduced in a tone or noise carrier (Fig. 4(C)). Because amplitude and frequency modulations are produced along different stimulus dimensions, the match of rBMFs in various stimulus conditions suggests an inherent temporal selectivity in cortical neurons that is applicable to a wide range of time-varying stimuli. The fact that auditory cortical neurons showed similar rMTF and rBMF (Fig. 4(C)) in response to amplitude-modulated tones and broad-band noises indicates that the observed

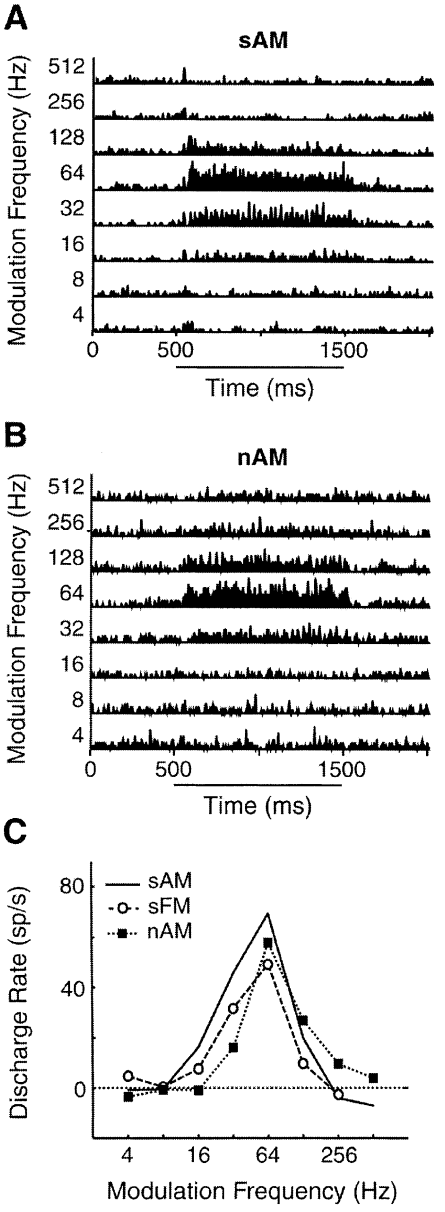


Fig. 4. Cortical responses to sAM, sFM and amplitude-modulated broad-band noise (nAM) stimuli in a representative neuron. (A) PSTHs of responses to sAM stimuli (onset at 500 ms, duration of 1000 ms). The display is arranged according to modulation frequency along the ordinate. (B) PSTHs of responses to nAM stimuli. (C) rMTFs measured for sAM (solid line), sFM (dashed line with open circles) and nAM (dotted line with filled squares). The stimulus parameters were: $f_c = 4.7$ kHz (equal to the neuron's CF), sound level = 70 dB SPL, $d_{sAM} = 75\%$, $d_{sFM} = 512$ Hz; $d_{nAM} = 75\%$ ($d$: modulation depth).

modulation selectivity was indeed a temporal instead of a spectral phenomenon. Together, these data show that neurons in the auditory cortex of awake marmosets have a preferred *temporal modulation frequency* that is relatively independent of how the temporal modulation is introduced (in the time or frequency domain, by tone or noise carrier). The most frequently encountered rBMFs in A1 of awake marmosets ranged from 8 to 64 Hz. For the population of neurons we studied, the distributions of rBMFs were centered near 30 Hz for both sAM and sFM stimuli (Fig. 5(A)). In fact, A1 is maximally excited near this particular temporal modulation frequency (Fig. 5(B)). This suggests that the optimal temporal integration window for A1 as a whole is in the order of ~30 ms.

The example given in Fig. 3 also shows that discharges of cortical neurons, in response to sAM and sFM stimuli, could exhibit stimulus-synchronized temporal patterns (Fig. 3(A)). Discharge patterns that are synchronized to the modulation waveform of an sAM or sFM stimulus can be quantified by the vector strength (Goldberg and Brown, 1969) and the Rayleigh statistic (Mardia and Jupp, 2000). The values of the Rayleigh statistic greater than 13.8 are considered as statistically significant ($p < 0.001$) (Mardia and Jupp, 2000). Fig. 3(C) plots Rayleigh statistics as a function of modulation frequency, also referred to as the discharge synchrony-based modulation transfer function or temporal modulation transfer function (tMTF). The modulation frequency corresponding to the maximum of a tMTF is conventionally referred to as the discharge synchrony-based best modulation frequency (tBMF). This neuron responded to sAM stimuli with well-synchronized discharges at modulation frequencies up to 128 Hz (Fig. 3(C)). The temporal discharge patterns in response to sFM stimuli (Fig. 3(A), lower) in the same neuron, however, differed markedly from those to sAM stimuli (Fig. 3(A), upper) in that there were two clusters of firings within each modulation period. This was because both the upward and downward trajectory of the modulation waveform excited this neuron. As a result, the discharge synchrony was much stronger

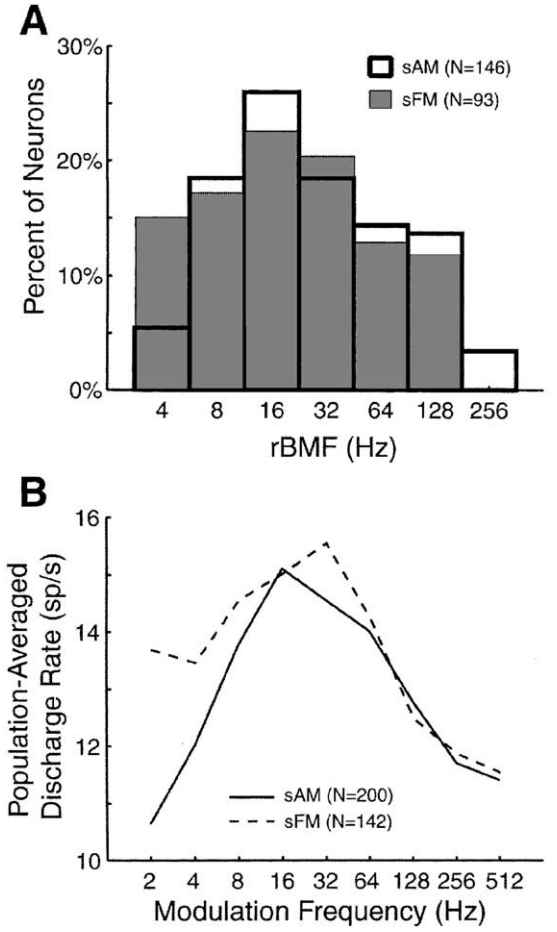


Fig. 5. Population properties for discharge rate-based modulation frequency selectivity (Liang et al., 2002). (A) Overlapping histograms showing distributions of rBMF derived from sAM (open) and sFM (shaded) stimuli, respectively. The neurons included in the histogram exhibited band-pass tMTF. The bin-widths of the histograms are on a base-2 logarithmic scale. The distributions of $rBMF_{sAM}$ and $rBMF_{sFM}$ are not statistically different from each other (Wilcoxon rank sum test, $p = 0.1$). The means of the two distributions are 25.9 Hz ($rBMF_{sAM}$) and 19.2 Hz ($rBMF_{sFM}$) on the base-2 logarithmic scale and 48.8 Hz ($rBMF_{sAM}$) and 35.1 Hz ($rBMF_{sFM}$) on the linear scale, respectively. The medians of the two distributions are 22.6 Hz ($rBMF_{sAM}$) and 18.1 Hz ($rBMF_{sFM}$), respectively. (B) Discharge rates averaged over the entire population of sampled neurons are plotted versus modulation frequency for sAM (solid line) and sFM (dashed line) stimuli. Spontaneous rates were not subtracted in the calculations.

when measured at *twice* the modulation frequency than at the modulation frequency of the stimulus (Fig. 3(C)). In general, discharges could be synchronized at a rate approximately equal to the modulation frequency when sAM stimuli were used, whereas for sFM stimuli, response synchronization could occur at a rate twice as large as the modulation frequency. Moreover, stimulus-induced synchronization was sometimes produced by one type of modulated sound, but not by another type in an individual neuron. Temporal modulation in a stimulus is, therefore, not always accurately reflected in the discharge synchrony of cortical neurons.

Across populations of A1 neurons, tBMF is lower than rBMF (comparing Fig. 6(A) with Fig. 5(A)). The distributions of tBMF, like those of rBMF, are similar for sAM and sFM responses. This observation again supports the proposition that auditory cortical neurons have a preferred temporal modulation frequency, regardless of how the modulation is introduced (in the time or frequency domain). Moreover, the selectivity for a particular temporal modulation frequency was often observed in average discharge rate in the absence of stimulus-synchronized discharges, indicating a temporal-to-rate transformation in the cortical coding of temporal modulations (Liang et al., 2002).

Another useful measure of discharge synchrony in a neuron is the maximum synchronization frequency ($f_{max}$), which is defined as the highest modulation frequency at which significant discharge synchrony exists. The distribution of $f_{max}$ is centered between 32 and 64 Hz (Fig. 6(B)), which is consistent with the distribution of the synchronization boundary of cortical neurons determined by click train stimuli (Fig. 2). Fig. 2 shows that the percentage of neurons with the synchronization boundary less than ~25 ms (equivalent to >40 Hz modulation frequency) drops rapidly.

In Fig. 6(C), we plot the percent of sampled neurons that exhibited statistically significant discharge synchrony as a function of modulation frequency. This figure shows that A1 is maximally synchronized to the temporal modulation at a modulation frequency of ~8 Hz, which is lower than the modulation frequency (~30 Hz) that elicits the maximum discharge rate (Fig. 5(B)).

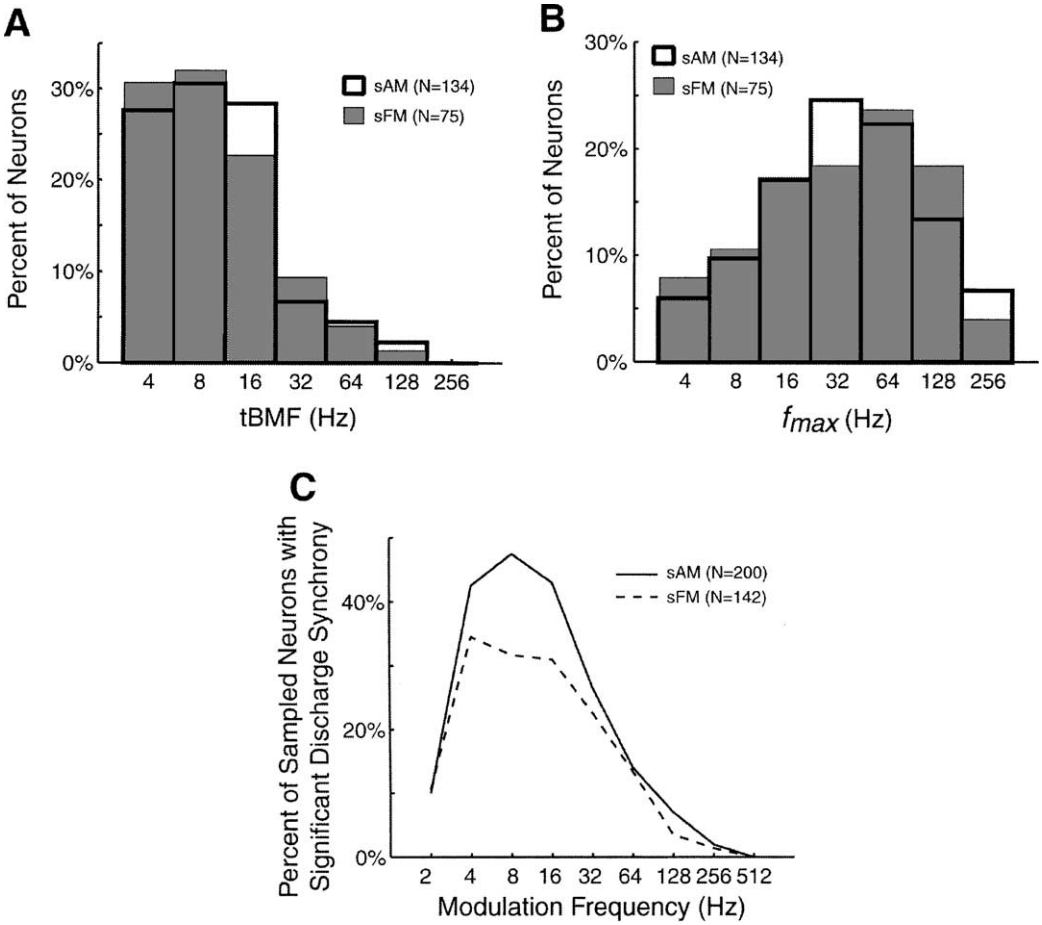


Fig. 6. Population properties for discharge synchrony-based modulation frequency selectivity (Liang et al., 2002). (A) Overlapping histograms showing distributions of tBMF derived from sAM (open) and sFM (shaded) stimuli, respectively. The bin-widths of the histograms are on a base-2 logarithmic scale. The distributions of $tBMF_{sAM}$ and $tBMF_{sFM}$ are not statistically different from each other (Wilcoxon rank sum test, $p = 0.82$). The means of the two distributions are 9.7 Hz ($tBMF_{sAM}$) and 9.2 Hz ($tBMF_{sFM}$) on the base-2 logarithmic scale and 15.6 Hz ($tBMF_{sAM}$) and 14.2 Hz ($tBMF_{sFM}$) on the linear scale, respectively. The medians of the two distributions are 9.6 Hz ($tBMF_{sAM}$) and 10.0 Hz ($tBMF_{sFM}$), respectively. (B) Overlapping histograms showing distributions of maximum synchronization frequency ($f_{max}$) derived from sAM (open) and sFM (shaded) stimuli, respectively. The bin-widths of the histograms are on a base-2 logarithmic scale. The distributions of $f_{max}$(sAM) and $f_{max}$(sFM) are not statistically different from each other (Wilcoxon rank sum test, $p = 0.97$). The means of the two distributions are 34.2 Hz (sAM) and 32.9 Hz (sFM) on the base-2 logarithmic scale and 58.9 Hz (sAM) and 57.4 Hz (sFM) on the linear scale, respectively. The medians of the two distributions are 34.2 Hz (sAM) and 39.4 Hz (sFM), respectively. (C) Percentage of neurons with statistically significant Rayleigh statistic (>13.8) over the entire population of sampled neurons is plotted versus modulation frequency for sAM (solid line) and sFM (dashed line) stimuli.

These observations further support the two-stage mechanism proposed on the basis of cortical responses to click trains (Lu et al., 2001b).

Do the response properties of cortical neurons revealed by temporally modulated signals bear any implications for the processing of complex sounds such as speech or species-specific vocalizations? As we know, speech and musical sounds contain prominent modulations in both amplitude and frequency domains. In particular, low-frequency (<30 Hz) modulations are important for melody recognition which, as suggested by our data, may be encoded on the basis of temporal discharge patterns of cortical neurons. One possible psycho-

physical correlate of the maximum synchronization frequency is the lower limit of pitch, which is defined as the lowest repetition rate that evokes a sensation of pitch and has been found near 30 Hz (Krumbholz et al., 2000; Pressnitzer et al., 2001). The temporal integration window of cortical neurons on the order of 20–30 ms suggests that spectro-temporal transients, such as formant transitions, may be integrated in the auditory cortex and implicitly coded by discharge rates. For modulation frequencies below ~30 Hz, modulation periods can be resolved temporally because there are significant synchronized discharges to the modulation periods (Liang et al., 2002). For higher modulation frequencies that correspond psychophysically to the sensation of roughness (Zwicker and Fastl, 1999), they are likely represented by discharge rate instead of temporal discharge patterns. The rBMF can be as high as 256 Hz in unanesthetized auditory cortex (Bieser and Müller-Preuss, 1996; Liang et al., 2002). Our recent study using click train stimuli showed rate-coding for even higher repetition frequencies (Lu et al., 2001b). The primate species studied, the common marmoset, produces two types of vocalizations (*trill* and *trillphee*) that display prominent sinusoidal frequency modulations (Wang, 2000). The distributions of the modulation frequency in both types of vocalizations are centered close to 30 Hz (Agamaite and Wang, 1997), near the temporal modulation frequency preferred by most cortical neurons (Liang et al., 2002). Temporal modulation frequency in these vocalizations differed among individual marmosets, suggesting that this parameter may be used in caller identification. Behaviorally, trill and trillphee calls are considered contact calls that marmosets use during intra-species vocal exchanges. Having the calls' modulation frequencies close to the temporal modulation frequency that maximally excites the auditory cortex should facilitate cortical processing of these vocalizations.

## 4. Cortical responses to stimulus transients within the temporal integration window

The experiments discussed above suggest that A1 neurons integrate stimulus components within a time window of ~30 ms and treat components outside this window as discrete acoustic events. Humans and animals are known to discriminate changes in acoustic signals at time scales shorter than the temporal integration window of A1 neurons, suggesting that cortical neurons must be able to signal such rapid changes. We have investigated sensitivity of cortical neurons to rapid changes within the putative temporal integration window (Lu et al., 2001a) using a class of temporally modulated signals termed *ramped and damped sinusoids* (Patterson, 1994a,b) (Fig. 7(A)). A damped sinusoid consists of a pure tone amplitude-modulated by an exponential function. It has a fast onset followed by a slow offset. The rate of amplitude decay is determined by the exponential half-life. A ramped sinusoid is a time-reversed damped sinusoid. Both types of sounds have identical long-term Fourier spectra. Our experimental stimuli consisted of ramped or damped sinusoid segments, typically with a period of 25 ms, repeated consecutively for 500 ms. For each neuron, the carrier frequency was set to a neuron's CF and the half-life was varied from 0.5 to 32 ms. Most cortical neurons that we studied showed a clear preference for either ramped or damped sinusoids (i.e., they responded more vigorously to a single stimulus type), with a greater portion of neurons preferring ramped stimuli (Lu et al., 2001a). Some neurons responded nearly exclusively to one stimulus type. A representative example of a neuron preferring ramped sinusoids to damped sinusoids is shown in Fig. 7(B). This neuron responded more strongly to ramped sinusoids at most half-lives (Fig. 7(B), a and b). The response asymmetry was observed in average discharge rate, but not in stimulus-synchronized discharges (Fig. 7(B), c). Fig. 7(C) shows a neuron that responded preferentially to damped sinusoids. Generally, preference for stimulus type was consistent across half-lives as illustrated by these examples. These observations demonstrated that temporal characteristics within the temporal integration window can profoundly modulate a cortical neuron's responsiveness.

We found that neurons in both synchronized and non-synchronized populations (Fig. 2) were sensitive to temporal asymmetry within a brief

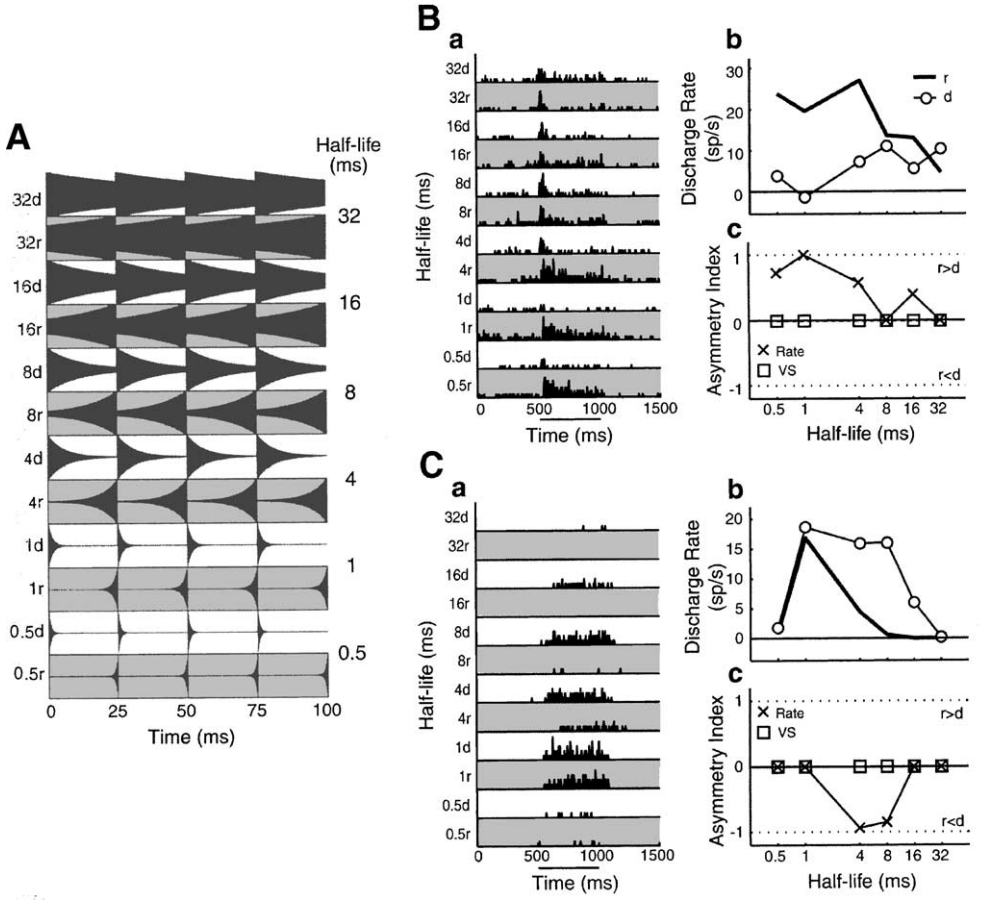


Fig. 7. Cortical responses to ramped and damped sinusoidal stimuli (Lu et al., 2001a). (A) Ramped and damped sinusoidal stimuli (Patterson, 1994a) with half-life ranging from 0.5 to 32 ms. The examples shown have a period of 25 ms and carrier frequency of 10 kHz. Only the first 100 ms of the stimuli are shown. Ramped sinusoids are shown in the shaded blocks. (B) A representative example of neurons with a preference for *ramped* sinusoids recorded in marmoset A1 (Lu et al., 2001a). (a) PSTHs are plotted in the order of increasing half-life along the ordinate, with alternating responses to ramped (shaded) and damped (unshaded) sinusoids. The heights of PSTHs are normalized to the maximum bin count over all stimulus conditions. Stimulus onset was at 500 ms, and duration was 500 ms. (b) Driven discharge rates are plotted as a function of half-life for ramped (thick line) and damped (thin line with open circles) sinusoids, respectively. (c) Asymmetry index is shown for calculations based on discharge rate (cross) or vector strength, VS (square) and is defined as $(R_r - R_d)/(R_r + R_d)$, where $R_r$ and $R_d$ are average discharge rates or VS to ramped and damped stimuli, respectively. Non-significant index values were set to zero ($p \geqslant 0.05$, Wilcoxon rank-sum). Significant asymmetry preference based on discharge rate was present over most half-lives tested in this neuron. No statistically significant asymmetry index was found for stimulus-synchronized activity in this neuron. (C) A representative example of neurons with preference for *damped* sinusoids. The format is the same as in (B).

time window (~25 ms) as measured by their average discharge rates. Fig. 8 shows responses of two representative A1 neurons to sequences of ramped and damped sinusoids at different interstimulus intervals or repetition periods (3–100 ms), one with stimulus-synchronized discharges (Fig. 8(A)) and the other with non-synchronized discharges (Fig. 8(B)). At repetition periods longer than the presumed temporal integration window (>20–25 ms), discharges of the neuron shown in Fig. 8(A) were synchronized to each period of ramped and damped sinusoids while the neuron showed stronger responses to damped sinusoids (Fig. 8(A), top). When repetition periods were shorter than the presumed temporal integration window, synchronized discharges disappeared but

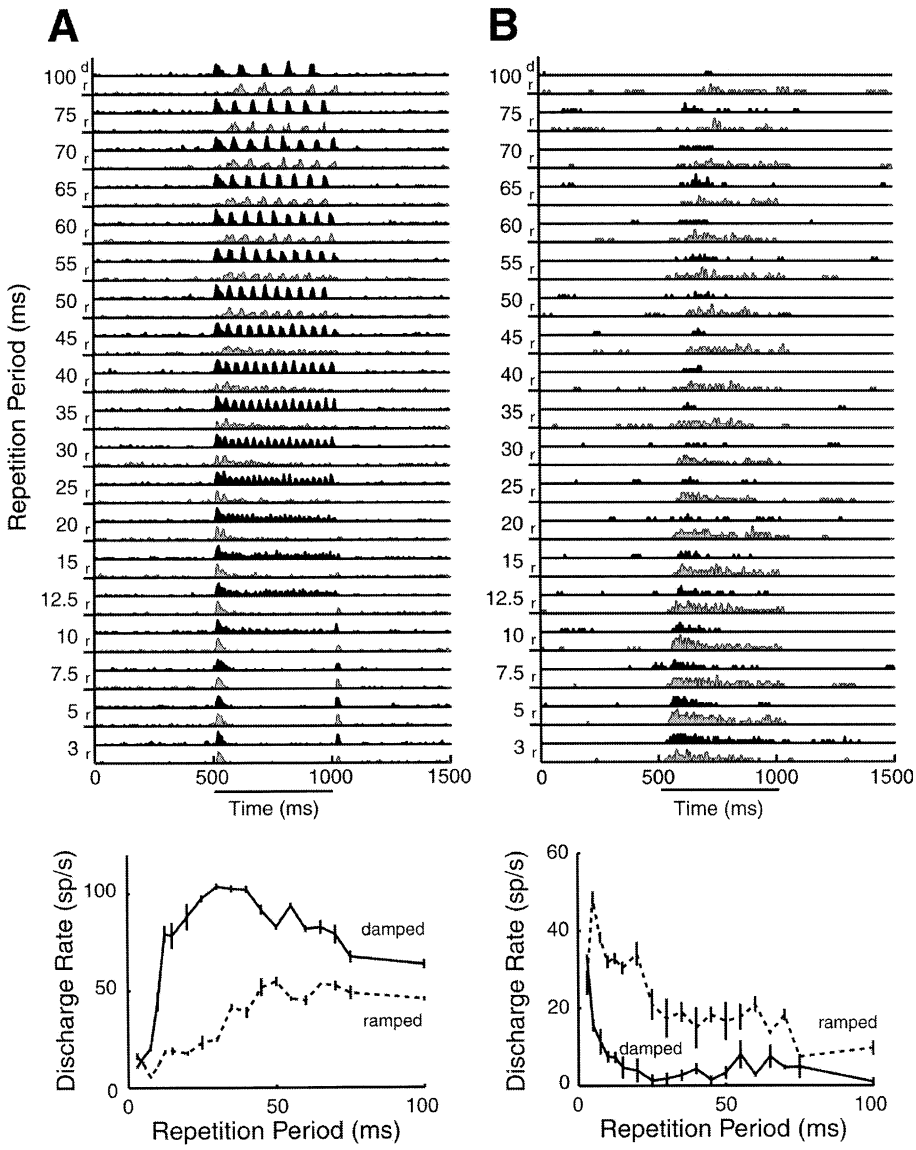


Fig. 8. Examples showing that the selectivity of A1 neurons for temporal asymmetry is unchanged at different inter-stimulus intervals (Lu et al., 2001b). The stimuli used were ramped or damped sinusoids with different repetition periods (3–100 ms) and a fixed half-life. (A) Response of a neuron with stimulus-synchronized discharges. This neuron responded more strongly to damped sinusoids across different repetition periods. (top): PSTHs are plotted in the order of increasing repetition period along the ordinate, with alternating responses to ramped (gray) and damped (black) sinusoids. The heights of PSTHs are normalized to the maximum bin count over all stimulus conditions. Stimulus onset was at 500 ms, and duration was 500 ms. (bottom): Average discharge rates for ramped (dashed line) and ramped (solid line) stimuli are plotted as functions of stimulus repetition period. Vertical bars represent standard errors of the means (SEM). (B) Example of a non-synchronized neuron that responded more strongly to ramped sinusoids across different repetition periods. The format of the plots is the same as in (A).

the overall preference for the damped sinusoids was maintained (Fig. 8(A), bottom). Fig. 8(B) shows a neuron of the non-synchronized population that responded more strongly to ramped

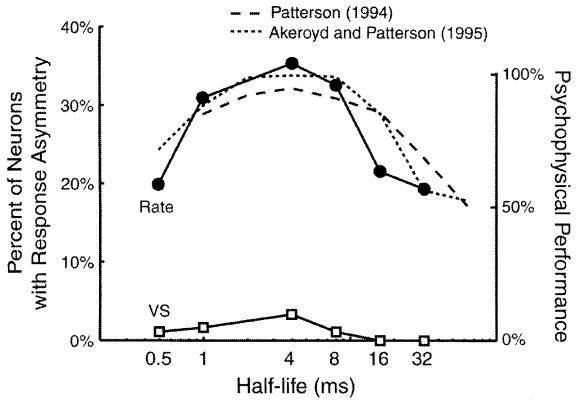


Fig. 9. Comparison between asymmetry preference of cortical neurons and human psychophysical performance in discriminating ramped and damped sinusoids (Lu et al., 2001a). The percentages of neurons (left ordinate) having significant asymmetry indices based on discharge rate are indicated as *rate* (solid line with filled circles). The percentages of neurons with significant asymmetry indices in vector strength are indicated as VS (solid line with open squares). Human psychophysical performance curves (right ordinate) using ramped and damped sinusoids is shown for tone carriers (dashed line), averaged over the different carrier frequencies (Patterson, 1994a), and noise carriers (dotted line) (Akeroyd and Patterson, 1995).

sinusoids at all repetition periods tested. These observations demonstrate that the sensitivity to temporal asymmetry within the temporal integration window is independent of a cortical neuron's ability to synchronize to stimulus events, suggesting a rate-based coding mechanism for time scales shorter than the temporal integration window for A1 neurons.

In Fig. 9 we compare response asymmetry of populations of A1 neurons with psychophysical performance in discriminating ramped versus damped sinusoids by humans (Lu et al., 2001a). The shape of the curve based on average discharge rate is qualitatively similar to psychophysical data with both tone carriers (Patterson, 1994a) and wide-band noise carriers (Akeroyd and Patterson, 1995). Psychophysical performance across half-life appears to be related to the percentage of A1 neurons that showed significant response asymmetry in their average discharge rates. A population measure based on discharge synchrony, on the other hand, reveals that only a very small portion of A1 neurons (<5%) showed response asymmetry in their temporal discharge patterns for the stimulus period used (25 ms).

## 5. The role of temporal selectivity and its behavioral relevance in cortical processing of species-specific vocalizations

The neural mechanisms involved in producing the temporal selectivity discussed above (at long and short-time scales) may possibly contribute to neural selectivity of complex vocalizations (Esser et al., 1997; Margoliash, 1983; Wang et al., 1995a). It has been shown that natural vocalizations of marmoset monkeys produced stronger responses in A1 than do spectrally similar, but temporally altered vocalizations (Wang et al., 1995a). Fig. 10 shows that a subpopulation of neurons in A1 of anesthetized marmosets responded more strongly to a natural twitter call than to its time-reversed version. Responses of this subpopulation of neurons were found to have a clearer representation of the spectral shape of the call than did responses of the non-selective neurons (Wang, 2000; Wang et al., 1995a). While these observations demonstrate a role of temporal selectivity in cortical responses to complex vocalizations, they also suggest that marmoset auditory cortex may preferentially respond to sounds with behavioral significance. To further test this notion, we have directly compared responses to natural and time-reversed calls in the auditory cortex of the cat, an extensively studied mammalian species, whose A1 shares similar basic physiological properties (e.g., CF, threshold, latency, etc.) to that of the marmoset (Aitkin and Park, 1993; Schreiner et al., 2000). Unlike neurons in A1 of marmosets, however, neurons in cat A1 did not differentiate natural marmoset vocalizations from their time-reversed versions (Wang and Kadia, 2001). Together, these observations suggest that temporal selectivity of cortical neurons may be dependent on the behavioral relevance of acoustic signals that a species encounters. A particular form of temporal selectivity under one behavioral context may not exist under other behavioral contexts. One potential mechanism that may give rise to such behavioral dependence is experience-dependent cortical plasticity (Buono-

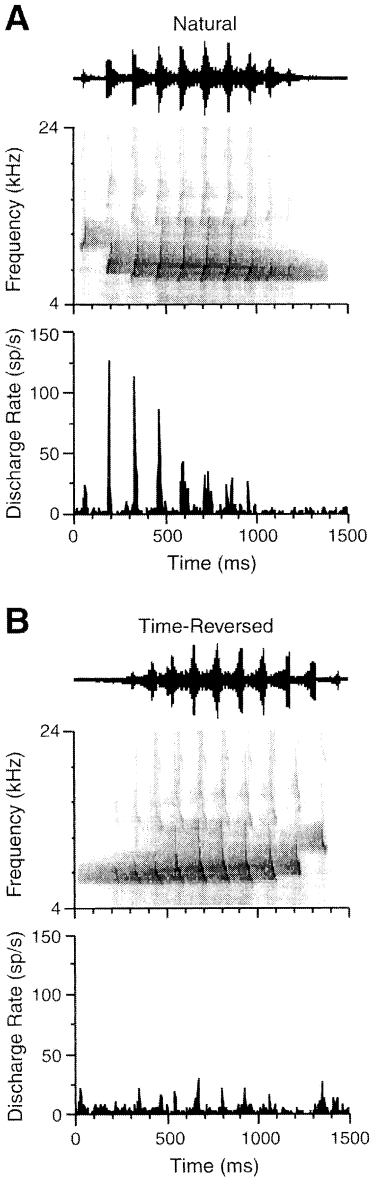


Fig. 10. Temporal discharge patterns of responses to a natural marmoset twitter call (A) and its time-reversed version (B) recorded from marmoset A1 (Wang et al., 1995a). Both the waveform (top) and spectrogram (middle) of the stimuli are shown. The corresponding cortical responses are shown in the form of a composite PSTH (bottom), averaged over a population of neurons that responded more strongly to the natural twitter call than to the time-reversed twitter call.

mano and Merzenich, 1998; Merzenich et al., 1984). In fact, it has been shown in the somatosensory cortex that temporal aspects of a complex stimulus are crucial in determining specific forms of learning-induced cortical reorganization (Wang et al., 1995b). We have argued (Wang, 2000) that the two central issues in our understanding of cortical representation of communications sounds are: (a) neural encoding of statistical structure of communication sounds and (b) the role of behavioral relevance in shaping cortical representations. Behaviorally relevant temporal modulations are clearly important acoustic features to be considered in the exploration of neural encoding mechanisms for speech-like signals in the auditory cortex.

## 6. Summary

Based on our findings from the studies discussed above, we suggest a two-stage model for processing temporal modulations by the auditory cortex. In this model, the auditory cortex integrates continuous acoustic streams over a temporal integration window of $\approx$30 ms. Temporal patterns that are separated by intervals longer than this integration window are explicitly coded by temporal discharge patterns of cortical neurons. Rapid time-varying components within the temporal integration window are instead represented implicitly by a discharge rate-based code. The combination of both temporal and rate codes should sufficiently encode the wide range of temporal modulations of biologically important complex sounds.

The significant reduction in the temporal limit on stimulus-synchronized discharges at the auditory cortex, as compared with the auditory periphery, has an important functional implication. It suggests that cortical processing of sound streams operates on a "segment-by-segment" basis rather than on a "moment-by-moment" basis as found in the auditory periphery. This is perhaps necessary for complex integration and comparison to take place at this level of the auditory system, since higher-level processing tasks require a broader view of acoustic events preceding and following a particular time of interest. Moreover, auditory information is encoded at the periphery at a much

higher temporal modulation rate than the rates at which the visual or tactile information is encoded. The slow-down of temporal response rate along the ascending auditory pathway allows fast-paced auditory information to be integrated in the cerebral cortex with information from other sensory modalities that is intrinsically slower.

## Acknowledgements

We would like to dedicate this article to Dr. Murray B. Sachs, a regular participant in the previous Utrecht Speech Perception workshops, whose work on neural encoding mechanisms at the auditory-nerve and cochlear nucleus has paved the way for our understanding of their transformations at higher auditory centers. This research was supported by Whitaker Foundation Research Grant RG-96-0268, NIH-NIDCD Grant DC03180 and by a Presidential Early Career Award for Scientists and Engineers (X. Wang). We thank Dr. Ross Snider and Steve Eliades for technical assistance, Ashley Pistorio for colony management, animal training, graphic assistance and proofreading the manuscript. Publications from our laboratory referred to in this article can be obtained at http://www.bme.jhu.edu/~xwang/papers.html.

## References

Agamaite, J.A., Wang, X., 1997. Quantitative classification of the vocal repertoire of the common marmoset (*Callithrix Jacchus Jacchus*). Assoc. Res. Otolaryng. Abstr. 20, 144.

Aitkin, L., Park, V., 1993. Audition and the auditory pathway of a vocal new world primate, the common marmoset. Prog. Neurobiol. 41, 345–367.

Akeroyd, M.A., Patterson, R.D., 1995. Discrimination of wideband noises modulated by a temporally asymmetric function. J. Acoust. Soc. Amer. 98, 2466–2474.

Bieser, A., Müller-Preuss, P., 1996. Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. Exp. Brain Res. 108, 273–284.

Blackburn, C.C., Sachs, M.B., 1989. Classification of unit types in the anteroventral cochlear nucleus: post-stimulus time histograms and regularity analysis. J. Neurophysiol. 62, 1303–1329.

Blackburn, C.C., Sachs, M.B., 1990. The representations of the steady-state vowel sound /ɛ/ in the discharge patterns of cat anteroventral cochlear nucleus neurons. J. Neurophysiol. 63 (5), 1191–1212.

Buonomano, D.V., Merzenich, M.M., 1998. Cortical plasticity: from synapses to maps. Annu. Rev. Neurosci. 21 (4), 149–186.

Creutzfeldt, O., Hellweg, F.-C., Schreiner, C., 1980. Thalamocortical transformation of responses to complex auditory stimuli. Exp. Brain Res. 39, 87–104.

de Ribaupierre, F., Goldstein Jr., M.H., Yeni-Komshian, G., 1972. Cortical coding of repetitive acoustic pulses. Brain Res. 48 (1), 205–225.

de Ribaupierre, F., Rouiller, E., Toros, A., de Ribaupierre, Y., 1980. Transmission delay of phase-locked cells in the medial geniculate body. Hearing Res. 3 (1), 65–77.

Eggermont, J.J., 1991. Rate and synchronization measures of periodicity coding in cat primary auditory cortex. Hearing Res. 56 (1–2), 153–167.

Eggermont, J.J., 1994. Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity. Hearing Res. 74 (1–2), 51–66.

Esser, K.-H., Condon, C.J., Suga, N., Kanwal, J.S., 1997. Syntax processing by auditory cortical neurons in the FM–FM area of the mustached bat pteronotus parnellii. Proc. Natl. Acad. Sci. USA 94, 14019–14024.

Evans, E.F., Whitfield, I.C., 1964. Classification of unit responses in the auditory cortex of the unanesthetized and unrestrained cat. J. Physiol. 171, 476–493.

Frisina, R.D., Smith, R.L., Chamberlain, S.C., 1990. Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement. Hearing Res. 44, 99–122.

Gaese, B.H., Ostwald, J., 1995. Temporal coding of amplitude and frequency modulation in the rat auditory cortex. Eur. J. Neurosci. 7, 438–450.

Goldberg, J.M., Brown, P.B., 1969. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. J. Neurophysiol. 32, 613–636.

Goldstein Jr., M.H., Kiang, N.Y.-S., Brown, R.M., 1959. Responses of the auditory cortex to repetitive acoustic stimuli. J. Acoust. Soc. Amer. 31, 356–364.

Houtgast, T., Steeneken, H.J.M., 1973. The modulation transfer function in room acoustics as a predictor of speech intelligibility. Acoustica 28, 66–73.

Johnson, D.H., 1980. The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. J. Acoust. Soc. Amer. 68, 1115–1122.

Joris, P.X., Yin, T.C.T., 1992. Responses to amplitude-modulated tones in the auditory nerve of the cat. J. Acoust. Soc. Amer. 91 (1), 215–232.

Krumbholz, K., Patterson, R.D., Pressnitzer, D., 2000. The lower limit of pitch as determined by rate discrimination. J. Acoust. Soc. Amer. 108 (3), 1170–1180.

Langner, G., Schreiner, C.E., 1988. Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms.. J. Neurophysiol. 60, 1799–1822.

Liang, L., Lu, T., Wang, X., 1999. Temporal encoding of amplitude modulated sounds with noise carrier in the lateral belt areas of the auditory cortex in awake marmoset monkeys. Soc. Neurosci. Abstr. 29.

Liang, L., Lu, T., Wang, X., 2002. Neural representations of sinusoidal amplitude and frequency modulations in the auditory cortex of awake primates. J. Neurophysiol. 87, 2237–2261.

Lu, T., Wang, X., 2000. Temporal discharge patterns evoked by rapid sequences of wide- and narrow-band clicks in the primary auditory cortex of cat. J. Neurophysiol. 84, 236–246.

Lu, T., Liang, L., Wang, X., 2001a. Neural representation of temporally asymmetric stimuli in the auditory cortex of awake primates. J. Neurophys. 85, 2364–2380.

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

Mardia, K.V., Jupp, P.E., 2000. Directional Statistics. Wiley, New York.

Margoliash, D., 1983. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. J. Neurosci. 3, 1039–1057.

Merzenich, M.M., Nelson, R.J., Stryker, M.P., Cynader, M.S., Schoppmann, A., Zook, J.M., 1984. Somatosensory cortical map changes following digital amputation in adult monkeys. J. Comp. Neurol. 224, 591–605.

Palmer, A.R., 1982. Encoding of rapid amplitude fluctuations by cochlear-nerve fibers in the guinea-pig. Arch. Otorhinolaryngol. 236, 197–202.

Patterson, R.D., 1994a. The sound of a sinusoid: Spectral models. J. Acoust. Soc. Amer. 96, 1409–1418.

Patterson, R.D., 1994b. The sound of a sinusoid: Time-interval models. J. Acoust. Soc. Amer. 96, 1419–1428.

Pressnitzer, D., Patterson, R.D., Krumbholz, K., 2001. The lower limit of melodic pitch. J. Acoust. Soc. Amer. 109 (5), 2074–2084.

Rall, W., Agmon-Snir, H., 1998. Cable theory for dendritic neurons. In: Segev, C.K.a.I. (Ed.), Methods in Neuronal Modeling. The MIT Press, Cambridge, MA.

Rosen, S., 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos. Trans. Roy. Soc. Lond. Ser. B, Biol. Sci. 336 (1278), 367–373.

Sachs, M.B., Young, E.D., 1979. Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate. J. Acoust. Soc. Amer. 66, 470–479.

Sachs, M.B., Blackburn, C.C., Banks, M.I., Wang, X., 1992. Processing of the auditory-nerve code for speech by populations of cells in the anteroventral cochlear nucleus. In: Schouten, M.E.H. (Ed.), The Auditory Processing of Speech: From Sounds to Words. Mouton De Gruyter Publishers, Berlin, pp. 47–60.

Schreiner, C.E., Urbas, J.V., 1988. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. Hearing Res. 32 (1), 49–63.

Schreiner, C.E., Read, H.L., Sutter, M.L., 2000. Modular organization of frequency integration in primary auditory cortex. Annu. Rev. Neurosci. 23, 501–529.

Wang, X., 2000. On cortical coding of vocal communication sounds in primates. Proc. Natl. Acad. Sci. USA 97, 11843–11849.

Wang, X., Kadia, S.C., 2001. Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. J. Neurophysiol. 86, 2616–2620.

Wang, X., Sachs, M.B., 1993. Neural encoding of single-formant stimuli in the cat. I. Responses of auditory-nervefibers. J. Neurophysiol. 70, 1054–1075.

Wang, X., Sachs, M.B., 1994. Neural encoding of single-formant stimuli in the cat. II. Responses of anteroventral cochlear nucleus units. J. Neurophysiol. 71, 59–78.

Wang, X., Sachs, M.B., 1995. Transformation of temporal discharge patterns in a VCN stellate cell model: Implications for physiological mechanisms. J. Neurophysiol. 73, 1600–1616.

Wang, X., Merzenich, M.M., Beitel, R., Schreiner, C.E., 1995a. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J. Neurophysiol. 74 (6), 2685–2706.

Wang, X., Merzenich, M.M., Sameshima, K., Jenkins, W.M., 1995b. Remodeling of hand representation in adult cortex determined by timing of tactile stimulation. Nature 378, 71–75.

Whitfield, I., Evans, E., 1965. Responses of auditory cortical neurons to stimuli of changing frequency. J. Neurophysiol. 28, 655–672.

Young, E.D., Sachs, M.B., 1979. Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. J. Acoust. Soc. Amer. 66, 1381–1403.

Zurita, P., Villa, A.E., de Ribaupierre, Y., de Ribaupierre, F., Rouiller, E.M., 1994. Changes of single unit activity in the cat's auditory thalamus and cortex associated to different anesthetic conditions. Neurosci. Res. 19 (3), 303–316.

Zwicker, E., Fastl, H., 1999. Psychoacoustics. Spinger, Berlin.