Spectrotemporal Receptive Fields in Anesthetized Cat Primary Auditory Cortex Are Context Dependent

In order to investigate how the auditory scene is analyzed and perceived, auditory spectrotemporal receptive fields (STRFs) are generally used as a convenient way to describe how frequency and temporal sound information is encoded. However, using broadband sounds to estimate STRFs imperfectly reflects the way neurons process complex stimuli like conspecific vocalizations insofar as natural sounds often show limited bandwidth. Using recordings in the primary auditory cortex of anesthetized cats, we show that presentation of narrowband stimuli not including the best frequency of neurons provokes the appearance of residual peaks and increased firing rate at some specific spectral edges of stimuli compared with classical STRFs obtained from broadband stimuli. This result is the same for STRFs obtained from both spikes and local field potentials. Potential mechanisms likely involve release from inhibition. We thus emphasize some aspects of context dependency of STRFs, that is, how the balance of inhibitory and excitatory inputs is able to shape the neural response from the spectral content of stimuli.

Keywords: inhibition, local field potential, multipeaked neurons, multiunits, spectral edges

Introduction

Auditory spectrotemporal receptive fields (STRFs) have increasingly been used to characterize the transfer function of a neuron and thus be able to predict its response to complex stimuli (Aertsen and Johannesma 1981b; Eggermont et al. 1981). This way of characterizing neurons relies on the linearity hypothesis for the neural response and has been criticized because the response to natural sounds with their numerous complex spectral components was poorly explained by STRFs (Aertsen and Johannesma 1981a; Theunissen et al. 2000). However, STRFs remain a very useful first tool in investigating how the information is “processed, encoded, and mapped to guide perception and behavior” (Fritz et al. 2007).

Although STRFs are assumed to be time invariant in order to maintain a stable and consistent representation of the sensory information, several studies hinted at the possibility that STRFs can be modified, temporarily or not, by the environment. Long-term modifications of the tuning and STRFs have been obtained following behavioral or sensory conditioning (Weinberger and Diamond 1987; Edeline 1999; Fritz et al. 2005), following neuromodulation (Edeline 2003), and after presentation of a spectrally enhanced acoustic environment for several weeks (Noreña et al. 2006) that induce massive reorganization of the cortical tonotopic map.

Fast neural mechanisms accounting for shaping STRFs likely rely on synaptic properties and adaptation processes. Very short-term adaptation (a few seconds) was found to induce systematic changes in STRFs obtained from inferior colliculus or primary auditory cortex (AI; Kvale and Schreiner 2004; Shechter and Depireux 2007). However, 2-tone studies revealed that STRFs are mainly shaped by inhibitory mechanisms related to forward suppression (Brosch et al. 1999; Brosch and Schreiner 2000). In this context, STRFs heavily depend on the spectrotemporal content of the stimulus used to derive them. For instance, increased spectrotemporal density revealed inhibitory sidebands in STRFs (Blake and Merzenich 2002; Valentine and Eggermont 2004). Recently, spectral shape (deCharms et al. 1998; Qin, Chimoto, et al. 2004) and spectral edge preference (Qin, Sakai, et al. 2004) of some cortical neurons have led to the idea that STRFs could be modified according to the spectral profile of stimuli and the resulting functional distribution of excitatory and inhibitory inputs to the neuron. These neural properties have potential functional roles, such as pitch detection (Qin et al. 2005).

Given that many natural sounds such as conspecific vocalizations have a limited critical bandwidth, we use narrowband stimuli in this study to generate STRFs. We show that STRFs obtained from narrowband and broadband stimuli differ by at least 2 points: 1) non-best frequency (BF) stimulation allows some residual peaks to appear in STRFs and 2) some specific spectral edges of narrowband stimuli induce increased firing. These 2 phenomena illustrate the context dependency of STRFs, which has many potential consequences for coding of complex spectral patterns and consequently for the use of STRFs as predictors of the neural response.

Material and Methods

Animal Preparation

Details about the anesthesia, the stimulus types, and the protocol have previously been reported in Tomita and Eggermont (2005) and Eggermont (2006). Summarizing, all animals were deeply anesthetized with the administration of 25 mg/kg of ketamine hydrochloride and 20 mg/kg of sodium pentobarbital, injected intramuscularly. A mixture of 0.2 ml of acepromazine (0.25 mg/ml) and 0.8 ml of atropine methyl nitrate (25 mg/ml) was administered subcutaneously at 0.25 ml/kg body weight. Lidocaine (20 mg/ml) was injected subcutaneously prior to incision. The tissue overlying the right temporal lobe was removed, and the dura was resected to expose the area bounded by anterior and posterior ectosylvian sulci. The cat was then secured with one screw cemented on the head without any other restraint. Additional acepromazine/atropine mixture was administered every 2 h. The cats were between 2.7 and 6.5 kg in weight. The ketamine dose to maintain a state of areflexive anesthesia was in the range of 6-13 mg/kg h. The care and the use of animals reported in this study was approved (BI 2001-021) and reviewed on a yearly basis by the Life and Environmental Sciences Animal Care Committee of the University of
Cortical Area Boundaries

The following properties were used in the assessment of cortical area boundaries: reversal of the characteristic frequency (CF) gradient in the tonotopic map and along the electrode array, minimum latency values, the shape of the STRF, and the peak value of the cross-correlation coefficient for recordings straddling boundaries. For delineating the border between AI and anterior auditory field (AAF), we first used the sign and/or reversal of the gradient of CF along the electrode array with distance in the anterior direction (Noreña and Eggermont 2005). The general shorter minimum latency in AAF compared with AI and, particularly, the much higher frequency-tuning curve bandwidth at 20 dB above threshold in AAF (Eggermont 1998) were used as well. For the distinction between AI and posterior auditory field (PAF) or potentially intermediate part of the posterior auditory field (Eggermont 1998) were used as well. For the distinction between AI and posterior auditory field (Eggermont 1998) were used as well.

Stimuli

Stimuli were generated in MATLAB (Natick, MA) and transferred to the DSP boards of a TDT-RP2 (Tucker Davis Technologies, Alachua, FL) sound delivery system. Acoustic stimuli were presented in an anechoic room from a speaker system (Fostex [Boonton, NJ] RM765 in combination with a Realistic super tweeter that produced a relatively flat spectrum [±5 dB] up to 40 kHz measured at the cat’s head) placed 30 cm from the midline into the contralateral field and 50 cm from the cat’s left ear. STRFs were obtained by presenting multifrequency stimuli consisting of randomly presented gamma-tone pips. The envelope of the gamma-tone pips is given by \( \gamma_t(t) = (1/4)e^{-t/4} \) with \( t \) in milliseconds. The duration of the gamma-tone pip at half-peak amplitude is 15 ms, and the envelope is truncated at 50 ms (Valentine and Eggermont 2004). Here, gamma-tone pips for each of 16 frequencies per octave in 5 or 7 octaves were randomly presented according to a Poisson process (Blake and Merzenich 2002), with similar average rate but different realization for each frequency. Each gamma-tone pip frequency was presented at an average rate of 1.5 Hz so that the aggregate gammatone-pip rate was 24 pips per octave per s. This broadband stimulus was presented during 900 s at 65 dB sound pressure level (SPL).

A narrowband stimulus was also used and consisted of 6 two-octave-wide frequency bands containing the same random tone pips at 24 pips per octave s. The overlapping frequency bands are [0.3125; 1.25], [0.625; 2.5], [1.25; 5], [2.5; 10], [5; 20], and [10; 40] kHz. Each band was presented 10 times during about 15 s with a silence interval of 250 ms between 2 bands so that the total duration of the stimulus is also 900 s.

Finally, a single-tone stimulus was used consisting of individual gamma-tone pips randomly presented every 250 ms at 65 dB SPL, 10 times repeated and covering 5 or 7 octaves (with 5–6 frequencies per octave).

Data Analysis

STRFs were determined by constructing frequency-dependent peri-stimulus time histograms for each of the gamma-tone pips in a 100-ms window that preceded the spike (Fig. 1). For that purpose, each spike elicited was plotted one time (single-tone stimulus) or several times (multitone stimuli) in the appropriate frequency bins and in the 100-ms time window after the onset of each of the preceding gamma-tone pips (because spike latency is a priori unknown). If the multitone stimuli had no effect on a spontaneously firing neuron, the entire matrix of 81 (5 octaves) or 112 (7 octaves) frequency bins by 100 (1 ms wide) time bins would be filled uniformly. If certain frequencies consistently produce excitation (respectively inhibition) in a certain latency window, then this part of the frequency time plane would receive more hits (respectively fewer hits) than average. STRFs were subsequently smoothed by a 5 × 5 bin uniform filter. Before smoothing, STRFs were enlarged by replicating edge values so that the smoothing does not weaken STRFs edge values. The frequency time bin where the most hits were recorded for the broadband stimulus is defined as the BF. Peaks of neural response in STRFs were defined for a firing rate greater than 4 spikes per second per stimulus above the baseline. This value was based on visual inspection on the entire database. Among all the spike trains available, the 184 multiunits showing at least one peak in the STRF obtained from broadband stimulation were analyzed.

For narrowband stimuli, STRFs were obtained by 1) counting spikes for all frequency bands together, leading to one STRF similar to that obtained for the broadband stimulus and 2) separating spikes for each frequency band leading to 6 STRFs, one per band (Fig. 1).

For LFPs, STRFs were obtained by representing for each frequency the average time course of potentials following the presentation of tone pips. When a BF is assigned to LFPs, it is that of the spike-based edge values. The frequency time bin where the most hits were recorded for the broadband stimulus is defined as the BF. Peaks of neural response in STRFs were defined for a firing rate greater than 4 spikes per second per stimulus above the baseline. This value was based on visual inspection on the entire database. Among all the spike trains available, the 184 multiunits showing at least one peak in the STRF obtained from broadband stimulation were analyzed.

For narrowband stimuli, STRFs were obtained by 1) counting spikes for all frequency bands together, leading to one STRF similar to that obtained for the broadband stimulus and 2) separating spikes for each frequency band leading to 6 STRFs, one per band (Fig. 1).

For narrowband stimuli, STRFs were obtained by 1) counting spikes for all frequency bands together, leading to one STRF similar to that obtained for the broadband stimulus and 2) separating spikes for each frequency band leading to 6 STRFs, one per band (Fig. 1).

For narrowband stimuli, STRFs were obtained by 1) counting spikes for all frequency bands together, leading to one STRF similar to that obtained for the broadband stimulus and 2) separating spikes for each frequency band leading to 6 STRFs, one per band (Fig. 1).

In addition to the use of the well-known unilateral Wilcoxon matched-pairs signed-ranks test (Wilcoxon 1945), we also use the one proportion test in this paper (Fleiss et al. 2003). It is based on the statistic \( F = \frac{k}{n} \), where \( k \) is the count of successes among \( n \) observations. Under the hypothesis that the theoretical proportion of successes is \( p \), we have approximately \( F \sim N \left( 0, \sqrt{\frac{p(1-p)}{n}} \right) \) when \( n > 30 \).

Results

Individual Examples

Figure 2 shows several examples of STRFs obtained from spike trains and LFPs in response to single-tone, broadband (7 octaves) and narrowband (2 octaves) stimuli. These individual examples represent single units in order to emphasize that the observed changes in the STRFs for multiple single-units, as reported in the remainder of the paper, are not due to different units’ properties. Most of the time, STRFs obtained in response to
to the single-tone stimulus reveal broad and poorly selective tuning, whereas the use of the broadband stimulus reveals one main frequency peak (the BF). Presentation of the narrowband stimulus enhances some residual peaks that were barely visible in the broadband-related STRF or even causes new peaks to appear at frequencies exciting the multiunit during single-tone presentation. Showing STRFs for each band reveals that these new peaks can appear entirely within the 2-octave frequency band (Fig. 2A,B,E,F) and/or specifically at the edges of the band (Fig. 2C,D). These peaks illustrate the capacity of some neurons

---

**Figure 1.** Stimulus presentation. Single-tone stimulus (top): randomized tone pips with a frequency range of 5 or 7 octaves are presented every 250 ms and each tone pip is repeated 10 times. Total duration of the stimulus is 67.5 and 95 s for the 5 and 7 octave stimuli, respectively. Broadband stimulus (top): random tone pips at 24 pips per octave per s are presented with a frequency range of 5 or 7 octaves during 900 s. Narrowband stimulus (bottom): same random tone pips within bands of 2 octaves wide are presented for 15 s; each of the 6 bands ([0.3; 1.25], [0.6; 2.5], [1.25; 5], [2.5; 10], [5; 20], and [10; 40] kHz) is presented 10 times. About 250 ms of silence separates every band presentation. Total duration of the stimulus is 900 s. For the narrowband stimulus, STRF are obtained for all bands together as well as for each band. In this latter case, the 6 resulting STRFs are displayed according to their frequency range.

**Figure 2.** Individual examples illustrating the modifications in the STRFs induced by the presentation of narrowband sounds. Each row (A and B; C and D; and E and F) corresponds to one recording site. (A, C, E) Single unit-based STRFs obtained from the presentation of (left to right) (column 1) single tones separated of 250 ms, (column 2) multiple tones at 24 pips per octave per s over 7 octaves (broadband stimulus), (columns 3 and 4) the same multiple tones organized in 15-s-long bands of 2-octaves wide (narrowband stimulus). Each band shows a 1-octave overlap with the previous and the following bands. The fourth STRF on the right is obtained by concatenation of STRFs for each band as described in Figure 1. Well sorted, presumably single units were used. The percentiles at 5% and 95% of the set of waveforms are represented in the small white rectangle for single-tone (blue solid lines), multitone broadband (dashed red lines), and multitone narrowband (dotted green lines) stimuli. (B, D, F) Same as (A, C, E) for LFPs, respectively. Same scales (spikes per stimulus per s or mV) have been used for multiple tones broadband- and narrowband-evoked STRFs in each case. (A, B, C, D) For single-tone stimulus, only frequencies above 1.25 kHz were presented.
to show selective responses within frequency bands far from their BF. For instance, in response to the presentation of the [10–40] kHz band, the units’ response peaks at 10 kHz, whereas its broadband STRF shows a narrow peak response at 7 kHz (Fig. 2C). If frequencies below 10 kHz are not present in the stimulus, the multunit thus shows an adapted response to slightly higher frequencies. It is assumed that STRFs obtained from LFPs partly show the local excitatory input distribution, which extends from at least 2.5 to 20 kHz in this case (Fig. 2D).

One should note the numerous residual peaks emphasized in LFP-based STRFs by narrowband stimulation and broadband stimulation to a lesser extent compared with single-tone-based STRFs (Fig. 2B,D,F).

Basically, when peaks that emerge at the edges of the narrowband stimuli are not taken into account, 3 basic results are observed in STRFs obtained from narrowband sounds, as illustrated in Figure 2: 1) 1 peak (Fig. 2C,D), 2) 2 peaks (Fig. 2A,B), and 3) 3 peaks or more (Fig. 2E,F). A reliable estimate of the proportion for each case is hard to obtain given the low signal-to-noise ratio of some peaks for several multiunits. Visual inspection of unambiguous cases (122 cases = 66.3%) gave, respectively, 33% (40 cases), 36% (44 cases), and 31% (38 cases).

### Firing Rate at the BF

Examples A and C in Figure 2 illustrate the increase in firing rate at the BF during presentation of narrowband and broadband stimuli. The firing rate remains lower than that obtained with the single-tone stimulus. Given that the "local" density of tones remains the same between the broadband and the narrowband stimuli (24 pips per octave per s), this effect is due to the removal of tones with frequencies distant to the BF in the narrowband stimuli. Figure 3 shows global results of the ratio of the maximal response at the BF in the STRF obtained from the broadband stimulus to that obtained during presentation of the narrowband stimulus. At all frequencies, removing distant tones generally induces an enhancement of the response at the BF at local (spikes) and global (LFPs) scales.

There is a small but significant correlation between the variation at the BF in the LFP-related STRF and that in the spike-based STRF (correlation coefficient = 0.317, $P < 10^{-4}$ for the Fisher’s $z$-transformation test).

### Unmasking of Inputs

We investigated the frequency distribution of peaks that were unmasked by the use of 2-octave-wide stimuli. Time frequency bins of STRFs that result in a stronger neural response for narrowband stimulus presentation than for broadband stimulus presentation are shown in Figure 4. For multiunits with low BFs (Fig. 4A), unmasked peaks occur at a large range of frequencies up to 20 kHz. As the BF increases, the proportion of peaks at 3 kHz appearing among the unmasked peaks increases (Fig. 4C,D). Similar results are obtained for LFPs for which a specific distribution of unmasked peaks at 3, 5, and 10 kHz is observed for high-BF neurons (Fig. 4H). Figure 4I–P shows the percentage of time frequency bins of STRFs obtained from broadband stimulation having a firing rate (respectively average response for LFPs) stronger than 10% of the maximum firing rate (respectively negative extrema) above (respectively below) the baseline. The wider bandwidth of LFP-based STRFs in average is confirmed. A large proportion of high-BF neurons also show activity around 3 kHz (LFPs and spikes), 5 kHz, and 10 kHz (mostly LFPs) for this low threshold of 10% of the maximum firing rate/negative extrema. More generally, for any BF, areas of unmasked inputs in spike- or LFP-based STRFs (Fig. A–H) are included in the frequency range of LFP-based STRF bandwidth (Fig. 4M–P). Thus, it appears that the specific distribution of unmasked peaks in narrowband-based STRFs (Fig. 4A–H) correspond to responses that were already present in the broadband-based STRFs (Fig. 4I–P), albeit weaker. In the individual examples from Figure 2, for instance, in one case (Fig. 2A,B) the residual peak at 3 kHz is weak but visible in the broadband LFP-based STRF, not in the spike-based one, and in another case (Fig. 2E,F), residual peaks are present, albeit very weakly, in both broadband LFP- and spike-based STRFs. This result also shows that the frequency edges initially chosen for narrowband stimuli (0.6, 1.2, 2.5, 5, 10, and 20 kHz) are not the source for the specific distribution of unmasked peaks at 3, 5, and 10 kHz.

In any case, these results emphasize the distribution of excitatory inputs and shows that input circuitry may be very different for low- and high-BFs neurons. In particular, high-BFs neurons are unexpectedly associated to overrepresented neural responses at 3, 5, and 10 kHz. Overall, a high percentage of neurons shows stronger responses at non-BF frequencies when using narrowband stimuli (see % scales in Fig. 4), which suggests that more than half neurons show a context dependency of their STRF. Peaks can appear up to 2 or 3 octaves above or below any BF as obtained for broadband stimuli.

We now investigate frequency profiles of peaks at the BF (Fig. 5), that is, average firing rate for frequencies at and around the BF. For multiunits, presentation of narrowband sounds

---

**Figure 3.** Ratio of the (A) firing rate of spikes and (B) average amplitude of LFPs at the BF during the presentation of the narrowband stimulus and during the broadband stimulus. BF is chosen as the frequency inducing the strongest response during the broadband stimulus. Response at the BF during the narrowband stimulus was considered to be in the band where the BF is the farthest from an edge. Dashed line shows locally weighted average. Log scales are used for the ordinate.
results on average in a much larger response bandwidth, albeit not as large as obtained for the single-tone stimulus. Interestingly, this enlarged bandwidth appears as a stronger response at frequencies far from the BF rather than a broadening of the peak at the BF. Similar results hold for LFPs but the peak bandwidth is always larger than for multiunits. For very low BFs, response profile enlargement is toward frequencies above BF, whereas the opposite happens for very high BFs. Peaks at 2 and 3 octaves below the BF in Figure 5G,H correspond to specific peaks emphasized in Figure 4G,H.

**Edge Enhancement**

For responses detected in the STRFs obtained from narrowband stimuli, we now compare the neural response frequency to the
The first result also suggests that release from inhibition is generally stronger close to the BF than further away from it and, consequently, the synaptic inhibitory input distribution would then generally be decreasing with the distance to the BF. Also, ratios \( R_{\text{below}}/P \) and \( R_{\text{above}}/P \) are strongly correlated for spikes and LFPs (0.555, \( P < 10^{-4} \), 0.587, \( P < 10^{-4} \), respectively), which suggests that this property of edge enhancement may not be local.

Amazingly, LFP latencies of peaks at spectral edges are much shorter than those at the BF (Fig. 7B, Wilcoxon test, peak below the edge \( P < 0.001 \), above the edge \( P < 0.001 \)). The phenomenon is also visible for spikes, although it is not significant (peak below the edge \( P = 0.96 \), above the edge \( P = 0.12 \)). This result suggests that processes leading to spectral edge preference differ from those generating the main peak response at the BF and might involve early subcortical or thalamic processing. In any case, edge enhancement seems to be a stable process and thus is likely not the result of some kind of progressive neural adaptation over a few seconds. When splitting each stimulus presentation block of 15s (see methods) into 2 intervals [0, 7.5] s and [7.5, 15] s, no significant difference between peak latencies at the edges was found (Fig. 7C, Wilcoxon test, peak below the edge \( P = 0.149 \), above the edge \( P = 0.182 \)).

Simulating Synaptic Input Distributions

In order to investigate the mechanisms potentially explaining the above results, we designed a simplified model, which attempts to reproduce the main principles of spikes generation related to the balance of excitatory and inhibitory inputs. This model is inspired by mechanisms described in Oswald et al. (2006) and therefore endeavors to directly use curves of excitatory and inhibitory inputs distribution \( R_E(f) \) and \( R_I(f) \), respectively. We set a postsynaptic potential (PSP) at a resting level of 0 (arbitrary) and provide a stimulus at frequency \( f_1 \) that adds to the current PSP an excitatory postsynaptic potential (EPSP) of amplitude \( R_E(f_1) \) (decreasing with exponential time constant 10 ms), and then after a delay of 3 ms, an inhibitory postsynaptic potential (IPSP) equal to \( R_I(f_1) \) (exponential time constant 25 ms) is added. The threshold for the PSP to elicit a spike was set at \( L = +0.2 \). After a spike discharge, the threshold \( L \) is set to 1 and decays exponentially to its resting level \( L = +0.2 \) with a time constant of 50 ms in order to reproduce effects of refractoriness and forward suppression. Spikes obtained are finally delayed by a value randomly chosen from a uniform distribution between 20 and 30 ms in order to simulate the physiological delays between the ear and the AI. The time course of EPSPs, IPSPs, membrane potential and firing
threshold, explicit equations describing them, as well as frequency response and temporal features of the model are shown in the Supplementary Material online.

We now use this simplified model for the 3 typical cases shown in Figure 2. The excitatory input distribution is based on STRFs obtained with LFPs (see Fig. 2). The model is able to closely reproduce the effects induced by narrowband sounds (Fig. 8), which are 1) appearance or increase of firing rate at residual peaks (Fig. 8C,E,M,O) and 2) increase of firing rate at frequency borders of the stimulus (Fig. 8H,J). This result can be obtained if the inhibitory input extends to a wider frequency area than the excitatory input (lateral inhibition model), even if these inputs are cotuned. The intermediate case when the inhibition input is broader than the excitatory one within
a similar spectral range (Wu et al. 2008) also leads to similar results (Supplementary Material online and Fig. 11).

**Discussion**

**Context Dependency**

Mammals show acoustic abilities over a large range of frequencies, typically up to 20 kHz for humans, up to 35 kHz for guinea pigs, and up to 40 kHz for cats. This range easily covers the frequency range of most natural sounds or conspecific sounds like vocalizations or speech. With respect to the tonotopy, which reflects the spatial organization of neurons' characteristic frequencies in a gradient along the AI of mammals, narrowband sounds should only excite neurons with corresponding CFs in the auditory pathways and thus induce responses in a limited number of cortical neurons. Yet, it is clear from a careful examination of "neurograms" (linear time frequency maps of the distributed responses to a particular stimulus [Nagarajan et al. 2002]) that vocalizations provoke some response, even limited, in a large percentage of neurons with BFs close to but outside the frequency content of the stimulus (see Fig. 4B, 9, and 10 in Gourevitch and Eggermont [2007], Fig. 4 in Wang [2000], and Fig. 5 in Nagarajan et al. [2002]). Part of the explanation may be the response of neurons to non-BF stimuli through cochlear nonlinearities and through at least 2 phenomena resulting from release from inhibition as shown in this study. These are 1) the appearance of response peaks in STRFs far from the BF and 2) the probability of discharge increases close to frequency borders of stimuli when its spectral content does not include frequencies close to the BF. These neural properties are not restricted to multiunits but are shared by large groups of neurons insofar as LFP-based STRFs show the same features at the same spectral edges or the same residual peaks as spike-based STRFs. Indeed, LFPs are compound excitatory PSPs and reflect important features of local synaptic activity (Mitzdorf 1985), even if there is uncertainty about the volume over which summed neural activity is recorded, which is probably less than 1 mm³ (Kaur et al. 2004). In any case, these 2 neural properties, mentioned above, underlie a context dependency of the STRF insofar as the STRFs are shaped according to the spectral bandwidth of the stimulus. The roles of such neural properties are discussed in the following 2 sections.

**Edge Enhancement**

Psychoacoustical studies have shown that spectral edges in masking noise were used by the listener to detect pure tones with frequency close to these edges (Emmerich et al. 1986; Fantini and Emmerich 1987; Allen et al. 1998) and for sound localization (van Schaik et al. 1999). Moreover, broadband noise with a sharp spectral edge creates a pitch percept in humans (Small and Daniloff 1967; Bilsen 1977; Klein and Hartmann 1981). The pitch is very close to the edge frequency. Similarly, a broadband noise containing a suppressed frequency band induces an auditory sensation similar in quality to a sinusoidal tone and its pitch always falls within the suppressed frequency band of the noise (Zwicker Tone, Zwicker [1964]).

A simple way of signaling a spectral edge of a stimulus could be the activation of neurons with BFs included in the stimulus spectrum compared with the background activity of non-activated neurons. However, it has been suggested that some AI neurons could specifically encode information about spectral

---

**Figure 7.** Latency of (A) maximum firing rate (multiunits) and (B) negative extrema (LFPs) at spectral edges as a function of the latency of the peak at the BF. Same latencies at spectral edges for (C) multiunits and (D) LFPs when only considering the first 7.5 s of stimulus presentation (ordinate) compared with those obtained only using the last 7.5 s. [A-D] Results have been split according to the fact that the peak is below \( P_{\text{below}} \) (see Fig. 6) or above \( P_{\text{above}} \) a spectral edge. Only peaks at edges showing a maximum firing rate >12 spikes per s per stimulus were considered. Peaks kept for LFPs are the same as for multiunits.
Systematic investigation of spectral edge preference in AI of awake cats with narrowband tones revealed that 34% of neurons (with BFs < 8 kHz) better responded to one cutoff side than the other and that 70% of neurons produced an edge effect response (Qin, Sakai, et al. 2004), a percentage comparable to but slightly less than in our study (Fig. 6B, D). The main difference with our study is that we found that edge enhancement extended up to 3 octaves below or above the CF (Fig. 6E–L).

In the study of Qin et al. although not as emphasized as in our study, neurons showing a response to stimulus edges also show enhanced responses when frequency bands of tones do not include the BF (see Fig. 3 in Qin, Sakai, et al. 2004), a percentage comparable to but slightly less than in our study (Fig. 6B, D). The main difference with our study is that we found that edge enhancement extended up to 3 octaves below or above the CF (Fig. 6E–L).

We also found an enhanced response above spectral edges in neurons with BFs below the edge and vice versa (Fig. 6A–D). This result seems true independent of the BF (Fig. 6E–L). This phenomenon and the spectral edge preference may be related to the asymmetry of inhibitory sidebands. Indeed, in a study in ferret AI (Shamma et al. 1993), cells that were strongly inhibited by frequencies higher than the BF responded best to stimuli that contained the least spectral energy above the BF, that is, stimuli with the opposite asymmetry as in our study. Similarly, in the rat AI, the width of the lower inhibitory sideband increased with CF, whereas the width of the upper inhibitory sideband decreased with CF (Zhang et al. 2003). However, no relation was found or shown between asymmetry of inhibitory sidebands and CFs in cat AI (Sutter et al. 1999). The relationship between direction selectivity, which is heavily related to inhibitory sidebands, and CFs is also controversial insofar as no significant relationship was found in cat AI (Heil et al. 1992; Mendelson et al. 1993), whereas up-sweep (respectively down-sweep) selectivity predominates in cells tuned to low (respectively high) frequencies in the rat AI (Orduna et al. 2001; Zhang et al. 2003).
All these results suggest that specific information carried by neurons with BFs close but outside the spectral content of the stimulus may help representing the frequency borders of the stimulus. One of the mechanisms for coding such borders may then be by releasing inhibition in neurons with neighboring BFs. Indeed, given the results of Figure 6, it is possible to speculate on the neural image of a narrowband stimulus in the AI: "some" neurons may show a higher firing rate if their BF is outside the frequency content of the stimulus than if it is inside. In particular, the property of asymmetry leads neurons with lower BFs to rather detect the bottom spectral edge of a stimulus and vice versa for neurons with high BFs. Theoretically, the strength of activated neurons related to their tuning properties might thus constitute a code for the detection of notch-like parts in stimuli spectra, which are present in many natural or conspecific sounds. This coding appears stable and likely is not the result of an adaptation process (Fig. 7C,D).

It is also striking that increased neural response in LFPs to spectral edges occur at short latencies (between 13 and 25 ms, Fig. 7B). It is thus likely that some spectral edge detection occurs at thalamic or more subcortical levels. Nevertheless, there is a strong difference with multiunit processing for which latencies are more comparable with the responses to the BF (Fig. 7A). One should note the crucial role of the high-spectrotemporal density of our stimulus in this result: release from inhibition of a neural response to a given spectral edge frequency can only be visible if the inhibition caused by the presence of distant frequencies to this edge is strong enough during broadband stimulation. This is only possible if these distant frequencies (including the BF) were presented less often during the 100–200 ms before the spectral edge-related frequency (the duration of maximal forward suppression found in Wehr and Zador [2005]). Thus, the spectrotemporal density has likely to be greater than a dozen of pips per octave per s. This is the reason why spectral edge features are not visible with single-tone stimuli, which lead to broad STRFs (Valentine and Eggermont 2004).

Unmasking Peaks

It is assumed that residual peaks in narrowband-sound-evoked STRFs reflect locally strong excitatory inputs masked by the predominance of the BF and the associated forward suppression in broadband STRFs. The absence of BF stimulation induces a release from inhibition illustrated by the result that the neural response to broadband stimuli is stronger on average than the response to broadband stimuli and weaker than that to single-tone stimuli at frequencies aside the BF (Fig. 5). The high-spectrotemporal density used in the narrowband stimulus allows the emergence of residual peaks, whereas the frequency selectivity is poorer when single-tones stimuli are used (Fig. 2). It is assumed that residual peaks in narrowband-sound-evoked STRFs reflect locally strong excitatory inputs masked by the predominance of the BF and the associated forward suppression in broadband STRFs. Interestingly, the release from inhibition associated to residual peaks is not uniform in frequency. If the asymmetric narrowband results were due to sampling differences one would expect a similar distribution (or scaled version thereof) of activity around the BF as obtained for single-tone and broadband stimulations. Looking at the extreme cases (Fig 5A,D,E,H), one observes that this is clearly not the case. This suggests that the asymmetries are not the result of sampling.

Because peaks tend to appear above BF for very low BFs and below BF for very high BFs (Fig. 5) leads to the unveiling of peaks with frequencies in the midfrequency range (1–20 kHz). It is possible that this result correlates with the better hearing sensitivity of the cat in this range (Liberman 1978; Heffner RS and Heffner HE 1985). However, this hypothesis would require that residual peaks are also found at lower levels of the auditory system, which remains speculative. Part of the explanation may also stem from the asymmetry of inhibitory sidebands described in the previous section. Local holes in the largest inhibitory sidebands, that is, those in the midfrequency range, could lead to peaks emerging in the STRFs when the stronger inhibition at the BF is released by the use of narrowband stimuli. Given that these unmasked peaks do not appear or only weakly if the neuron is also excited at its BF (by the broadband stimulus), it is possible that unmasked peaks reflect a specific response to the simultaneous presence or absence of particular spectral components in the stimulus.

Detection of complex components in conspecific sounds has been mainly studied through neural sensitivity to sequences of 2 tones in AI of monkeys or cats (Brosch et al. 1999; Brosch and Schreiner 2000). These latter studies revealed that such neural sensitivity is best when the interstimulus interval is around 100 ms and is generally less marked or even absent when tones are simultaneously presented. As a consequence, the neural response to the second tone depends on forward suppression mechanisms. Our results generalize the 2-tone studies by showing that multitone stimulation with various intervals between tones preserves some local response peaks as long as there is absence of BF stimulation.

What remains unclear in our study is why we observed a specific set of peaks around 3, 5, and 10 kHz for neurons with high BFs, whereas peaks (as in Norena et al. [2008]) were more uniformly distributed in frequency for neurons with lower BFs (Fig. 4). This overrepresentation at 3, 5, and 10 kHz involves residual peaks and thus cannot be detected easily through flat distributions of characteristic or best frequencies in other cat studies (see Fig. 4L, where this overrepresentation is barely visible and, e.g., Rajan et al. [1993] and Ehret and Schreiner [1997]). This result might only appear in studies of multipeaked neurons. Indeed, examination of multipeaked tuning curves found in cat’s AI also shows a large proportion of peaks at 5 and 10 kHz (Fig. 3 in Sutter and Schreiner [1991]) despite the limited sampling. Further investigation of the basic properties of hundreds of neurons in response to single-tone stimulation is clearly required in order to map the excitatory synaptic inputs in the AI of the cat. In any case, results shown in Figure 4 suggest that a specific distribution of excitatory synaptic inputs may occur for high-frequency neurons, whereas lower BF neurons reveal enhanced responses to a larger range of frequencies far from the BF.

Potential Mechanisms

STRFs obtained from narrowband stimuli show firing rate enhancement at the BF, at the spectral edges of the stimulus, or the appearance of new peaks at frequencies sometimes distant to the BF. The modifications in STRFs emphasized in this study are based neither on permanent functional changes following
behavioral or sensory conditioning (Weinberger and Diamond 1987; Fritz et al. 2005, 2007), long-term presentation of spectrally enhanced auditory environments (Noreña et al. 2006), nor on mechanisms linked to synaptic depression because narrowband and broadband stimuli used in our study share the same local spectrotemporal density (24 pips per octave per s). Knowing that, any enhancement at a given frequency in the STRFs obtained with narrowband stimuli should result from a decrease of the inhibitory synaptic inputs tuned to distant frequencies.

At the cortical level, effects of inhibition are progressively more understood. When stimulation is done by temporally isolated single tones at the same intensity level (65 dB SPL), the neural response is not very selective especially in dorsal and ventral Al of cats (Schreiner et al. 2000), as demonstrated by the large bandwidth of tuning curves at 65 dB SPL (single-tone STRFs in Fig. 2 and Valentine and Eggermont [2004]). At least 2 modifications of the stimulation induce narrower bandwidth: 1) decrease of intensity restricts excitatory input strength and reveals CFs of neurons and 2) increase of spectrotemporal density (Blake and Merzenich 2002; Valentine and Eggermont 2004). Temporally dense stimuli ranging from multiple tone stimulation as in our study or from continuous random harmonic sounds (Dynamic Moving Ripples, Temporally Orthogonal Ripple Combinations [Shamma et al. 1995; Klein et al. 2000; Depireux et al. 2001]) are generally used as a way to obtain STRFs by correlation techniques. In this case, forward suppression as well as lateral inhibition sharpens the peak of neural response, which is then usually called the BF. It is assumed that BFs are associated with the strongest excitatory inputs to the neuron. One should notice that peak bandwidth around BF is not dramatically modified by narrowband sounds (Fig. 2 and 5). Two-octave-wide band stimuli are thus wide enough to shape the peak at the BF in the STRF and to provide the maximum information about the stimulus frequency.

The balance between inhibition and excitation has led to 2 simplified models for the synaptic mechanisms shaping the receptive fields (Oswald et al. 2006): 1) lateral inhibition and 2) cotuning. In both cases, BF stimulation elicits a strong EPSP followed after a short delay (1–6 ms), by a strong IPSP. An action potential is elicited by a suprathreshold EPSP and is restrained in any case to the time interval before the IPSP appears. Also, in both cases, the cell receives strong inhibitory inputs during non-BF stimulation. In the lateral inhibition model, the cell does not receive strong excitatory inputs during non-BF stimulation, whereas excitatory and inhibitory inputs are roughly similar for any frequency stimulation in the cotuning hypothesis. Although the question is not completely solved, recent in vivo intracellular recordings in auditory cortical areas using voltage clamp techniques found cotuned inhibitory and excitatory inputs (Zhang et al. 2003; Tan et al. 2004; Wehr and Zador 2005). In contrast, a recent study found broader tuning for inhibitory inputs compared with excitatory ones, likely stemming from less-selective outputs associated to fast-spike neurons supposed to characterize inhibitory inputs (Wu et al. 2008). However, in this latter paper, inhibition and excitatory inputs shared a similar spectral range which makes this model essentially a hybrid of the lateral inhibition and co-tuned models.

Our model was designed to generate spikes according to the basic mechanisms described above. A delay of 3 ms between the EPSP and the IPSP was the average value found in intracellular recordings in rat AI (see Table 1 in Tan et al. [2004]). Time constants of 10 ms and 25 ms for EPSP and IPSP, respectively, were estimated from results in Wehr and Zador (2005) as well as time constant of 50 ms for forward suppression, which was found to last more than inhibitory conductance decrease. Finally, the shape of the excitatory input function was chosen according to response breadth in LFP-related STRFs insofar as LFPs are supposed to reflect important features of local synaptic activity. Additional descriptions of the model behavior are available in Supplementary Material online.

The results shown in Figure 8 have at least 2 implications: 1) the spectral edge preference and the unmasking of residual peaks are both well reproduced by our spike-generation model in response to narrowband sounds. This result suggests that both phenomena are relative simple consequences of the weights of inhibitory and excitatory inputs as well as forward suppression mechanisms considered in the model. In particular, these latter neural properties are sufficient to explain the level of release from inhibition observed in this study without having to account for any additional mechanism related to synaptic depression. Knowledge of the functions related to inhibitory and excitatory inputs as well as the time course of forward suppression thus helps to predict the neural response to complex stimuli, which is generally poor from knowledge of classical STRFs only (Aertsen and Johannesma 1981b; Theunissen et al. 2000) and 2) spectral edge preference and unmasking of residual peaks are both compatible with lateral inhibition and cotuning models for synaptic excitatory and inhibitory inputs (Fig. 8) as well as the intermediate case like that in Wu et al. (2008; Fig. 11 and Supplementary Material online). This result also suggests that both phenomena occur whenever the inhibitory synaptic input is equal or larger than the excitatory one. In particular, the spectral edge preference might be a general mechanism occurring as soon as the stimulus spectrum includes well-marked energy bands. One should note that peaks in our simulated STRFs rarely are as narrow as in real recordings. It is possible that additional nonlinear inhibitory processes contribute to sharpen response peaks close to BF and thus reduce ambiguity of frequency coding by the neuron.

Basic principles used here for simulating STRFs suggest that both edge effects and residual peaks are actually consequences of the same phenomenon: Integrated inhibition for a narrowband stimulus excluding BF frequencies is less than that for a narrowband stimulus including the BF frequencies. Consequently, any local peak in the excitatory input is enhanced by releasing inhibition from the BF. In the case of residual peaks, this local peak is a consequence of the connectivity network, which might favor specific frequencies at local scales. In the case of an edge effect, it is assumed that excitation input strength generally is decreasing with distance to the BF (Wu et al. 2008). As a consequence, for a narrowband stimulus, the spectral profile of the strength of the excited inputs shows a peak at the closest edge to the BF. In both cases, peaks can then appear on the condition that the spectrotemporal density of the stimulus is small enough to avoid massive inhibition due to forward suppression and high enough to avoid peaks with large bandwidth in STRFs as in single-tone stimulation.

Limitations and Perspectives

The use of multiunits for global results rather than single units may appear as a limitation at first sight. Obviously, the
identification of the same units between recordings obtained from stimuli sometimes separated by an hour is a difficult task, due to possible minor electrode movement modifying the amplitude or shape of action potentials, especially when considering that no perfect algorithm has ever been found for spike sorting. However, above all, we had 2 main reasons to use multiunits. 1) The narrowband stimulus lasts 15 min, a somewhat long period which yet means that each frequency band of narrowband stimuli was only presented for 150 s. Such a short period is associated with a low signal-to-noise ratio in the STRF. Moreover, 67% of our recordings showed a peak firing rate below 10 spikes per s per stimulus above the baseline (equivalent to 2.25 spikes above the baseline in the 1-ms time bin corresponding to the BF). Reliable peaks identification in STRFs obtained from single-units with necessary lower firing rates would thus be highly questionable, so would be the real existence of residual peaks. 2) Basically, the use of multiunits does not affect the 2 main results of our study. Edge effects as well as residual peaks likely happen for most of the single-units if they are visible for multiunits. Different residual peaks can still possibly stem from different units. Nevertheless, the existence of residual peaks unmasked is not in question insofar as each of these units would then necessarily induce a residual peak associated to a peak at the BF given the tonotopic axis. After visual inspection on well-isolated single units (2–4 putative single units were found per multiunit), we estimate that residual peaks appear in 35.3% (190/538) of cases, whereas edge effects seem to be present in 50.7% (273/538) of cases. Such estimates are slightly lower than for multiunits (78/184 = 42.4% and 117/184 = 63.6% respectively, also after visual inspection) but one also has to notice that roughly 20% of STRFs obtained from single units were very noisy and led to ambiguous interpretations.

Another potential limitation of the study is the unknown sites of this interaction between excitation and inhibition in the auditory system. There is no evidence that what we observe only occurs in the AI and not already at a subcortical level. Intensity tuning (firing rate decrease at higher intensity) progressively appears at cortical levels (primary cortex and mainly nonprimary areas) and is thus thought to imply central inhibition mechanisms (Sutter et al. 1999). With respect to frequency tuning, there is increasing evidence for a frequency match of inhibitory inputs to the excitatory ones at subcortical levels, especially in anteroventral cochlear nucleus and dorsal cochlear nucleus (pharmacological studies [Evans and Zhao 1993; Caspary et al. 1994]) and in the central nucleus of the inferior colliculus (chinichillas [Palombi and Caspary 1996]). The match also occurs at the cortical level according to intracellular recordings (Volkov and Galazjuk 1991; Tan et al. 2004; Wehr and Zador 2005; Wu et al. 2008). Further investigation at other cortical levels is clearly required, especially for spectral edge detection, which was found to occur in LFP-related STRFs as well (Fig. 6B).

Another limitation concerns the random nature of sounds used in this study. Natural sounds such as conspecific vocalizations contain more regular spectral components such as harmonic sounds, long-lasting frequencies, as well as frequency modulation sweeps (e.g., Gourevitch and Eggermont 2007)). It is thus somewhat speculative how phenomena found in this study combine and what their precise role and effect is when natural sounds are processed. In any case, it is clear that some parameters of narrowband sounds used in this study should be more systematically studied: increasing the number of bands and varying their width should help characterizing the spectral edge sensitivity of neurons. Given the common finding of multiple inhibitory bands found in tuning curves of neurons in AI (Sutter et al. 1999), it is also likely that using multiple separated frequency bands may modify STRFs.

Even with all these cautions, one major use of STRFs remains the prediction of neural responses to complex stimuli. We showed that classical and context-dependent STRFs are easily obtained from specific distributions of inputs applied to a simple model of spike generation. As a consequence, it is possible that improvements in predicting the neural response to complex spectral or temporal components of natural sounds may stem from the use of estimates of distributions of inhibitory and excitatory inputs rather than the use of classical STRFs. Consistently, modifications of the balance of excitatory and inhibitory inputs in cortical integration models allowed reproducing many variations of the neural discharge patterns (Xing and Gerstein 1996; Shadlen and Newsome 1998). For the model, we designed excitatory inputs curves based on LFP-based STRFs obtained from narrowband stimuli. At first sight, it seems that this curve would be more reliably estimated using LFP-based STRFs obtained from single-tone stimulation. However, it is well known that spike-based STRFs obtained from single tones show a large bandwidth making the BF and the residual peaks barely visible (Valentine and Eggermont 2004). In LFP-based STRFs, this becomes an issue insofar as single-tone stimulation seems to “saturate” the peak response over frequency. In our examples from Figure 2, the peak response is flat over frequency and the residual peaks are completely invisible. A possible explanation is the volume integration performed by LFPs which smoothes the frequency response. In any case, by sending multiple tones within a narrow frequency band, the small variations of excitatory strength with frequency tend to be more visible. This was a rationale for our choice of using narrowband-related STRFs to illustrate the excitatory input curve. Nonetheless, it is clear that estimation of excitatory/inhibitory inputs distribution using extracellular recordings remains a delicate issue and that further investigation in this domain is clearly required.

Also, one interesting perspective of such study is the context dependency in nonprimary areas. Indeed, it is assumed that such areas process complex spectral patterns (Gourevitch and Eggermont 2007) and should thus benefit of neural coding of spectral edges, for instance, in lower cortical areas. Massive synaptic integration from subcortical levels as well as AI occurs in nonprimary areas as reflected for instance in the cat PAF by the larger bandwidth of frequency tuning and longer latencies (Heil and Irvine 1998; Loftus and Sutter 2001). Moreover, the latter study found that inhibitory STRFs have the same complexity in PAF than in AI. It can thus be hypothesized that context dependency should also be found in nonprimary areas.

Conclusion

STRFs evoked by narrowband sounds revealed context dependency through at least 2 distinct phenomena: spectral edge sensitivity of neurons and appearance of residual peaks far from the BF. These intrinsic properties of STRFs are assumed to
directly stem from the release from inhibition and thus from the balance between inhibitory and excitatory inputs. Subcortical coding of spectral edges is suspected. Knowledge of synaptic distribution of inputs should allow prediction of the effects observed in STRFs and thus should help refine the prediction and understanding of neural responses to complex stimuli.

Supplementary Material
Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding
Alberta Heritage Foundation for Medical Research; The Natural Sciences and Engineering Research Council: 1206-05; Canadian Institutes of Health Research New Emerging Team Grant: NET-54023; The Campbell McLaurin Chair of Hearing Deficiencies.

Notes
The authors thank Martin Pienkowski for helpful discussions. Conflict of Interest: None declared.

Address correspondence to Jos J. Eggermont, PhD, Department of Psychology, 2500 University Drive Northwest, University of Calgary, Calgary, Alberta, Canada T2N 1N4. Email: eggermon@ucalgary.ca.

References


