Behavioral/Systems/Cognitive

Level Invariant Representation of Sounds by Populations of Neurons in Primary Auditory Cortex

Srivatsun Sadagopan and Xiaoqin Wang
Laboratory of Auditory Neurophysiology, Departments of Neuroscience and Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

A fundamental feature of auditory perception is the constancy of sound recognition over a large range of intensities. Although this invariance has been described in behavioral studies, the underlying neural mechanism is essentially unknown. Here we show a putative level-invariant representation of sounds by populations of neurons in primary auditory cortex (A1) that may provide a neural basis for the behavioral observations. Previous studies reported that pure-tone frequency tuning of most A1 neurons widens with increasing sound level. In sharp contrast, we found that a large proportion of neurons in A1 of awake marmosets were narrowly and separably tuned to both frequency and sound level. Tuning characteristics and firing rates of the neural population were preserved across all tested sound levels. These response properties lead to a level-invariant representation of sounds over the population of A1 neurons. Such a representation is an important step for robust feature recognition in natural environments.

Key words: **primary auditory cortex; frequency tuning; level invariance; neural coding; awake primate; marmoset**

Introduction

The range of sound intensities in the environment is vast, but behavioral studies have shown that auditory perception is robust despite large variations in sound level. In humans, psychophysical measures of speech recognition in quiet show a monotonic and saturating increase in performance with increasing sound level (Hirsh et al., 1954; Studebaker et al., 1993). In noise, performance saturates (Hirsh et al., 1954) or slightly degrades (Studebaker et al., 1999) with increasing sound level. These observations suggest that the representation of sounds at some stage of the auditory pathway should be relatively independent of sound level. However, frequency tuning curves of neurons in the auditory periphery and brainstem widen at higher sound levels (Nomoto et al., 1964; Kiang, 1965; Popper and Fay, 1992). It is unclear where and how such level-intolerant tuning is transformed into a level-invariant representation that could account for behavioral observations.

The response of a neuron to pure tones of varying frequency and level, known as its frequency response area (FRA), is a fundamental receptive field measure in the auditory system. Subcortical FRAs, up to the inferior colliculus (IC), are mostly level intolerant in that FRAs become wider at higher sound levels (Rose et al., 1963; Popper and Fay, 1992; Davis, 2005), with a few exceptions. Level-tuned units have been reported in the dorsal cochlear nucleus (Spirou and Young, 1991) and the IC (Yang et al., 1992; Ramachandran et al., 1999; LeBeau et al., 2001), but their best levels (BLs) (sound level eliciting maximal response) are confined to a narrow range of near-threshold values. In the IC of the echolocating bat, Suga observed neurons that were narrowly tuned to tones ("pure-tone selective" units) and in some cases to sound level ("closed" response maps) as well (Suga, 1965a, 1969). Rouiller et al. (1983) reported that 74% of neurons in the auditory thalamus of anesthetized cats were nonmonotonically tuned to stimulus intensity using a metric based on slope changes in the rate-level function.

At the cortical level, Suga (1965b) observed that a small number (4 of 140) of units in the auditory cortex of anesthetized bats exhibited closed response areas. In subsequent studies of the Doppler-shifted constant frequency (DSCF) area of the unanesthetized bat, many level-tuned units were found that had an overrepresentation of the frequency and amplitude range corresponding to commonly observed echo frequencies and amplitudes (Suga, 1977; Suga and Manabe, 1982). The notion of "level-tolerant" receptive fields that maintain sharp frequency tuning with increasing level was introduced in these studies. Suga hypothesized that level-tuned units may create an amplitude-spectrum representation of stimuli on the cortical surface and aid in maintaining sharp frequency tuning over all sound levels. Based on these observations, Suga recognized the necessity of level-tolerant selectivity of neural responses, especially because communication and biosonar sounds could occur over different distances (Suga, 1992).

In awake primate auditory cortex, neurons with nonmonotonic rate-level functions spanning a wide range of best levels have been observed (Brugge and Merzenich, 1973; Pfingst and O'Connor, 1981). In particular, 78% of units in auditory cortex were found to have nonmonotonic rate-level functions in the

Received June 16, 2007; revised Feb. 20, 2008; accepted Feb. 20, 2008.
This work was supported by National Institutes of Health Grant DC-03180 (X.W.). We thank Elias Issa and Dr. Yi Zhou for many helpful comments and suggestions and Dr. Cory Miller for comments on this manuscript. We also thank Ashley Pistorio and Jenny Estes for assistance with animal care.
Correspondence should be addressed to Srivatsun Sadagopan or Xiaoqin Wang, Johns Hopkins University School of Medicine, 720 Rutland Avenue, 412 Traylor, Baltimore, MD 21205. E-mail: vatsun@jhu.edu or xiaoqin.wang@jhu.edu.
DOI:10.1523/JNEUROSCI.2743-07.2008
Copyright © 2008 Society for Neuroscience 0270-6474/08/283415-12$15.00/0

study by Pfingst and O'Connor (1981). However, these early studies did not systematically explore the relationship between frequency and level tuning, leaving open the possibility that nonmonotonic rate-level tuning is limited to frequencies at or near the best frequency (BF) of a unit. In the ~25 years since these studies, however, many studies have described FRAs of primary auditory cortex (A1) neurons as level intolerant in a variety of species such as cats (Calford et al., 1983; Heil et al., 1992; Schreiner et al., 2000; Moshitch et al., 2006), rats (Kilgard and Merzenich, 1999; Tan et al., 2004), mice (Linden et al., 2003), ferrets (Shamma et al., 1993), and macaque monkeys (Recanzone et al., 2000), to name a few. Except for the study by Recanzone et al. (2000), the above cited studies were all conducted in anesthetized animals. In A1 of anesthetized cats, a small number of studies observed neurons with closed FRAs (~20% of sampled units in the study by Sutter, 2000) or nonmonotonic rate-level functions (~23% of sampled units in the study by Phillips and Irvine, 1981).

Outside A1, a large fraction of units in the posterior field of the auditory cortex of the cat were found to exhibit nonmonotonic rate-level functions (Phillips and Orman, 1984). Polley et al. (2006) reported that 76% of units in the ventral auditory field of the rat had nonmonotonic rate-level functions, but their best levels were restricted to a near-threshold range. Thus, although a few studies have reported nonmonotonic level tuning at the level of auditory cortex, the majority of studies have reported a level-intolerant representation of sounds in A1. The significance of level-tuned neurons in A1 and their implications for a level-invariant representation remain unclear.

In the present study, we systematically measure FRAs in A1 of awake marmosets and show that a majority of FRAs exhibit level-tuned properties in quiet and noise. On the basis of these observations, we propose a representational framework for sounds in A1 that is different from previously described systems. Furthermore, the FRAs measured in A1 of awake marmosets are much more sharply tuned in frequency than shown in previous studies, resulting in fine frequency resolution at all sound levels. This could reflect a significant transformation in the representation of sounds that takes place between the subcortical pathway and the auditory cortex. We suggest that level-invariant coding in A1 is an important mechanism for coding sounds in dynamic and noisy acoustic environments.

Materials and Methods

Neurophysiology. We recorded from the right hemispheres of two awake marmoset monkeys. Details of surgical and experimental procedures are described in a previous publication (Lu et al., 2001). All experimental procedures were in compliance with the guidelines of the National Institutes of Health and approved by the Johns Hopkins University Animal Care and Use Committee. A typical recording session lasted 4–5 h, during which an animal sat quietly in a specially adapted primate chair with its head immobilized. The experimenter monitored the behavior of the animal via a television camera mounted inside the sound-proof chamber. When the animal was observed to close its eyes for a prolonged period of time, the experimenter ensured the animal opened its eyes before the next stimulus set was presented. A tungsten microelectrode (impedance 2–4 MΩ; A-M Systems, Carlsborg, WA) was positioned within a small craniotomy (~1 mm diameter) using a micromanipulator (Narishige, Tokyo, Japan) and advanced through the dura into cortex using a hydraulic microdrive (Trent-Wells, Los Angeles, CA). The contact of the electrode tip with the dural surface was verified visually, which allowed us to estimate the depth of each recorded unit. The estimated recording depth may not accurately reflect true recording depth because of tissue growth on the surface of the dura over multiple recording sessions. The experimenter typically advanced the electrode by ~25–50 μm and waited for a few minutes to allow the tissue to settle. During this period, a set of search stimuli were played that typically consisted of pure tones (~5 steps per octave), bandpassed noise, linear frequency modulated (lFM) sweeps, and marmoset vocalizations at multiple sound levels. This strategy of "burst" electrode movements with long waits while playing a wide array of stimuli helped us detect and isolate single units with very low spontaneous activity and avoid biases toward any particular kind of units. Proximity to the lateral sulcus, clear tone driven responses in the middle cortical layers and tonotopic relationship with other recorded units were used to determine whether a recording location was within A1.

Acoustic stimuli. Stimuli were generated digitally in Matlab (MathWorks, Natick, MA) at a sampling rate of 97.7 kHz using custom software, converted to analog signals (Tucker-Davies Technologies, Alachua, FL), attenuated (Tucker-Davies Technologies), power amplified (Crown Audio, Elkhart, IN), and played from a loudspeaker (Fostex FT-28D or B&W-600S3) situated ~1 m in front of the animal. The loudspeaker had a flat frequency response curve (±5 dB) across the range of frequencies of the stimuli used, with a calibrated level (at 1 kHz) of ~90 dB sound pressure level at a set level of 0 dB attenuation. Stimuli consisted of pure tones, usually 100 ms long with 5 ms cosine ramps, delivered every 300 ms in pseudorandom order and repeated five to eight times. We sampled the frequency axis in 0.1 octave steps (over a range of 2 octaves) and the level axis in 20 dB steps (over a range of 80 dB). These sampling intervals were chosen so that, in most cases, responses in at least two sampling bins were ensured. In some cases, we measured rate-level functions at the BF of a unit by delivering BF tones in 5 or 10 dB steps of sound level over a range of 90 dB. Wideband noise masker was generated using a real-time digital signal processor (RX6; Tucker-Davies Technologies), gated by a cosine window and started at least 500 ms before beginning probe tone delivery. The root mean square level of the masking noise was matched to peak-to-peak tone level and normalized every 5 s of delivery to ensure consistent presentation levels. The noise signal (attenuated appropriately) was added to the probe signal before amplification using a custom-built signal summer.

Analysis. We recorded the responses of 460 single units in A1 of two awake marmosets to pure-tone stimuli that spanned a wide range of frequencies and sound levels with sufficient sampling density. However, not all tested neurons responded significantly to these stimuli because of the high selectivity of auditory cortex neurons in the awake condition (Wang et al., 2005). We first verified that at least one stimulus evoked a significant excitatory response within a 150 ms window beginning 15 ms after stimulus onset using the t test ($p < 0.01$) compared with mean spontaneous activity over the entire stimulus set. A total of 348 of 460 units met this criterion. We used 15 ms as response latency for our windowing because the median latency (defined as time to 10% maximum response after stimulus onset) for "I"/"V" and "O" units was 21 and 25 ms, respectively, and the minimum latency was 16 ms over the entire population. Seventy-three units were excluded from additional analyses because they either responded only at the loudest level tested ($n = 34$) or their frequency tuning at one or more sound levels differed by more than two estimated bandwidths from the BF calculated at the best level ($n = 39$). The latter group of excluded units had "patchy" or ill-defined FRAs, for which estimating bandwidth, level-tuning width, and related parameters were rendered noisy by sudden "jumps" in frequency tuning with sound level. There was no systematic relationship between BF and sound level in the FRAs of these units. Frequency tuning curves of the remaining 275 units were then computed at each level, thresholded at 20% of peak response, and smoothed using a five-point moving window. We then fit an area-matched rectangle to a tuning curve at a given sound level by fixing its position at the centroid of the tuning curve and its height at maximum firing rate. The calculated widths of the rectangles were then taken as measures of bandwidth at that sound level. For a Gaussian function, this measure closely approximates the full-width at half-maximum, with the added advantage of not assuming any underlying tuning curve shape. The rectangle fit also takes into account long-tailed tuning curves, leading to conservative bandwidth estimates. The bandwidth of the unit was taken as the bandwidth at BL. For comparison with previous studies, we also measured quality factor (Q) values (Schre-

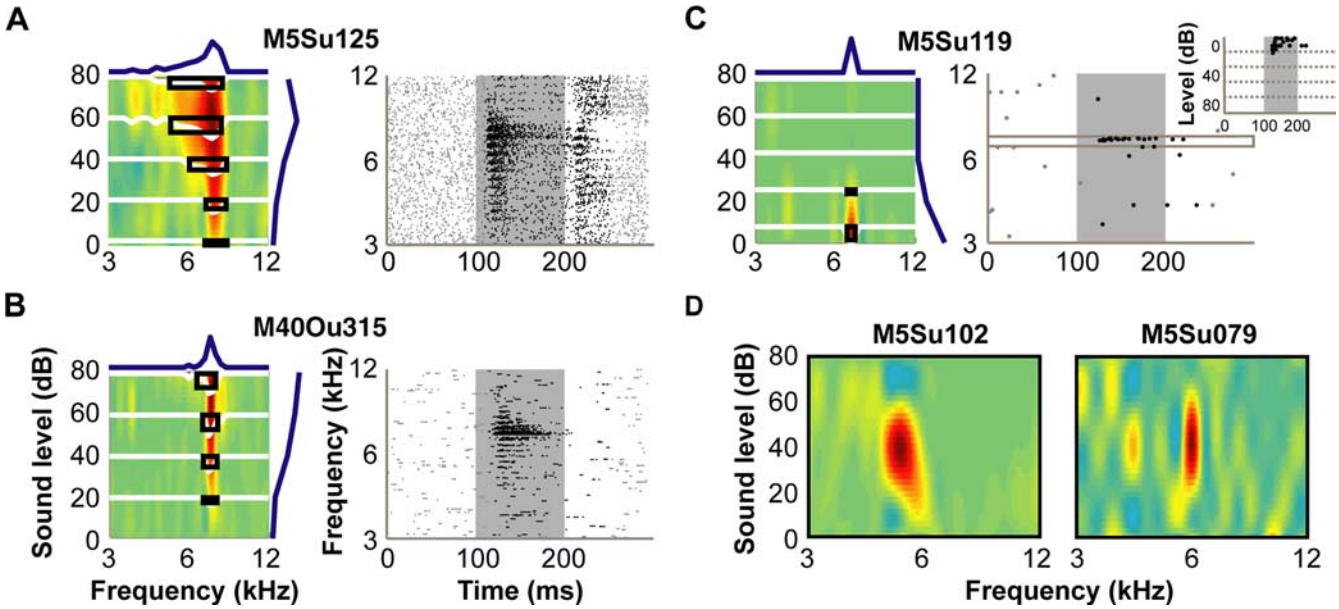

Figure 1. A1 FRA shapes. Frequency tuning curves (white lines) were computed at different levels, and area-matched rectangles (black boxes) were fit to them to determine bandwidth at each level and used to compute an SI. Blue lines on margins are "slices" of the FRA and denote frequency tuning at best level (top) and level tuning at best frequency (right). These were later used to determine separability of the FRA. ***A–D***, Based on FRA shape, neurons were classified as V (***A***), I (***B***), and O (***C***, ***D***) units (color map corresponds to normalized response rate). Corresponding spike rasters show stimulus-dependent temporal dynamics of the response (shaded area is stimulus duration; black dots are spikes falling inside our analysis window). Stimuli are ordered in blocks of increasing tone frequency. Within each frequency block, sound level is varied (gray box in ***C*** corresponds to a single frequency bin, expanded in inset).

iner et al., 2000), defined as the BF of a unit divided by its bandwidth, typically at 10 or 40 dB above threshold. Based on the bandwidth measures at each level, we then computed a shape index (SI), defined as the ratio of the bandwidth computed at the loudest level to the bandwidth at the BL. O-shaped FRAs tended to have decreasing bandwidths with increasing levels, leading to small SI values (in most cases close to zero). I- and V-shaped FRAs had increasing bandwidths that led to SI values equal to or greater than one. We used SI of 0.75 as the boundary to classify whether a unit belongs to the O or V and I categories. A visual inspection of a subset of the data verified this to be a satisfactory criterion.

To measure how separable the frequency and level tuning curves were, we determined what fraction of variance of the original FRA could be accounted for by a simple cross product of frequency and level tuning at BL and BF. O-shaped FRAs with a single peak were readily decomposed into a single frequency and level tuning curve leading to high separability (r^2) values. I-shaped FRAs were similarly separable, but V-shaped FRAs were usually not. We used $r^2 = 0.75$ to separate single and multi-peaked O units. To measure shifts caused by masking noise, we determined the ratio of shift caused in the BL to the magnitude of noise masker relative to the BL. Rate-level functions in these experiments were measured by presenting CF tones in 10 or 5 dB steps. In most cases, units exhibited shifts in BL when the noise masker level was equal to the BL (relative masker level of zero). In these cases, any positive shift in BL >5 dB was taken to be a 100% shift. We typically used Wilcoxon's rank-sum tests to compare medians and Kolmogorov–Smirnoff tests to compare distributions when more than one peak was observed in a distribution.

Model. We modeled the A1 neurons studied here as simplified, spectrally linear integrators with difference-of-Gaussian shaped frequency tuning. O units were modeled with Gaussian tuning to a BL, and V units had linearly increasing response rate and bandwidth with level. We added balanced inhibition such that a stimulus with bandwidth covering the excitatory and both inhibitory regions would produce zero response. In both cases, parameters were drawn with replacement from our experimental data. To avoid units being tuned completely outside the range covered by stimulus spectra, we constrained stimulus frequencies and unit BFs to a range of 2–16 kHz. The maximum firing rate of the O and V units were then adjusted to match the data medians, with the rates adapting to 40% of maximum level (time constants, 20 ms for O units, 15 ms for V units to simulate transient responses). We used 300 stimuli consisting of tone pips, lFM sweeps, random frequency contours, natural and reversed vocalizations, and environmental sounds, most of which we regularly used for other experiments in our laboratory. Stimuli were windowed every 10 ms using overlapping and causal exponential kernels. In each segment, Fourier transforms were computed, and stimulus levels were determined. These were then convolved with the FRA of a model unit to produce the response rate in each bin. The average of these rates was considered as the firing rate of that neuron in response to the stimulus. Because of the nonparametric nature of the test stimulus set, we used excess kurtosis of the firing rate distribution as a measure of selectivity and sparseness (Lehky et al., 2005). Briefly, if a single neuron responded with more or less the same mean firing rate to all 300 stimuli in the set, the distribution of the firing rates of this neuron across all stimuli would be well described by a normal distribution with an excess kurtosis of zero. However, if the neuron were highly selective, it would respond with high rates to very few stimuli and not respond to most stimuli, leading to a highly skewed firing rate distribution with high excess kurtosis. Similarly, response sparseness can be defined as the excess kurtosis of the distribution of firing rates of all 300 neurons in response to a single stimulus. Note that, by sparseness, we mean that few units of the population respond to a given stimulus simultaneously, and this term does not imply that each unit fires only a few spikes. To measure the response precision, we first computed the magnitude spectrum of the response histogram of each unit in response to each stimulus. The precision of each response was then defined as the fraction of power at frequencies >10 Hz. This measure characterizes the "peakiness" of the peristimulus time histogram (PSTH); high values reflect high-frequency modulations of the PSTH resulting from temporally precise firing.

Results

The analyses presented here were based on 275 units that responded significantly to at least one stimulus, responded to at least one sound level other than the loudest level tested, and did not have "patchy" FRAs that resulted from scattered frequency tuning across sound level (see Materials and Methods). We sampled over a large range of BFs (0.5–28 kHz).

Distinct types of FRAs in A1

Figure 1 shows representative examples of the three types of FRAs that we observed in A1 of awake marmosets. The neuron in Figure 1*A* is referred to as a V unit in this paper. The V shape of its FRA reflects the dependence of frequency tuning bandwidth on sound level. Figure 1*B* depicts an I unit, which exhibited the same bandwidth at all levels tested, resulting in a characteristic I-shaped FRA. Together, these two types of responses are similar to traditional descriptions of FRAs in A1 (Recanzone et al., 2000; Schreiner et al., 2000; Moshitch et al., 2006). In the present study, however, only 100 of 275 FRAs (36%) were classified as V or I shaped. The majority of units we observed (175 of 275, 64%) had O-shaped FRAs, such as the examples in Figure 1, *C* and *D*. These O units were narrowly and separably tuned to both frequency and sound level.

To classify these FRA types, we measured an SI, defined as the ratio of bandwidth at the highest sound level tested to the bandwidth at the best sound level (see Materials and Methods). This measure determines whether frequency tuning widens or sharpens with increasing sound level. Because O units were level tuned, we expected the bandwidths of these units to drop close to zero at loud sound levels, leading to SI values near zero. Because V and I units responded better at louder sound levels, we expected the bandwidths of these units at the loudest sound level to be equal to, or greater than, their bandwidths at the BL, leading to an SI of one or greater. Our data revealed a bimodal distribution of SI (Fig. 2*A*) with peaks at zero and one. We separated O units from I and V units using a criterion of SI $\leq$0.75. This classification boundary took into account O units tuned to high sound levels whose bandwidths decreased, but did not drop to zero, at the loudest level tested.

Analyses revealed that the SI alone could not distinguish between V and I units because the BL was often the loudest level tested for both types. Therefore, we measured to what extent the entire FRA could be reconstructed using independent frequency and level tuning components (Fig. 2*B*) (see Materials and Methods). Because frequency and level tuning of I units did not interact, their FRAs were separable ($r^2 \geq 0.75$), whereas the FRAs of V units were usually not separable. However, the r^2 values of V and I units were distributed smoothly along a continuum, and we could not reliably distinguish between these two FRA types. Therefore, we lumped V and I units together for all subsequent analyses. Most O units, however, could be expressed as a cross product of a level tuning curve (at BF) and a frequency tuning curve (at BL), but, consistent with a previous study in awake marmoset A1 (Kadia and Wang, 2003), a minority of O units ($n = 50$, ~20% of the total population) were multipeaked and could not be separated ($r^2 < 0.75$) into independent frequency and level tuning curves. We did not observe significant differences in the distributions of SI, BL, monotonicity, or frequency tuning widths between O units with $r^2 < 0.75$ and $r^2 \geq 0.75$ and therefore did not separate O units based on separability for the purposes of the present study. Median separabilities of O ($r^2 = 0.9$) and I/V ($r^2 = 0.69$) FRAs were significantly different ($p < 0.0001$, Wilcoxon's rank-sum test).

Figure 2*C* shows the distribution of the BLs of all units. The BLs of O units covered a large range, whereas the V units usually preferred the highest tested levels. We measured the ratio of the response rate of a unit at the loudest level tested to its response rate at the BL, termed it monotonicity index (MI), to establish a correlate with previous studies (Pfingst and O'Connor, 1981). An MI near zero indicates highly nonmonotonic rate-level characteristics, whereas an MI near one indicates a monotonic rate-level function. Most O units were strongly nonmonotonic (median MI of 0.25), and most I/V units were strongly monotonic (median MI of 1) (Fig. 2*D*). The fraction of nonmonotonic units we observed (MI $\leq$ 0.75; ~64% over entire response duration, ~76% during sustained response) was similar to that found in a previous study of A1 in awake and behaving macaques (78%) (Pfingst and O'Connor, 1981). We did not observe any differences between the BF distributions of O and I/V units (Fig. 2*E*).

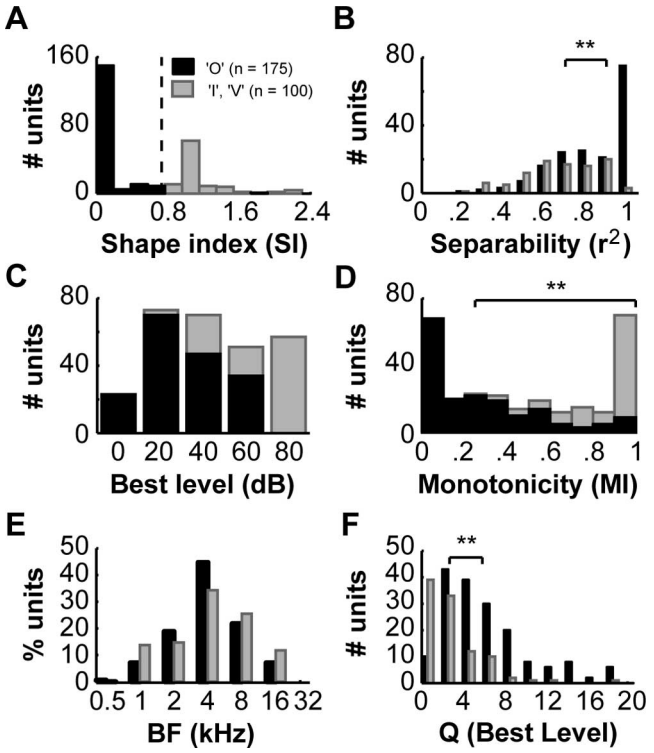

Figure 2. Distinct response types to pure-tone stimuli. ***A***, The distribution of SIs computed from the FRAs of 275 tone-responsive A1 units was bimodal. We used SI of 0.75 as a boundary to separate O (black; $n = 175$) from I and V (gray; $n = 100$) units. ***B***, Distributions of separability (r^2) computed for O and I/V units. The majority of O units were separable into frequency and level tuning components. ***C***, Distribution of best sound levels separated by unit type. O units covered lower sound levels, whereas most I/V units preferred high sound levels with a few exceptions. ***D***, MI was strongly correlated with FRA shape. Most O units were strongly nonmonotonic (MI of 0.25), and most V units had monotonic rate-level functions (MI of 1). ***E***, No significant difference was observed between the BF distributions of O and I/V units. ***F***, Q values calculated at best level show that O units are much more sharply tuned than I/V units. All values are medians; $^{**}p < 0.01$, statistical significance determined with Wilcoxon's rank-sum tests for equal medians.

In Figure 3*A*, we analyze the frequency tuning bandwidth and level tuning width for all units. Whereas I/V units populated this two-dimensional parameter space broadly and showed some overlap, O units were more selective, as evidenced by their clustering near low values. The median bandwidth of I/V units at BL was 0.52 octaves ($n = 100$), comparable with data from A1 of anesthetized cats (Schreiner et al., 2000; Moshitch et al., 2006) or awake macaques (Recanzone et al., 2000). In terms of Q values, I/V units exhibited a median $Q_{10\ dB}$ of 3.8 and median $Q_{40\ dB}$ of 1.8, similar to data from anesthetized cat A1 (Schreiner et al., 2000).

The median bandwidth of O units, however, was only 0.25 octaves ($n = 175$), significantly smaller than that of I/V units ($p < 0.0001$, Wilcoxon's rank-sum test). We also measured level tuning widths of O and I/V units by fitting rectangles to the rate-level functions at BF (see Materials and Methods). O units

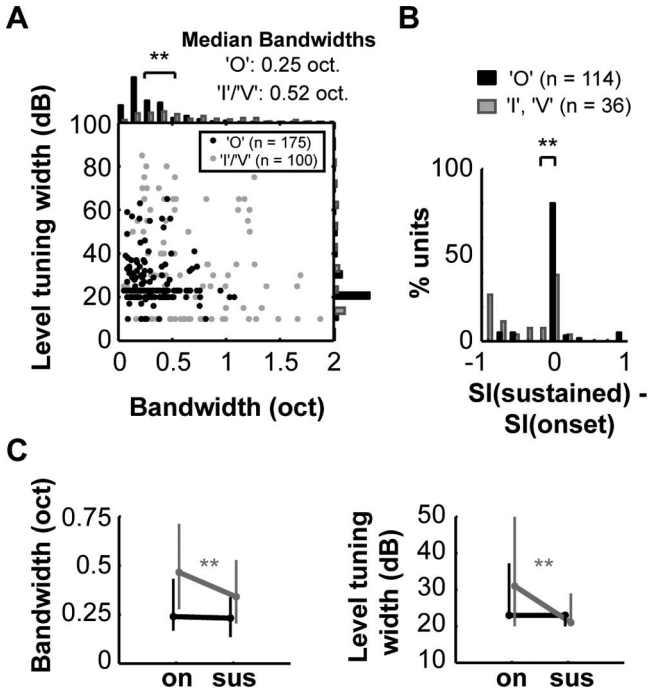

Figure 3. Temporal dynamics of O and I/V unit tuning properties. ***A***, O and I/V units encode frequency-level space differently (distributions are on margins). O units were finely tuned in both frequency (0.25 octaves) and level (25 dB), whereas I/V units were broadly tuned to frequency (0.52 octaves) and responded to a broader range of levels (32 dB). ***B***, Distribution of SI difference between the onset (first 50 ms) and sustained (next 100 ms) of the response. O units retain shape throughout response duration, whereas V units show decreased SI during the sustained response window, indicating sharpening of the FRA. ***C***, When the onset (on, first 50 ms) and sustained (sus, next 100 ms) portions of the response were analyzed separately (medians and interquartile range are plotted), the bandwidth and level tuning width of I/V units dropped by 25 and 30%, respectively, and approached the resolution of the O units as the response developed over time. O units, conversely, retained the same resolution over the entire response duration. All values are medians; $^{**}p < 0.01$, statistical significance determined with Wilcoxon's rank-sum tests for equal medians.

responded to a narrow range of sound levels (median level tuning width of 25 dB), but, unlike O-shaped FRAs in the IC (Ramachandran et al., 1999), their BLs covered a wide range of sound levels (Figs. 2*C*, 4*A*). In comparison, I/V units responded to a broader range of sound levels (median level tuning width of 32 dB). For O units, median $Q_{10\ dB}$ and $Q_{40\ dB}$ values were 11.8 and infinity, much higher than those of I/V units. We preferred the bandwidth at BL to describe O unit tuning rather than Q values for two reasons: (1) O units could not be conceptualized in terms of a threshold and quality factor and (2) $Q_{40\ dB}$ for most O units was infinity and therefore unsuitable for comparisons. However, to facilitate comparison with previous data, we calculated a Q value at the BL of I/V and O units that reflects the bandwidth of a neuron at its widest point. At BL, I/V and O units had median Q values of 2.7 and 5.9, respectively. The distributions of Q values at BL are shown in Figure 2*F*. These data suggest that the selectivities of most neurons in A1 with respect to frequency and level tuning are much sharper than described previously (Calford et al., 1983; Recanzone et al., 2000; Schreiner et al., 2000; Moshitch et al., 2006).

Functionally, the different tuning properties of O and I/V units described here may be associated with different scales of representation in A1. Because I/V units populate frequency-level parameter space broadly, they may be better suited for detecting sounds at a coarse resolution, especially during early response periods. In contrast, O units may be better used for discriminating sounds at a finer scale throughout the duration of the response. These unit types also exhibited different response dynamics that could change the scale of representation over time. The raster plots in Figure 1, for example, suggest that the temporal evolution of the response is stimulus dependent. Typically, V and I units exhibited transient responses at many frequency-level combinations and sustained responses near their preferred stimuli (Wang et al., 2005) (Fig. 1*A*,*B*).

To quantify response dynamics, we compared tuning properties of O and I/V units that were active both during the onset (≤50 ms after stimulus onset) and sustained (next 100 ms) portions of the response for at least one stimulus (114 O units and 36 V units). Figure 3*B* is a plot of the distribution of the difference between the SI during the sustained and onset response for O and I/V units. The FRAs of most O units did not change shape over time (median SI difference of 0). Conversely, the SI of V units showed a significant reduction during the sustained portion of response (median SI difference of −0.2) ($p < 0.001$, Wilcoxon's rank-sum test), indicating a sharpening of their FRAs. Although the median bandwidth and level tuning width of the O units remained unchanged during the sustained portion of response, these two metrics decreased (sharpened) by 25 and 30%, respectively, in the V units ($p < 0.01$, Wilcoxon's rank-sum test) as the response became more sustained (Fig. 3*C*).

During the sustained portion of the response, the resolution of the I/V unit population approached that of the O unit population (Fig. 3*C*). In fact, the fraction of O units increased from 64 to 76% (114 of 150 neurons) when only the sustained response was analyzed. To summarize, whereas O units maintained a fine scale of processing throughout response duration, I/V units started out at a coarse scale and moved toward a finer processing scale with time. However, the sharpening of tuning properties did not affect the BLs of I/V units. It has been suggested in a previous study that such stimulus-dependent changes in response dynamics might create an initial widespread activation in auditory cortex that become more restricted depending on how well stimulus parameters match the spectral and temporal tuning characteristics of a neuron (Wang et al., 2005). Here we suggest a similar initial detection activation of many I/V units that narrows over time to a finer scale of processing based on the tuning properties of these neurons that adds to the population of already fine-tuned O units. This sharpening of receptive field size may be comparable with observations in the primary visual cortex (Malone et al., 2007). The sharpening occurs at a timescale similar to those observed in some higher visual areas such as posterior inferotemporal cortex (Brincat and Connor, 2006) and the middle temporal area (Pack et al., 2001).

The apparent restriction of the best level of O units to a 0–60 dB range (Fig. 2*C*) may be an experimental edge effect. In the reported experiments, the loudest level tested was 80 dB, and the level axis was sampled at 20 dB intervals. This meant that all units with a best level of 80 dB were by definition classified as I/V units, because we did not record responses at higher sound levels. It is likely that some of these I/V units would have been classified as O units had we been able to observe responses at higher levels. Therefore, our estimate of the proportion of O units among the responsive neurons (175 of 275, 64%) should be viewed as a lower bound of this type of A1 units. In addition, during the sustained portion of the response, the proportion of O units increases to 76%, and the bandwidth and level tuning width of the I/V units approaches that of the O units (Fig. 3*C*). Therefore, we suggest that a larger proportion of neurons participate in level-invariant

coding in A1 over a wider range of sound levels than indicated by the data shown in Figure 2*C*. The number of O units we observed was substantially greater than what has been reported previously in anesthetized animals. For example, Sutter (2000) reported that 20.4% of neurons in anesthetized cat A1 had circumscribed FRAs. Similarly, a previous study (Phillips and Irvine, 1981) reported that 23% of neurons in anesthetized cat A1 had nonmonotonic rate-level functions. The number of nonmonotonic units we found in our study is comparable with a previous study in awake behaving macaques (Pfingst and O'Connor, 1981).

The O units also differed from I/V units in other physiological properties that may explain why the predominance of O units in A1 has not been reported previously, even in unanesthetized preparations (Recanzone et al., 2000). Median spontaneous rate (O, 1.74 spikes/s; I/V, 4 spikes/s) and maximum driven rate (O, 22 spikes/s; I/V, 42 spikes/s) of O units were approximately half of the corresponding measures in I/V units. This suggests that O units would be harder to find than I/V units if the strategy for isolating a neuron during an experiment involved listening for spontaneous spikes while changing electrode position in a continuous manner. Because of their narrower tuning and lower maximum driven rate, O units would also be harder to drive if experimental stimuli did not sample the frequency and level axes finely. In comparison, Recanzone et al. (2000) report a mean spontaneous rate of 8.2 spikes/s and a mean peak driven rate of 39.3 spikes/s in A1. These values resemble our I/V units, lending support to our hypothesis that O units may have been missed previously because of search biases. We reduced the magnitude of these biases in our experiments by not relying on a fixed set of search stimuli while searching for neurons, sampling the frequency axis at a relatively high density, testing stimuli across a wide range of sound levels, and advancing the electrode in bursts with relatively long pauses (see Materials and Methods, Discussion). Because of technical limitations, we could not accurately determine spatial or laminar distributions of O and I/V units. Additional experiments are necessary to address questions regarding anatomical organization of these response types.

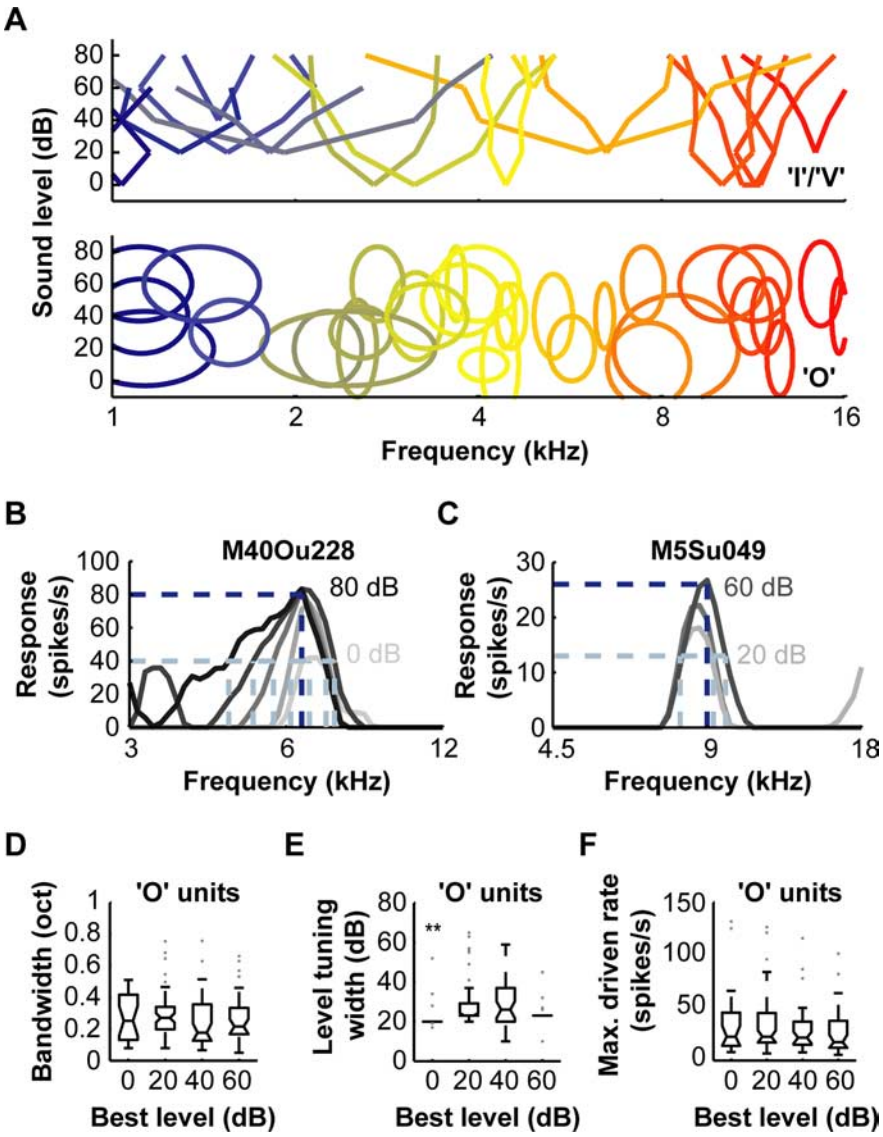

Figure 4. Invariance of O unit tuning properties with level. ***A***, Comparison of I/V and O unit coverage of the frequency and level axes. Lines (top) correspond to receptive field extents of randomly selected I/V units at half-maximal firing rate. This represents a traditional view of auditory cortex in which stimulus amplitude is represented by units with different thresholds. Frequency resolution decreases with increasing level. Ellipses (bottom) correspond to O unit receptive fields at half-maximal firing rate. This representation of stimulus space is different from the traditional view, in which amplitude in each frequency range is encoded by multiple units tuned to a smaller range of levels and frequency tuning width is independent of level (colors of lines and ellipses correspond to best frequency for clarity). ***B***, The dependence of frequency tuning on level in I/V units leads to confusion in the readout. Frequency tuning curves are plotted for an example low-threshold V unit at different levels (lines of increasing saturation correspond to tuning curves at increasing levels). At maximum firing rate (dashed blue lines), there is only one possible readout, but when the firing rate changes to a half-maximal level (dashed cyan lines), a number of possibilities covering a frequency range of ~1 octave and an amplitude range of ~80 dB must be resolved. ***C***, In the O unit population, however, this readout is much simpler. Even for an O unit responding over a 40 dB range, the number of possibilities at half-maximal rate and the spread of possible parameter values are restricted. ***D***, ***E***, Population summary of the invariance of tuning properties of O units with level. Regardless of best sound level, O units were narrowly tuned in frequency and level. ***F***, The maximum firing rate of O units in response to pure tones also did not change with sound level. (**$p < 0.01$, Wilcoxon's rank-sum test).

Level invariance can be implemented by the O unit population

Figure 4*A* illustrates the different strategies by which populations of I/V and O units cover frequency-level space. Individual O units within a given frequency range are tuned to different sound levels and completely cover the range of sound levels collectively but maintaining similar frequency and level tuning widths. I/V units, conversely, clearly exhibit a loss of frequency resolution with increasing sound level. The covariance of frequency and level tuning in V units may make it difficult to parse frequency and level solely based on firing rate. The V unit in Figure 4*B*, for example, had a low threshold (0 dB) and monotonically increasing firing rate with level. For this unit, frequency tuning curves at

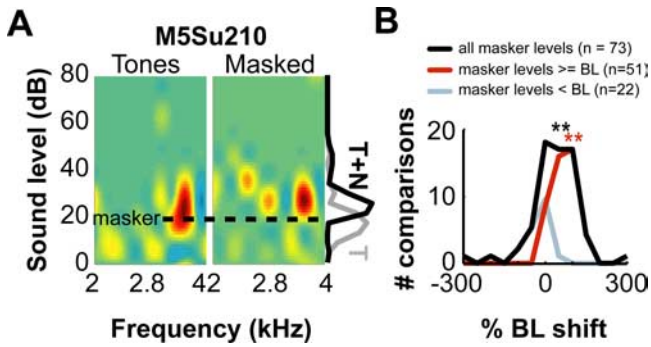

Figure 5. Effect of a continuous noise masker on O units. ***A***, When continuous masker noise (dashed line; masker level of 10 dB) was added while presenting tones to an O unit, the best level of that unit shifted while maintaining frequency and level tuning. ***B***, Over the population of strongly nonmonotonic units (black histogram; $n = 73$ comparisons from 26 neurons, 2–3 masker levels per neuron), we observed a 50% shift in best level. However, because of our low sampling resolution, it is unclear whether shifts occurred at low masker levels (cyan; $n = 22$). When these data are plotted for maskers at or louder than the best level of the unit (red; $n = 51$), the observed shift was higher (67%). This implies that the same units continue to encode sounds when there are dynamic shifts in noise conditions without altering their tuning properties (**$p < 0.01$, Wilcoxon's rank-sum test).

five levels, from 0 to 80 dB were obtained in 20 dB steps. At the maximal firing rate of the unit of ~80 spikes/s, there was no ambiguity in readout with respect to frequency. When the firing rate of this unit changed to a half-maximal level (~40 spikes/s), however, a number of possible conditions, involving frequency-level combinations over a range of 80 dB and ~1 octave had to be resolved. The number of possible frequency-level combinations to be disambiguated for a typical O unit (Fig. 4*C*), however, were much fewer and spanned a smaller parameter range than the V unit shown in Figure 4*B*. In a population-average-based readout scenario, it would require pooling of a significantly smaller number of O units to arrive at the "correct" stimulus parameters. Thus, a separable response to frequency and level is an advantageous coding strategy.

Over the population of O units, we observed that frequency tuning bandwidth (Fig. 4*D*) and level tuning width (Fig. 4*E*) did not depend on the best sound level around which the unit responded. The level tuning width at the lowest level (0 dB) was significantly smaller ($p < 0.01$, Wilcoxon's rank-sum test), but this is likely attributable to an edge effect (because we did not sample lower levels, we could observe only half of the level tuning curve). Importantly, the maximum firing rate of the units in response to pure tones also did not vary with sound level (Fig. 4*F*). This effectively normalizes the neural representation of the stimulus at all sound levels and removes the influence of sound level on firing rate. In other words, independent of mean stimulus intensity, similar tuning widths and firing rates ensure that the stimulus is processed and represented in the same way at all sound levels. This enables similar read-out mechanisms to be used by downstream neurons without dependence on sound level. Together, these data form the basis for our argument that the O unit population is level invariant.

We explored this concept further by studying pure-tone responses in the presence of a continuous wideband noise masker. Figure 5*A* shows FRAs of an O unit measured without and with the addition of the masking noise at its BL. Although this unit preferred 3.2 kHz tones presented at 10 dB in quiet (Fig. 5*A*, left), addition of noise (at 10 dB) resulted in a shift of the BL of the unit to 20 dB (Fig. 5*A*, right) without changing its frequency or level tuning widths. We measured the shift in BL caused by adding a noise masker at multiple sound levels in 26 units that remained strongly nonmonotonic in both quiet and noise (MI ≤ 0.5) (Fig. 5*B*). The addition of a noise masker at levels equal to or greater than BL of these units caused a 67% shift ($p < 0.0001$, Wilcoxon's rank-sum test) of the BL toward louder levels. Addition of the masker at levels lower than the BL of the units produced no observable shift. Because observing a small shift is dependent on sampling resolution, it is unclear whether any BL shifts occurred at low masker levels. These observations are consistent with shifts measured in anesthetized cat A1 (Phillips and Cynader, 1985). They suggest that O units are robust to dynamic shifts in environmental noise conditions. One interpretation of these data are that the population of O units is also signal-to-noise ratio (SNR) tolerant. Although large changes in SNR cause different neurons to become active, O units accommodate smaller changes by dynamically shifting their BL to account for prevailing noise conditions. We focused on O units in testing the effect of adding background noise in the present study and did not collect sufficient data to rigorously analyze threshold shifts of the I/V population. In a few examples tested, I/V units also showed threshold shifts in the presence of a noise masker, similar to shifts observed in previous studies in anesthetized cat A1 (Phillips and Cynader, 1985).

Tuning to sound level and shifts in best level with the addition of a noise masker typically generalized to stimuli more complex than tones. For example, the units in Figure 6*A* show similar best levels and level tuning widths when tested with pure tones at BF or with upward or downward lFM sweeps passing through the units BF. We typically observed that best level, level tuning width, and monotonicity index were preserved between tones and complex stimuli (data for $n = 13$ units using lFM sweeps shown in Fig. 6*B*). In a few tested units (Fig. 6*C*), similar shifts in best level were observed when a noise masker was added to lFM stimuli.

A model of level-invariant coding in A1

Figure 7 summarizes our interpretation of these data as a conceptual model. Consider the representation of pure tones or lFM sweeps by a population of entirely O or V units on the cortical surface. For pure tones at low sound levels (0 dB) (Fig. 7*A*, left panel), only O units encoding that frequency-level combination and V units with sufficiently low threshold respond. On the cortical surface, the representations of the stimulus in both populations are similar. When the sound level is increased (to 80 dB) (Fig. 7*A*, right panel), O units that responded previously to the 0 dB sound stop firing, and a new subpopulation of O units, tuned to the 80 dB sound level, are now responsive. However, the nature of the representation of the stimulus on the cortical surface by O units does not change. In comparison, many more units in the V unit population become active, including units not tuned (at threshold) to the stimulus frequency, leading to widespread activation of the cortical surface.

For time-varying stimuli such as lFM sweeps (Fig. 7*B*), a loss of temporal resolution also occurs at high sound levels. When the stimulus is presented at a low level (left panel), its representation on the cortical surface is a tightly propagating wave of activity across O units that respond at stimulus level and low threshold V units ("freeze frames" at 50 and 100 ms after stimulus onset are shown). When sound level increases, the pattern of activity is retained in the O units because of the property of level invariance. However, the neural activity becomes smeared out in frequency and time in the V population. The neuron highlighted in black, for example, responds over a 50 ms duration. It "sees" the lFM sweep before the instantaneous frequency of the stimulus has reached its BL (at threshold), simply as a result of increased band-

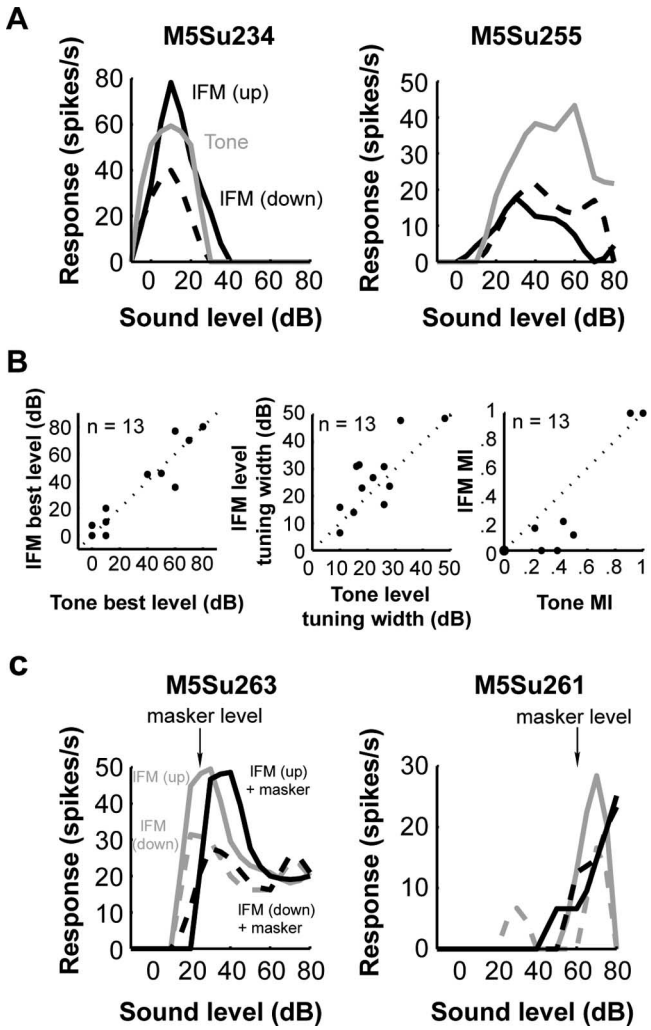

Figure 6. Level tuning typically generalized to more complex stimuli. ***A***, Two example units whose level tuning curves were similar for pure-tone and IFM stimuli. ***B***, Correlation of level tuning parameters (best level, level tuning width, and monotonicity index) derived using pure-tone and IFM stimuli measured from 13 single units. ***C***, In a few units tested, best level shifts attributable to the addition of wideband noise masker also occurred for IFM stimuli with magnitudes similar to pure-tone stimuli.

width at loud levels. Figure 7*C* illustrates this temporal smearing of response that results from increasing bandwidth. For narrowly tuned units (bottom panel), the stimulus crosses the excitatory portion of the receptive field of the neuron quickly, resulting in a temporally precise response. When frequency tuning bandwidth is broad (top panel), the stimulus stays in the excitatory receptive field for a proportionately longer duration, resulting in a response that is smeared out in time. This ultimately leads to lower spectral and temporal resolution and fundamentally changes the representation of the stimulus at the population level. For example, in a study of lFM responses in anesthetized cat A1, Heil et al. (1992) report that, at higher intensities, neural responses to FM sweeps were initiated at earlier instantaneous frequencies. Such systematic influences of level on response timing may be attributed to a widening FRA.

This conceptual model helped us visualize how robustly different unit types code simple signals (like tones). However, to generalize our observation to a broader class of sounds, measures of population activity that reflect an underlying loss of resolution in single neurons were necessary. Two such measures are the selectivity and sparseness of a population of neurons (Lehky et al., 2005). The loss of spectral and temporal resolution of V units with level described above, for example, can be thought of as a loss of selectivity (a unit responding to more stimuli) and a loss of sparseness (more units in the population responding to a single stimulus). Based on these ideas, we hypothesized that selectivity and sparseness of a population of O units would remain the same with level, whereas both parameters would decrease with level in a population of V units when more complex stimuli are presented.

To verify these predictions, we simulated a population of O and V neurons with FRA parameters drawn from the aforementioned data. For simplicity, each neuron was modeled as a linear spectral integrator (Barbour and Wang, 2003) with difference-of-Gaussian-shaped frequency tuning curves (O'Connor et al., 2005) and balanced inhibitory sidebands. O units were also tuned to a BL, whereas V units exhibited linearly increasing response strength and tuning bandwidth with sound level. Figure 8*A* shows examples of model O and V units. The addition of inhibitory sidebands and assumption of linearity are parsimonious and necessary to ensure that the model behaves realistically. For example, most A1 units in our experiments are nonresponsive to wideband noise but are responsive to band-limited noise whose bandwidth restricted to the excitatory peak. Sample responses of model neurons to commonly used stimuli are shown in Figure 8*B*.

Figure 9*A* is a snapshot of the activity of 300 O (left panel) and V (right panel) neurons, respectively, in response to 300 complex stimuli at four different sound levels. The "grayness" of each pixel indicates the response strength of a single neuron to one of the stimuli. It is evident that the fraction of gray pixels, which reflects population activity across all stimuli, remains similar across all sound levels in the O unit population (Fig. 9*A*, left). This is enabled by the level-invariant nature of stimulus representation by O units. In the V unit population, conversely, an increase in level resulted in more V units responding to more stimuli (increasing number of gray pixels with level). We quantified these observations using nonparametric measures of selectivity and sparseness (Lehky et al., 2005).

Figure 9*B* is a plot of medians and interquartile distances of these parameters for populations of O and V units at four different sound levels. Selectivity and sparseness of the O unit population were similar at all sound levels and considerably higher than V units even at moderate levels (O and V medians at 40 dB: selectivity, 11 and 3.6; sparseness, 17.5 and 1.7, respectively). Although V units matched the performance of O units at threshold (median selectivity, 13; sparseness, 36), they exhibited a systematic degradation with rising level (distributions of both parameters at all sound levels are significantly different from all other levels for V units; $p < 0.0001$, Wilcoxon's rank-sum tests). This systematic degradation could be attributed directly to the increase of frequency tuning bandwidth of model V neurons with increasing sound level.

One testable prediction that emerged from this model was the different temporal precision of O and V unit responses. Because bandwidths of V units were broader at louder sound levels, their responses were modulated at lower frequencies. This occurred because stimuli tended to stay within their excitatory receptive fields for longer durations (Fig. 7*C*). In O units, however, the combination of narrowly tuned excitation flanked by inhibition shaped the response histograms precisely. When we measured precision (Fig. 9*C*), defined here as the fraction of high-frequency power (>10 Hz) in the response histogram (see Materials and

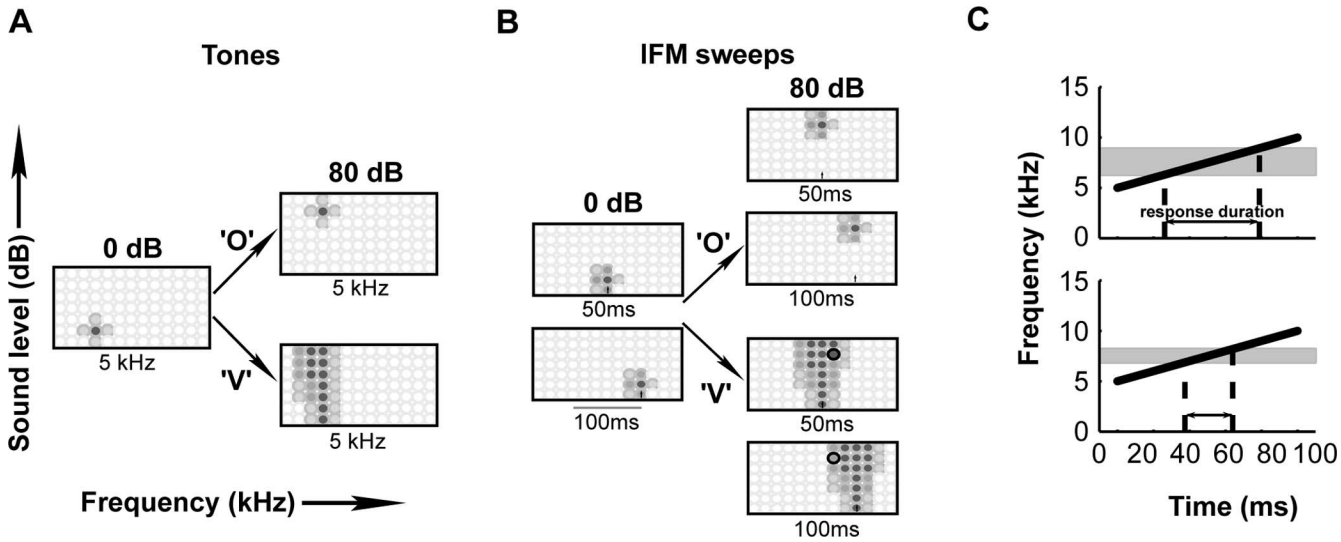

Figure 7. Conceptual model of a level-invariant representation in A1. ***A***, Diagram of a cortical sheet of neurons responding to a pure tone. Gray circles represent individual neurons, ordered by best frequency and best level, and grayscale fill represents response rate. When a pure tone (at 5 kHz, for example) is presented at a low level (0 dB; left), only O units tuned to low levels and low-threshold V units respond. In both cases, because of narrow frequency tuning, spread of activity is restricted to a small number of neurons. When sound level is increased (80 dB; right), O units tuned to this level start responding but the units tuned to low levels stop firing. The pattern of activity generated is just as restricted as low sound levels. However, activity spreads over a range of V units coding different frequencies and sound levels, leading to a loss of spectral resolution. ***B***, If a linear frequency modulated sweep (gray bar indicates spectral extent of sweep; small black arrows indicate instantaneous frequency of sweep) is presented as the stimulus at low levels (0 dB; left), a tight packet of activity propagates with the sweep across the cortical surface (snapshots of the population at 50 and 100 ms into the sweep are shown). When level is increased (80 dB; right), activity packets are just as well resolved in the O unit population. In the V unit population, there is temporal and spectral degradation. The neuron highlighted in black, for example, is active over a duration of 50 ms, starting to fire before the sweep "reaches" its BF (computed at threshold) and firing well after the sweep has crossed its BF. ***C***, Temporal precision of response is also affected by bandwidth. When an IFM sweep (black line) crosses the excitatory receptive field (shaded gray) of a narrowly tuned unit (bottom), response duration (double arrow) is short. However, for broadly tuned units (top), response is smeared out over time.

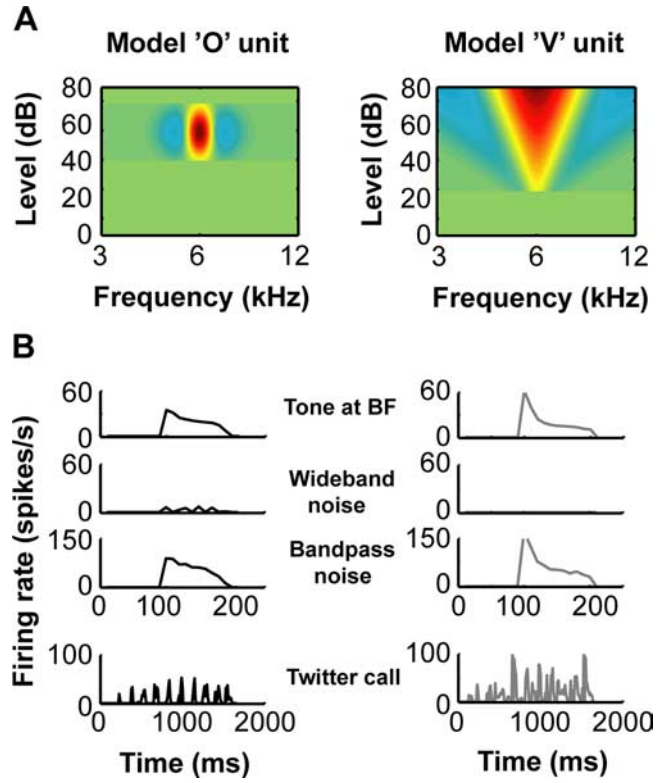

Figure 8. Model O and V units and their responses to common stimuli. ***A***, FRAs of O and V units that were modeled as spectrally linear integrators with difference of Gaussian-shaped receptive fields with response parameters drawn from our experimental data. ***B***, Model O and V unit responses to commonly used stimuli. Both types of model units were responsive to pure tones, bandpass noise (with bandwidth limited to the excitatory part of their receptive fields), and marmoset twitter calls. Neither class responded to wideband noise.

Methods), the V units showed a gradual decrease with increasing level (Fig. 9*C*, top), whereas the O units retained similar response precision at all sound levels (Fig. 9*C*, bottom). Precision of the O unit population at 0 and 60 dB were 0.7 and 0.7044, respectively, remaining relatively unchanged with varying level. However, precision of the V unit population at 0 and 60 dB were 0.8 and 0.6261, a 22% decrease over the range of sound levels tested. Precision of the V unit population decreased significantly ($p < 0.01$, Wilcoxon's rank-sum test) with each increase in level (in steps of 20 dB).

Data derived from these simulations helped us understand the activity of a population of O or V units in response to complex stimuli. Broadening of frequency tuning bandwidth with increasing level in single V units profoundly affected the nature of population activity (and, therefore, the neural representation of the stimuli) when complex stimuli were presented. These results further underscore the advantage of the level-invariant coding strategy that is implemented by the O unit population.

Discussion

The present study investigated single units in A1 of awake marmosets that were narrowly tuned to both frequency and sound level, distributed across a wide range of frequencies and sound levels. We propose based on these observations a framework for a robust representation of sounds across the large variations of intensity that are found in natural acoustic environments. Instead of thinking about these units tuned to sound level as encoding "best level" using their peak location, we propose that the O units remove the influence of sound level on frequency tuning so that variations in frequency content can now be coded in firing rate without the confounding influence of level. Whereas O units are involved in level-invariant processing, I/V units are well suited for an equally important function. The broad bandwidths

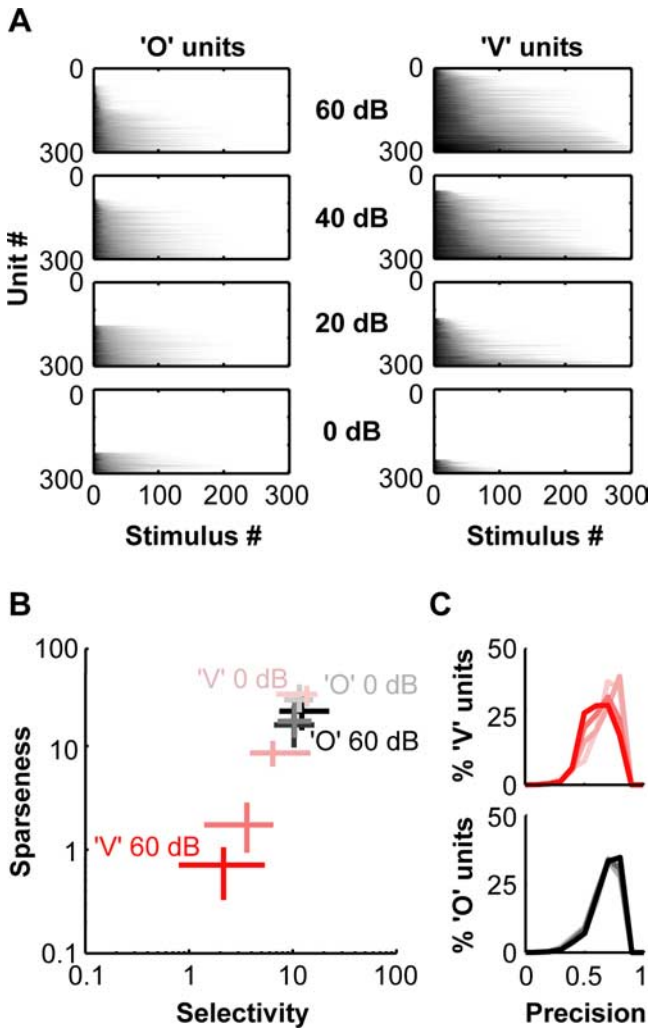

Figure 9. Simulation of O and V unit responses to test population level invariance. ***A***, Normalized responses of 300 O (left) or V (right) units to a battery of 300 complex stimuli at multiple levels. Units are sorted by selectivity and stimuli are sorted by efficacy (gray shading indicates response strength) for display clarity. Whereas similar numbers of O units responded to similar numbers of stimuli at al levels, more V units became active in response to more stimuli as level increased. ***B***, Indicators of population activity remained constant with level in the simulated O unit population but degraded in the V unit population. Selectivity and sparseness (lines are interquartile range, and intersection points are medians) of the O unit population (gray lines; black, 60 dB; lightest gray, 0 dB) remained constant, with sound level indicating level invariance of the population. The V unit population (red lines; red, 60 dB; lightest pink, 0 dB) resembled the O unit population at a low level (0 dB) but gradually lost selectivity and sparseness with increasing sound level. ***C***, This model predicts decreasing precision of V unit responses with increasing sound level as a consequence of increasing bandwidth, whereas precision of O unit responses remain constant with level.

of I/V units allow them to analyze the context in which a sound is heard over a wide parameter range and helps establish a reference point for overall sound level based on prevailing background conditions. They can thus act as "detectors" of sounds or background signals over a wide range of frequencies and levels.

In contrast to our results, previous studies in anesthetized cat (Schreiner et al., 2000; Moshitch et al., 2006) and awake primates (Recanzone et al., 2000) show that most A1 units have V-shaped FRAs. Several factors could lead to the discrepancy between the results of the present and previous studies: (1) anesthetic effects and related laminar biases, (2) coarse sampling with search stimuli, (3) search biases attributable to inherent differences in response properties between O and I/V units, and (4) potential species specificity. Most studies of anesthetized A1 were restricted to layer 4 because of the difficulty of driving layer 2/3 neurons under anesthesia, whereas the present study sampled the supragranular layers extensively in the awake marmoset. It is conceivable that the shape of the FRA is affected by anesthesia and laminar position. However, this does not explain why the study by Recanzone et al. (2000) in awake macaques did not encounter as many O units as we did.

One likely source of this discrepancy is a unit search bias. Because O units are narrowly tuned to frequency and level, experimental stimuli need to finely sample frequency and level axes. Using coarser spacing in frequency and searching only at one or two moderately loud levels would bias the experimenter toward V units because they are more widely tuned and have monotonic rate-level functions. The inherent differences in basic response properties such as the higher spontaneous and driven rates of V units make this bias toward V units stronger. This possibility is supported by the higher spontaneous and peak driven rates of neurons reported by Recanzone et al. (2000) in their study, which is comparable with those of our I/V units. We avoided these biases by searching at multiple sound levels, using a battery of search stimuli that was not restricted to tones and by depending less on stimulus-driven responses to isolate single units. A second possibility is the difference in stimulus delivery. Whereas Recanzone et al. (2000) delivered stimuli to the contralateral ear, we presented stimuli from a speaker situated directly in front of the animal.

It is unlikely that our results differ from previous results because of species specificity of this phenomenon. Neurons similar to O units are predominant in the DSCF of bats (Suga, 1977; Suga and Manabe, 1982), and high proportions of strongly nonmonotonic units have also been found in A1 of awake macaques (Brugge and Merzenich, 1973; Pfingst and O'Connor, 1981). One study reported that strongly nonmonotonic units were found more often when recording single-unit responses (28%) than multiunit responses (8%) in anesthetized cat A1 (Sutter and Schreiner, 1995). Furthermore, multiunit responses recorded from middle layers of barbiturate-anesthetized marmoset A1 showed predominantly V-shaped FRAs (X. Wang, unpublished data). Another multiunit study of the A1 in anesthetized marmosets (Philibert et al., 2005) also indicates a prevalence of V-shaped FRAs ($Q_{40\ dB} < Q_{10\ dB}$, suggesting V-shaped FRAs). Philibert et al. (2005) concluded that the functional organization of marmoset A1 is similar to that found in the owl monkey (Recanzone et al., 1999) and the squirrel monkey (Cheung et al., 2001). These studies suggest that, within the same species, different physiological states may lead to different response properties related to sound level.

Experimental data from several studies have revealed that the pattern of cortical activation as a function of level is not predicted by underlying single-unit tuning curves. For example, Heil et al. (1994) show that, in anesthetized cat A1, mean cortical activation increases with increasing intensity at low and moderate levels and saturates at high levels, whereas spatial patterns of activation along isofrequency contours continue to change at high levels. If all A1 neurons were O shaped with their thresholds and best levels equally distributed across sound levels, there would be no change in mean cortical activation with increasing sound level. The increase in mean activation observed by Heil et al. (1994) can be explained by the activity of I/V neurons that increase their firing rate with increasing level at low and moderate levels before saturating at high levels. An important difference is that the present study was based on single-units classified as either O or I/V

shaped, whereas Heil et al. (1994) study was based on multiunits recorded from middle cortical layers. It is not clear what proportions of O- and I/V-shaped units were included in the mean multiunit activity observed by Heil et al. (1994). The change in spatial patterns with increasing sound level observed by Heil et al. (1994) represents a different way of signaling changing sound level than the intensity coding mechanism proposed in the present study. Phillips et al. (1994) showed that the pattern of cortical activation bears little relationship to the threshold CF contour map in anesthetized cat A1. This may be primarily explained by the spatially unorganized occurrence of nonmonotonic units along isofrequency contours that are rendered inactive with a change in level. Another study in cat auditory cortex using consonant–vowel speech sounds found that, although spatial activation was strongly intensity dependent, the pattern of activation was patchy (Wong and Schreiner, 2003). Our results and our proposed model partially concur with these studies. The small population of monotonic I/V units might account for the increase in mean spatial activation, with "patchiness" being a consequence of different spatially unorganized O units firing at different sound levels.

Another interesting issue is how O-shaped FRAs are generated. It is as yet unclear to what extent O-shaped FRAs are created at A1 or inherited from subcortical stations. For example, a large proportion of nonmonotonic units have been found in the auditory thalamus (Rouiller et al., 1983). Alternatively, the evolution of V unit responses over time to O unit-like responses (Fig. 3*B*,*C*) might point to a cortical inhibition source as a likely candidate for shaping O-like responses. For example, intracellular data from nonmonotonic units in anesthetized cats revealed strong inhibition at loud sound levels (Ojima and Murakami, 2002). A more recent study demonstrated that unbalanced synaptic inhibition, likely of cortical origin, underlies intensity tuning in rat A1 (Tan et al., 2007). A recent study found that behavioral training could cause over-representation of target-specific sound levels in the auditory cortex of rats. This shows that the best level of cortical units is amenable to top-down influences, again suggesting a strong cortical influence on level tuning (Polley et al., 2006). Another likely possibility is a combination of both factors. Namely, thalamocortical inputs that are weakly nonmonotonic may be sharpened further by cortical inhibition, leading to the strongly level-tuned neurons that we observe.

It is also important to ask why O units are necessary for auditory processing and how they fit into the processing hierarchy. Many studies suggest that receptive field structure reflects an efficient representation of stimulus space (Simoncelli, 2003; Smith and Lewicki, 2006). Marmosets make both loud (phee calls) and quiet (trill calls) vocalizations that are highly tonal in nature (Agamaite, 1997; DiMattina and Wang, 2006). One might hear the same sound in a variety of listening conditions, such as distance from source, occlusion by intervening objects, and changing ambient noise conditions. Variability in all these conditions determines the effective intensity at which the sound is heard, but the quality of the sound remains unaffected. O units may reflect an internal representation of this statistical independence of frequency and level, exploiting this property to code sounds efficiently. Note that we are not proposing that the encoding of absolute sound level is not important in A1. Rather, this could be accomplished at subcortical processing stations or by the population of I/V units in A1. We only argue that implementing level-invariant coding at an early stage of cortical processing is a highly desirable property that can simplify computational goals of the auditory cortex, such as feature or object recognition. Level invariance at the population level among the O units in A1 might provide information to cortical areas higher in the auditory hierarchy in which level invariance can be seen in single neurons. For example, a recent study found scale-invariant multiunit responses in the auditory cortex of anesthetized bats (Firzlaff et al., 2007). In the context of echolocation, object scale is associated with changes in amplitude and delay of the reflected echolocation pulse. A level-invariant representation of such sounds could be a potential first stage in achieving this scale invariance.

It is unclear how the sound level invariance proposed here compares mechanistically with invariance to intensity dimensions in other modalities. This is attributable to the fact that the tuning to stimulus intensity observed in A1 has not been observed in other sensory cortices. For example, neural response rates increase monotonically as a function of stimulus indentation in primary somatosensory cortex (Wang et al., 1995), as a function of whisker deflection velocity in barrel cortex (Wilent and Contreras, 2004), and as a function of stimulus contrast in primary visual cortex (Skottun et al., 1987). Even neurons in higher cortical areas such as macaque area V4 respond with monotonically increasing rates as stimulus contrast increases (Williford and Maunsell, 2006). Therefore, the tuning to sound level exhibited by A1 units and the mechanisms by which intensity invariance is implemented by these units is possibly unique to auditory processing.

References

Agamaite JA (1997) A quantitative characterization of the vocal repertoire of the common marmoset. Master's thesis, Johns Hopkins University.

Barbour DL, Wang X (2003) Auditory cortical responses elicited in awake primates by random spectral stimuli. J Neurosci 23:7194–7206.

Brincat SL, Connor CE (2006) Dynamic shape synthesis in posterior inferotemporal cortex. Neuron 49:17–24.

Brugge JF, Merzenich MM (1973) Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. J Neurophysiol 36:1138–1158.

Calford MB, Webster WR, Semple MM (1983) Measurement of frequency selectivity of single neurons in the central auditory pathway. Hear Res 11:395–401.

Cheung SW, Bedenbaugh PH, Nagarajan SS, Schreiner CE (2001) Functional organization of squirrel monkey primary auditory cortex: responses to pure tones. J Neurophysiol 85:1732–1749.

Davis KA (2005) Spectral processing in the inferior colliculus. Int Rev Neurobiol 70:169–205.

DiMattina C, Wang X (2006) Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations. J Neurophysiol 95:1244–1262.

Firzlaff U, Schuchmann M, Grunwald JE, Schuller G, Wiegrebe L (2007) Object-oriented echo perception and cortical representation in echolocating bats. PLoS Biol 5:1174–1183.

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

Heil P, Rajan R, Irvine DRF (1994) Topographic representation of tone intensity along the isofrequency axis of cat primary auditory cortex. Hear Res 76:188–202.

Hirsh IJ, Reynolds EG, Joseph M (1954) Intelligibility of different speech materials. J Acoust Soc Am 26:530–538.

Kadia SC, Wang X (2003) Spectral integration in A1 of awake primates: neurons with single- and multipeaked tuning characteristics. J Neurophysiol 89:1603–1622.

Kiang NY (1965) Discharge patterns of single fibers in the cat's auditory nerve. Cambridge, MA: MIT.

Kilgard MP, Merzenich MM (1999) Distributed representation of spectral and temporal information in rat primary auditory cortex. Hear Res 134:16–28.

LeBeau FEN, Malmierca MS, Rees A (2001) Iontophoresis *in vivo* demonstrates a key role for $GABA_A$ and glycinergic inhibition in shaping fre-

quency response areas in the inferior colliculus of guinea pig. J Neurosci 21:7303–7312.

Lehky SR, Sejnowski TJ, Desimone R (2005) Selectivity and sparseness in the responses of striate complex cells. Vision Res 45:57–73.

Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM (2003) Spectrotemporal structure of receptive fields in areas A1 and AAF of mouse auditory cortex. J Neurophysiol 90:2660–2675.

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

Malone BJ, Kumar VR, Ringach DL (2007) Dynamics of receptive field size in primary visual cortex. J Neurophysiol 97:407–414.

Moshitch D, Las L, Ulanovsky N, Bar-Yosef O, Nelken I (2006) Responses of neurons in primary auditory cortex (A1) to pure tones in the halothane-anesthetized cat. J Neurophysiol 95:3756–3769.

Nomoto M, Suga N, Katsuki Y (1964) Discharge pattern and inhibition of primary auditory nerve fibers in the monkey. J Neurophysiol 27:768–787.

O'Connor KN, Petkov CI, Sutter ML (2005) Adaptive stimulus optimization for auditory cortical neurons. J Neurophysiol 94:4051–4067.

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

Pack CC, Berezovskii VK, Born RT (2001) Dynamic properties of neurons in cortical area MT in alert and anesthetized macaque monkeys. Nature 414:905–908.

Pfingst BE, O'Connor TA (1981) Characteristics of neurons in auditory cortex of monkeys performing a simple auditory task. J Neurophysiol 45:16–34.

Philibert B, Beitel RE, Nagarajan SS, Bonham BH, Schreiner CE, Cheung SW (2005) Functional organization and hemispheric comparison of primary auditory cortex in the common marmoset (*Callithrix jacchus*). J Comp Neurol 487:391–406.

Phillips DP, Cynader MS (1985) Some neural mechanisms in the cat's auditory cortex underlying sensitivity to combined tone and wide-spectrum noise stimuli. Hear Res 18:87–102.

Phillips DP, Irvine DR (1981) Responses of single neurons in physiologically defined primary auditory cortex (A1) of the cat: frequency tuning and responses to intensity. J Neurophysiol 45:48–58.

Phillips DP, Orman SS (1984) Responses of single neurons in posterior field of cat auditory cortex to tonal stimulation. J Neurophysiol 51:147–163.

Phillips DP, Semple MN, Calford MB, Kitzes LM (1994) Level-dependent representation of stimulus frequency in cat primary auditory cortex. Exp Brain Res 102:210–226.

Polley DB, Steinberg EE, Merzenich MM (2006) Perceptual learning directs auditory cortical map reorganization through top-down influences. J Neurosci 26:4970–4982.

Popper AN, Fay RR, eds (1992) The mammalian auditory pathway: neurophysiology. New York: Springer.

Ramachandran R, Davis KA, May BJ (1999) Single-unit responses in the inferior colliculus of the decerebrate cats. I. Classification based on frequency response maps. J Neurophysiol 82:152–163.

Recanzone GH, Schreiner CE, Sutter ML, Beitel RE, Merzenich MM (1999) Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey. J Comp Neurol 415:460–481.

Recanzone GH, Guard DC, Phan ML (2000) Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. J Neurophysiol 83:2315–2331.

Rose JE, Greenwood DD, Goldberg JM, Hind JE (1963) Some discharge characteristics of single neurons in the inferior colliculus of the cat. I. Tonotopical organization, relation of spike counts to tone intensity, and firing patterns of single elements. J Neurophysiol 26:294–320.

Rouiller E, de Ribaupierre Y, Morel E, de Ribaupierre F (1983) Intensity functions of single unit responses to tone in the medial geniculate body of cat. Hear Res 11:235–247.

Schreiner CE, Read HL, Sutter ML (2000) Modular organization of frequency integration in primary auditory cortex. Annu Rev Neurosci 23:501–529.

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

Simoncelli EP (2003) Vision and the statistics of the visual environment. Curr Opin Neurobiol 13:144–149.

Skottun BC, Bradley A, Sclar G, Ohzawa I, Freeman RD (1987) The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. J Neurophysiol 57:773–786.

Smith EC, Lewicki MS (2006) Efficient auditory coding. Nature 439:978–982.

Spirou GA, Young ED (1991) Organization of dorsal cochlear nucleus type IV unit response maps and their relationship to activation by bandlimited noise. J Neurophysiol 66:1750–1768.

Studebaker GA, Gilmore C, Sherbecoe RL (1993) Performance-intensity functions at absolute and masked thresholds. J Acoust Soc Am 93:3418–3421.

Studebaker GA, Sherbecoe RL, McDaniel MD, Gwaltney CA (1999) Monosyllabic word recognition at higher-than-normal speech and noise levels. J Acoust Soc Am 105:2431–2444.

Suga N (1965a) Analysis of frequency-modulated sounds by auditory neurones of echo-locating bats. J Physiol (Lond) 179:26–53.

Suga N (1965b) Functional properties of auditory neurones in the cortex of echo-locating bats. J Physiol (Lond) 181:671–700.

Suga N (1969) Classification of inferior collicular neurones of bats in terms of responses to pure tones, frequency-modulated sounds and noise bursts. J Physiol (Lond) 200:555–574.

Suga N (1977) Amplitude spectrum representation in the doppler-shifed-CF processing area of the auditory cortex of the mustached bat. Science 196:64–67.

Suga N (1992) Philosophy and stimulus design for neuroethology of complex-sound processing. Phil Trans R Soc Lond B Biol Sci 336:423–428.

Suga N, Manabe T (1982) Neural basis of amplitude-spectrum representation in the auditory cortex of the mustached bat. J Neurophysiol 47:225–255.

Sutter ML, Schreiner CE (1995) Topography of intensity tuning in cat primary auditory cortex: single-neuron versus multi-neuron recordings. J Neurophysiol 73:190–204.

Sutter ML (2000) Shapes and level tolerances of frequency tuning curves in primary auditory cortex: quantitative measures and population codes. J Neurophysiol 84:1012–1025.

Tan AY, Zhang LI, Merzenich MM, Schreiner CE (2004) Tone-evoked excitatory and inhibitory synaptic conductances of primary auditory cortex neurons. J Neurophysiol 92:630–643.

Tan AY, Atencio CA, Polley DB, Merzenich MM, Schreiner CE (2007) Unbalanced synaptic inhibition can create intensity-tuned auditory cortex neurons. Neuroscience 146:449–462.

Wang X, Merzenich MM, Sameshima K, Jenkins WM (1995) Remodeling of hand representation in adult cortex determined by timing of tactile stimulation. Nature 378:71–75.

Wang X, Lu T, Snider RK, Liang L (2005) Sustained firing in auditory cortex evoked by preferred stimuli. Nature 435:341–346.

Wilent WB, Contreras D (2004) Synaptic responses to whisker deflection in rat barrel cortex as a function of cortical layer and stimulus intensity. J Neurosci 24:3985–3998.

Williford T, Maunsell JH (2006) Effects of spatial attention on contrast response functions in macaque area V4. J Neurophysiol 96:40–54.

Wong SW, Schreiner CE (2003) Representation of CV-sounds in cat primary auditory cortex: intensity dependence. Speech Commun 41:93–106.

Yang L, Pollack GD, Resler C (1992) GABAergic circuits sharpen tuning curves and modify response properties in the mustache bat inferior colliculus. J Neurophysiol 68:1760–1774.