Receptive field dimensionality increases from the auditory midbrain to cortex

Craig A. Atencio,1,2 Tatyana O. Sharpee,3 and Christoph E. Schreiner1,2

1The UCSF Center for Integrative Neuroscience and 2Coleman Memorial Laboratory, Department of Otolaryngology-Head and Neck Surgery, University of California, San Francisco; and 3The Salk Institute for Biological Studies and the Center for Theoretical Biological Physics, University of California, San Diego, La Jolla, California

Submitted 8 November 2011; accepted in final form 8 February 2012

FOR AUDITORY NEURONS, the spectrotemporal receptive field (STRF) is a complete description of the stimulus preferences of a neuron. The STRF may be approximated as a set of linear filters, or components, that describe the spectrotemporal stimulus selectivity of auditory neurons. The number of filters represents the dimensionality of the STRF (Sharpee et al. 2004). In all layers of primary auditory cortex (AI), STRFs have been shown to have multiple components that capture concurrent processing of disparate stimulus aspects (Atencio et al. 2008, 2009). In contrast, primary visual cortex granular layers predominantly contain simple cells, which can be well described with a single stimulus filter, while extragranular layers are dominated by complex cells, which can only be described with multiple filters. Since the number of filters changes within the visual cortical hierarchy, this indicates that multicomponent receptive fields are constructed within cortex itself. Within AI, however, the situation is different; since neurons in all layers of AI have STRFs with multiple components, it is unknown whether this concurrent, multidimensional processing emerges in AI or whether it is inherited from subcortical stations.

The auditory forebrain comprises the cortex and the thalamus. The midbrain is positioned prior to the forebrain, and it is an obligatory station (Ehret 1997). Signals proceed along thelemniscal pathway from the midbrain, to the thalamus, and then to cortex. Midbrain STRFs have been extensively characterized with single-filter descriptions (Andoni et al. 2007; Escabi and Schreiner 2002; Lesica and Grothe 2008; Rodriguez et al. 2010). It is unknown whether midbrain STRFs can be more completely characterized by using more than one filter.

STRFs can be conceptualized as describing the stimulus filtering that a neuron performs. One approach to estimating STRFs is to calculate the spike-triggered average (STA). The STA is a single filter that may be used to describe spectrottemporal processing (Andoni et al. 2007; Escabi and Schreiner 2002). If the processing of the cell can be captured by one filter, then the STA may be a sufficient descriptor (de Boer and Kuyper 1968). If more than one filter is required, as in visual and auditory cortex (Atencio et al. 2008; Chen et al. 2007; Rust et al. 2005; Touryan et al. 2002), then this model is inadequate. In the auditory midbrain, the main station is the central nucleus of the inferior colliculus (ICC). For the ICC, a single linear filter, in conjunction with a static nonlinearity, may be used to describe some aspects of processing (Andoni et al. 2007; Lesica and Grothe 2008). However, without knowing the number of filters, or stimulus dimensions, that are required to model midbrain function, the adequacy of the model may not be discerned.

Here, to address these questions, we systematically examine spectrottemporal filters in the auditory midbrain. We first address the adequacy of the STA as a functional descriptor for ICC STRFs. Next, we estimate the number of filters that are needed to capture midbrain spectrottemporal processing. We then compare the spectrottemporal processing of midbrain neurons to auditory cortical neurons. This is necessary since more than one filter is present in granular and extragranular layers of auditory cortex, and the demonstration of the existence of multiple filters in earlier stations would constrain the sites of origins of this important processing principle. We conclude by demonstrating the feasibility of spike-triggered analyses for a continuum of ICC temporal response types, from phasic to tonic. Our results reveal that spectrottemporal stimulus encoding undergoes a fundamental transformation between midbrain and forebrain. This transformation is reflected in the emergence of multidimensional receptive fields beyond the auditory midbrain.

METHODS

Electrophysiology. Electrophysiological methods and stimulus design were similar to previous reports (Escabi and Schreiner 2002; Schreiner and Langner 1997). Young adult cats were given an initial dose of ketamine (22 mg/kg) and acepromazine (0.11 mg/kg), and anesthesia was maintained with pentobarbital sodium (Nembutal, 15–30 mg/kg) during the surgical procedure. The animal’s temperature was maintained with a thermostatic heating pad. A custom head...
holder was used to stabilize the head. Bupivacaine was applied to incision points. Surgery consisted of a tracheostomy, reflection of the soft tissues of the scalp, craniotomy over cortex, and durotomy. The cortex posterior to auditory cortex, and above the ICC, was then resected. After surgery, the animal was maintained in an anaesthetized state with a continuous infusion of ketamine-diazepam (2–10 mg·kg\(^{-1}\)·h\(^{-1}\) ketamine, 0.05–0.2 mg·kg\(^{-1}\)·h\(^{-1}\) diazepam in lactated Ringer solution). All procedures were in strict accordance with, and were administered under an experimental protocol approved by, the University of California of San Francisco Committee for Animal Research.

All recordings were made with the animal in a sound-shielded anechoic chamber (IAC, Bronx, NY), with stimuli delivered via a closed speaker system (diaphragms from Stax) that delivered sound through a hollow ear bar that was inserted into the ear canal contralateral to the recording site. Simultaneous extracellular recordings were made with multichannel gold-plated silicon recording probes (kindly provided by the University of Michigan Center for Neural Communication Technology). The probes contained 16 linearly spaced recording channels, with each channel separated by 0.15 mm. The impedance of each channel was 4–5 M\(\Omega\). Probes were positioned orthogonally with a microdrive (David Kopf Instruments, Tujunga, CA).

Neural traces were band-pass filtered between 600 and 6,000 Hz and were recorded to disk with a Neurolyx Cheetah A/D system at sampling rates between 18 and 27 kHz. The traces were sorted off-line for single units with a Bayesian spike-sorting algorithm (Lewicki 1994).

After spike sorting and analysis, neurons were determined to be in the central nucleus of the ICC by analyzing responses to pure tones and ripple stimuli (Merzenich and Reid 1974; Schreiner and Langner 1997). The ICC was identified with physiological criteria: As recording depth increased, there was an accompanying increase in characteristic frequency with pure tones and in best frequency with STRFs (see below for analysis procedures).

Stimulus. For any recording position, neurons were probed with pure tones and then with a 20-min dynamic moving ripple (DMR) stimulus. Pure tones were 50 ms in duration, had 5 ms cosine-squared onset/offset ramps, and were presented every 300 ms. The frequency and level of each tone were chosen from 45 frequencies and 15 levels, resulting in a set of 675 different tones. The set was presented in pseudorandom sequence (Schreiner and Sutter 1992). Level steps were 5 dB and covered a range of 70 dB. Frequencies were logarithmically spaced and encompassed the frequency preferences of encountered neurons. For the DMR stimulus, the maximum spectral modulation frequency was 4 cyc/oct, and the maximum temporal modulation frequency was 500 cyc/s (Escabí and Schreiner 2002). Maximum modulation depth of the spectrotemporal envelope was 40 dB. Mean intensity was set at 30–50 dB above the mean pure-tone threshold.

Analysis. Data analysis was carried out in MATLAB (MathWorks, Natick, MA). We used the reverse correlation method to derive the average spectrotemporal stimulus envelope preceding a spike (STA) (Aertsen and Johannesma 1981; deCharms et al. 1998; Escabí and Schreiner 2002). To obtain the maximally informative dimensions (MIDs), we followed previously reported methodologies (Sharpee et al. 2004, 2006). To find relevant stimulus dimensions, we searched through the stimulus space for those dimensions that maximized the mutual information between the stimulus and the spiking response. The first MID (MID1) is the direction in stimulus space that maximizes the mutual information between the stimulus and the response. The second MID (MID2) was then found as the dimension in the stimulus space that, together with the first MID, further maximized the information. The mutual information between projections onto individual filters, \(v\), and single spikes was computed according to

\[
I(v) = \int \mathrm{d}x P_s(x) \log_2 \frac{P(v|x)}{P(v)}
\]

where \(x\) represents projections onto the relevant dimension, \(v\), \(P_s(x)\) is the distribution of projections for all presented stimuli, \(P(v|x)\) is the distribution of projections for only those stimuli that led to a spike. The filter \(v\) was either MID1 or MID2. The one-dimensional input/output nonlinearity was calculated via

\[
P(v|x) = \frac{P(x|v)}{P(x)}
\]

where \(P(x)\) is the average firing rate of the neuron.

The mutual information between single spikes and both MIDs was calculated as

\[
I(MID1, MID2) = \int \int \mathrm{d}x_1 \mathrm{d}x_2 P(x_1, x_2) \log_2 \frac{P(x_1, x_2|v)}{P(x_1, x_2)}
\]

where \(x_1\) and \(x_2\) represent the projections of the stimulus onto the first and second MIDs, respectively. The two-dimensional nonlinearity was calculated via

\[
P(v|x_1, x_2) = \frac{P(x_1, x_2|v)}{P(x_1, x_2)}
\]

All estimates of relevant stimulus dimensions (STA, MID1, MID2) were computed as an average of four jackknife estimates. Each jackknife estimate was computed by using a different \(\frac{3}{4}\) of the data (the training data set), and thus leaving a different \(\frac{1}{4}\) of the data as a test data set. Information values were calculated using different fractions of the test data set for each neuron. To accomplish this, the information values were calculated over the first 80%, 90%, 92.5%, 95%, 97.5%, and 100% of the test data set. The information calculated from these data fractions was plotted against the inverse of the data fraction percentage (1/80, 1/90, etc.). We extrapolated the information values to infinite data set size by fitting a line to the plot and taking the \(y\)-axis intercept as the information value for unlimited data size (Brenner et al. 2000; Strong et al. 1998; Treves and Panzeri 1995).

From the STA and MID1, we obtained the latency, best frequency (BF), bandwidth (BW), and spectral tuning. To obtain the parameters, the filter was summed across time or frequency, which produced the frequency or time marginals, respectively. The time or frequency axis for each marginal was then upsampled 500 times (using MATLAB function linspace.m), and the marginals were interpolated (using MATLAB function interp1.m with “spline” option). Latency was defined as the peak in the time marginal. BF was defined as the peak in the frequency marginal, and BW was the width of the distribution at 25% of the peak height. Spectral tuning, defined as the quality factor \(Q\), was calculated from the BF and BW via \(Q = \text{BW}/\text{BF}\).

The similarity between filters was calculated with the Pearson correlation coefficient, \(r\), defined as

\[
r = \frac{\sum \sum A(i,j)B(i,j)}{\sqrt{\sum \sum A(i,j)A(i,j)} \sqrt{\sum \sum B(i,j)B(i,j)}}
\]

where \(A\) and \(B\) are matrices and may be either the STA or MID1. The similarity ranges between +1 and −1 and is a measure of the spectrotemporal correlation between the two filters.

The temporal evolution of pure-tone responses was evaluated with the phasic- tonic index (PTI):

\[
\text{PTI} = N_p/(N_B + N_E) = N_p/N_T
\]

where \(N_p\) is the number of spikes during the first half of the stimulus, \(N_E\) is the number of spikes during the last half of the stimulus, and \(N_T\)
is the sum of $N_t$ and $N_p$. For purely phasic neurons the PTI is 0, while for tonic responses the PTI is $-0.5$. The PTI includes data from all pure-tone presentations; thus it may include responses that are phasic at low sound levels or at the edge of a neuron’s frequency response area. In this manner, the PTI is a conservative measure.

The MID1 contribution was defined as $100·(\text{MID1})/(\text{MID1} + \text{MID2})$. The STA and MID1 were compared using the STA sufficiency metric, defined as $100·(\text{STA})/(\text{MID1})$. For both measures we used the information values from the extrapola projection.

We also followed previous work and computed a response precision index (RPI) for each neuron using the relation $\text{RPI} = \sqrt{\text{max}^2 - \text{min}^2}/\text{avg}$, where max(STA) and min(STA) are the maximum and minimum values in the STA and avg is the average firing rate (Escabí and Schreiner 2002). The RPI measures the temporal precision of a neuron by comparing the difference between the maximum and minimum in the estimated STA to the expected theoretical difference for stimuli that are perfectly aligned (Escabí and Schreiner 2002). The RPI ranges from 0 (not temporally precise) to 1 (temporally precise).

RESULTS

The ICC is an obligatory computational hub in the central auditory system (Aitkin 1986; Ehret 1997). Information from brain stem nuclei, such as the superior olivary complex and cochlear nucleus, converges onto the ICC (Oliver 2005; Schofield 2005). The projections to the ICC are localized into specific topographic domains (Malmierca et al. 2005; Oliver 2000; Oliver et al. 1995; Schofield 2005). Although our knowledge of these projection patterns is extensive, our understanding of ICC spectrotemporal processing is less comprehensive.

Our goal in this study was to estimate ICC STRFs. To characterize ICC neurons, we used two types of stimuli: pure tones and a dynamic, broadband sound. In the ICC, the temporal response pattern to pure tones fell along a gradient, from tonic to phasic. For tonic responses, vigorous discharges occurred throughout the stimulus duration (Fig. 1).

The responses can be more fully appreciated with the post-stimulus time histogram (PSTH). The PSTH was estimated by including responses to every pure-tone frequency-intensity combination. For predominantly tonic cells, the response duration was tightly locked to the pure-tone duration (Fig. 1; pure-tone duration of 50 ms indicated by gray boxes). Additionally, tonic neurons often had few discharges after 150 ms, which is 100 ms after tone offset, further indicating effective stimulus entrainment (Fig. 1, A and B).

Tonic responding neurons responded vigorously to the DMR. Although the DMR has no well-defined onset, tonic neurons exhibited high discharge rates in response to this continuous stimulus. Spike rates in response to the DMR often exceeded 20 spikes/s [median = 12.5, median absolute deviation (m.a.d.) = 9.4]. Despite these high rates, we were able to obtain the STA from DMRs for tonic responding neurons (Fig. 1).

A wide variety of temporal responses were found in the midbrain. Responses could be highly tonic with little background activity (Fig. 1, A–C); they could be tonic with much background activity (Fig. 1, D and E); in some cases they were onset with a sustained component (Fig. 1F); and in other cases responses were highly phasic (Fig. 1G). In all cases, we were able to estimate well-defined STAs (Fig. 1). Thus the phasic-tonic nature of the responses had no bearing on our ability to utilize spike-triggered averaging techniques.

The temporal response type of a neuron was not directly related to the structure of the STA. For tonic responses, diverse STA patterns were observed: excitation present without inhibition (Fig. 1A; excitation: red subfields; inhibition: blue subfields); excitation followed by inhibition (Fig. 1D); excitation flanked by inhibition in time and frequency (Fig. 1E); or excitation accompanied by strong spectral sidebands (Fig. 1F). Phasic neurons similarly exhibited temporal and spectral inhibitory sidebands (Fig. 1G). The overall impact of temporal response profiles on STRFs is quantified further below.

Single-dimensional STRFs. The receptive field estimated by the STA represents a single feature dimension in the spectrotemporal stimulus space. The STA captures the average stimulus feature that accounts for a neuron’s response. Another approach that may be used to obtain the STRF is MID analysis.

In this framework, an STRF, or linear filter, is estimated by maximizing the mutual information between the stimulus and the spiking response (Atencio et al. 2008, 2009; Sharpee et al. 2004). The filter is calculated by an iterative process, and the result is the stimulus feature that best describes the acoustic pattern that drives the neuron. The resulting filter, which accounts for the most mutual information, is the first MID (MID1). MID1 is an unbiased estimate since the iteration process removes the effects of stimulus correlations, which can confound STRF estimates when the stimulus ensemble contains higher-order correlations (Sharpee et al. 2004). For every neuron with an STA we were also able to calculate an MID1 (Fig. 2C). The STAs and MID1s were very similar, indicating that both approaches are adequate descriptors of a linear, single-dimensional STRF (Fig. 2, A and C).

The MID analysis may be expanded so that we can estimate the potential contributions of a second, independent filter, MID2. The presence of an MID2 further maximized the mutual information between the stimulus and the response (Fig. 2E). MID2s in the ICC, however, hardly increased the total information captured by the combined application of the MIDs and seldom revealed structured excitatory or inhibitory subfields (see below).

STA and MID1 example nonlinearities. For each spectrotemporal filter, a nonlinear function may be calculated that describes the firing rate of the neuron as a function of the similarity between the stimulus and the filter. This static, time-independent function is termed a nonlinearity. This nonlinearity represents a rule that describes how the neuron will respond, given the similarity between the stimulus and the filter. STA (Fig. 2B) and MID1 (Fig. 2D) nonlinearities were similar, with each having an asymmetric structure. Thus, when the stimulus is negatively correlated with the filter, the response of the neuron decreases below the mean driven firing rate (Fig. 2; dashed lines indicate mean rate). As the similarity increases, the response strength of the neuron increases. For some nonlinearities, the response rate plateaus (Fig. 2, 3rd row), although for most the response is highly monotonic. MID2 nonlinearities were less stereotyped in structure, and often did not deviate from the mean driven response rate (Fig. 2F).

To estimate how well the STA approximates MID1, the STA and MID1 filters and nonlinearities may be compared. We first compared basic parameters from each filter. The response...
latency was highly similar for the STA and MID1 (Fig. 3A; $r = 0.995$, $P < 0.001$, t-test). BF was even more highly correlated (Fig. 3B; $r = 0.999$, $P < 0.001$, t-test). Additionally, the BW of each filter was matched (Fig. 3C; $r = 0.982$, $P < 0.001$, t-test). Finally, the spectral tuning of each filter, defined as the quality factor $Q = \text{BW/ BF}$, was also significantly correlated across STA and MID1 filters (Fig. 3D; $r = 0.975$, $P < 0.001$, t-test).

For the STA and MID1, we also compared the global structure of the filters and nonlinearities. The correlation between the filter structure was never below 0.8 (Fig. 3E; median $= 0.948$, m.a.d. $= 0.016$), further indicating a strong agreement between MID1 and the STA. The correlation between the nonlinearities was also high (Fig. 3F; median $= 0.905$, m.a.d. $= 0.087$), implying that the STA is a sufficient approximation to MID1.

The performance of the STA and the MIDs may be estimated by calculating the mutual information. We measured the performance in the midbrain and then compared this to the performance in AI, using previously published data (Atencio et al. 2008). We used mutual information because it is an objective measure of the stimulus-response relationship that has been used extensively to evaluate neural processing. For midbrain neurons, the information provided by MID1 was highly similar to that provided by the STA, indicating that the performance of each STRF model was nearly indistinguishable (Fig. 4A; $P = 0.481$, rank sum test). By contrast, for cortical
neurons, there was much more scatter between the two estimates, and the STA information was often less than the MID1 information ($P < 0.001$, rank sum test). The ratio of the STA information to the MID1 information is the STA sufficiency. In the midbrain the STA sufficiency was often ~80%, while in the cortex the distribution was not dominated by such high values (Fig. 4B).

**Multidimensional STRFs.** One of the main advantages of the MID analysis is that a second filter, MID2, may be estimated (Atencio et al. 2008, 2009). This second MID defines an additional stimulus-based dimension that, in conjunction with MID1, further maximizes the mutual information between the stimulus and the response (Fig. 2E).

MID2s in ICC neurons seldom had discernible structures, and combining both filters did not add substantial explanatory
power over either the STA or the MID1 single-dimensional filter model.

Comparing the MID1 information values to the joint MID1 and MID2 information allows us to estimate the dimensionality of the STRFs in the ICC and the cortex (Fig. 4C). If the joint information is significantly greater than the MID1 information, then more than one MID, or stimulus feature dimension, is required to adequately describe ICC processing. Alternatively, when the MID1 information is approximately equal to the joint MID1 and MID2 information, then the second filter, MID2, provides little additional information. For midbrain neurons, the MID1 information was statistically indistinguishable from the joint MID1 and MID2 information (Fig. 4C; P = 0.594, rank sum test). By contrast, in the cortex, the MID1 information was significantly less than the joint filter information (P < 0.001, rank sum test).

The ratio of the MID1 information to the joint MID1 and MID2 information may be formed, and is termed the MID1 contribution (Fig. 4D). Across ICC neurons, MID1 accounted for >80% of the information in a two-MID model; this indicates that MID1 by itself may be quite adequate to describe the spectrotemporal processing of ICC neurons. The MID1 contribution fell below 75% for only one neuron (Fig. 4D). This result is in striking contrast to the much lower contribution values reported for AI (Atencio et al. 2008, 2009). Thus, for all neurons we encountered, the processing in the ICC may be accounted for by a single filter, while in the auditory cortex more than one filter may be required.

Response measures. One of the goals of our study was to determine the information processing of the different temporal response types found in the ICC. To characterize the responses of ICC neurons and their impact on STRF estimation, we utilized two metrics from the DMR responses and one from the pure-tone responses. The first DMR metric was firing rate (the number of spikes divided by the duration of the DMR). We encountered a broad range of firing rates, ranging up to 70 spikes/s over the duration of the DMR (Fig. 5A). These values are consistent with previous definitions of sustained responses in the auditory cortex: a firing rate >5 spikes/s during both the first and the second half of the stimulus (Wang et al. 2005). The second DMR metric was the RPI, which measures a neuron’s temporal precision (see METHODS). RPIs were distributed over a broad range (Fig. 5B; median = 0.114, m.a.d. = 0.049), which was consonant with previously described midbrain properties (Escabi et al. 2005).

The temporal evolution of pure-tone responses was evaluated with the PTI. For purely phasic neurons the PTI is 0, while for tonic responses the PTI is ~0.5. The responses of ICC neurons were distributed across a broad range of PTI values (Fig. 5C; median = 0.426, m.a.d. = 0.048), and sustained portions were frequent, accounting for a mean PTI near 0.5.

Firing rate and the RPI were hyperbolically related, with the lowest firing rates correlated with high RPI values (Fig. 5D; r = −0.428, P = 0.0005). The PTI was weakly correlated with firing rate, indicating that sustained responses could be recovered over a range of response strengths (Fig. 5E; r = 0.326, P = 0.0097). Finally, the RPI was not correlated with the PTI; thus tonic and phasic responses occurred across a wide range of RPIs (Fig. 5F; r = −0.135, P = 0.295).

Information analysis for response metrics. After we characterized the response types of ICC neurons we then related the types to receptive field processing. The information results for the STA and the MIDs generalized across response type in the ICC. The STA sufficiency was high regardless of the firing rate of ICC neurons, indicating that the STA information approximated that of MID1 across all response strengths (Fig. 6A; r = −0.165, P = 0.352). Additionally, the STA sufficiency was not affected by response precision; it was similar across different RPIs (Fig. 6B; r = 0.285, P = 0.103). Finally, the STA sufficiency was independent of either phasic or tonic response behavior (Fig. 6C; r = 0.053, P = 0.768).

The MID1 contribution measures the dimensionality of ICC processing, and it achieved similarly high values for the three response metrics. For firing rate, the MID1 contribution was not dependent on the response strength of the neuron (Fig. 6B; r = 0.231, P = 0.188). The response precision of the neuron was also not correlated with model complexity (Fig. 6D; r = 0.049, P = 0.784). Finally, the PTI was moderately correlated with the MID1 contribution (Fig. 6E; r = 0.484, P = 0.004). However, for almost all neurons, no matter the PTI, the MID1 contribution was >75%. Therefore, regardless of response strength or temporal response pattern, MID1 is an appropriate model for the spectrotemporal processing of ICC neurons.

Information analysis in ICC and primary auditory cortex. For ICC neurons, a single filter—combined with an appropriate input/output nonlinearity—is sufficient to describe spectrotemporal processing. One filter is appropriate regardless of the response strength; it is also sufficient for phasic-to-tonic response patterns. This finding contrasts with those from AI, where multiple filters are present in granular (~600–1,100
granular layers vs. midbrain:

midbrain: P m.a.d./H11005

than cortical STAs to MID1s (Fig. 7)

ation, midbrain STAs were better approximations to MID1s

AI granular (thalamic input) and nongranular layers. In addi-

single-filter characterization is sufficient for the ICC but not for

/ H9262

Fig. 6. Comparison between response metrics and information in the

2600 STRF DIMENSIONALITY IN AUDITORY MIDBRAIN

m) and nongranular layers (Atencio et al. 2009; Huang and

Winer 2000; Mitani et al. 1985; Rouiller et al. 1991; Winer

1984). To compare the midbrain and the cortex, we used


When we compared the STA sufficiency distributions, we

found that the midbrain population diverged significantly from

the cortical population. Compared with both cortical granular

and nongranular layers, the ICC population was significantly

shifted toward higher values [Fig. 7A; granular layers vs.

midbrain: P < 0.001, Kolmogorov-Smirnov (KS) test; non-

granular layers vs. midbrain: P < 0.001, KS test]. Thus a

single-filter characterization is sufficient for the ICC but not for

AI granular (thalamic input) and nongranular layers. In addi-

tion, midbrain STAs were better approximations to MID1s

than cortical STAs to MID1s (Fig. 7C; ICC: median = 89.7,

m.a.d. = 5.94; AI: median = 63.1, m.a.d. = 18.4; rank sum

test, P < 0.001).

Furthermore, the strength of the first MID, relative to a

combined first and second MID model, was substantially

different between midbrain and cortex (Fig. 7, B and D). The

midbrain had a stronger relative MID1 compared with either

granular or nongranular data (Fig. 7B; granular layers vs.

midbrain: P < 0.001, KS test; nongranular layers vs. midbrain:

P < 0.001, KS test). This indicates that the multi-filter

model of AI is not passively received (“inherited”) from

midbrain processing. At the population level, the MID1 con-

tribution was much greater in the ICC than in AI (ICC: median =

92.0, m.a.d. = 4.58; AI: median = 62.4, m.a.d. = 20.4; rank

sum test, P < 0.001). Thus the second MID contributes much

more to information processing in AI than in the ICC. There-

fore, the number of stimulus dimensions needed to describe AI

neurons is substantially different from the number needed for the

midbrain.

DISCUSSION

A single stimulus filter dominates the spectrotemporal pro-

cessing in the auditory midbrain, whereas primary cortical

neurons in all cortical layers are best characterized by at least

two filters (Atencio et al. 2008, 2009). This implies that the

nature of receptive fields undergoes a fundamental transforma-

tion from subcortical to cortical stations. We established this

result in several ways.

First, for each ICC response type we encountered, whether

phasic or tonic, a single filter accounted for the overwhelming

majority of the conveyed stimulus information. Additionally,

the single-filter description held irrespective of the response

metric we employed, indicating that neither levels of firing rate

nor the degree of response precision conflicts with this result.

Therefore, the single-filter description likely generalizes across

the population of midbrain cells.

Second, the STA was an adequate descriptor of spectrotem-

poral ICC processing. For cells with only one relevant stimulus

dimension, and with appropriate stimuli, the STA and MID

descriptions should be similar. On the basis of filter similarity

and information, the ICC STA was nearly equivalent to MID1.

Thus the STA and MID1 correspond to the same stimulus

dimension that accounts for ICC responses. This is striking,

since the STA is a much simpler measure than an MID, both to

compute and to interpret. It implies that simple reverse corre-

lation may be used to obtain an unbiased receptive field

estimate for the ICC.

Third, the single feature selectivity of the ICC was con-

firmed by assessing the impact of a second MID. The MID

analysis produced a second MID for ICC neurons. However,

the filter shape was diffuse, and the information contributed by

neurons is substantially different from the number needed for the

midbrain.
MID2 was negligible. By contrast, in AI, second MIDs with significant information contributions and reliable filter shapes were encountered for the vast majority of neurons in both the thalamic input/granular layers and the output/extragranular layers. The existence of multiple filters in the cortical input layers permits the possibility that the multiple-filter property is generated before AI. However, since the number of filters required to describe midbrain responses is different from that in the cortical input layers, this eliminates the possibility that cortical processing properties are inherited from collricular neurons. The possibility of the emergence of multiple spectrotemporal filters in thalamic neurons is still feasible.

Transformations in central auditory system. For sound processing, there are relatively few established transformations along the ascending central auditory pathway. The most widely recognized processing change is the transition from a temporal to a more rate-based coding scheme. For example, phase-locking to the stimulus fine structure degrades as information proceeds from lower to higher stations (Johnson 1980; Kuwada et al. 1984; Liu et al. 2006; Lu et al. 2001; Winter and Palmer 1990). Also, the highest temporal modulation frequency to which neurons may respond with a temporal code decreases across the populations (Joris et al. 2004).

Our analyses established a further transformation that is fundamentally different from those that are related to temporal coding: The manner in which spectrotemporal information is processed changes from midbrain to cortex. By showing that the number of STRF components increases from the ICC to AI, we have uncovered one of the first fundamental receptive field transformations in the central auditory system: the emergence of multidimensional receptive fields in the auditory forebrain. The forebrain is composed of both the auditory cortex and the auditory thalamus (medial geniculate body). We can only speculate at this time on whether multidimensional receptive fields will be found in the thalamus. We do note that the thalamus, like the ICC, has a laminated structure (Morest 1965). Additionally, thalamic circuitry appears to be more similar to that in the ICC than to the cortex, and the number of cell types is also restricted relative to cortex (Winer 1992). Given these admittedly cursory considerations, we expect that thalamic receptive field processing will be more similar to that in ICC than that in AI. Thus, if future work shows that the thalamus can also be sufficiently described with a single filter, then the multiple-filter model for AI neurons, which is present in all cortical layers, would be an emergent feature of the cortical circuit itself.

The change from single to multiple feature selectivity is conceptually significant since it indicates that even though the ICC is an obligatory processing station for auditory information, further emergent changes take place in the forebrain. The conclusiveness of previous characterizations of central processing principles has often been hampered, since it has been difficult to demonstrate and isolate truly emergent properties in AI with simple stimuli. A creation of de novo receptive field properties had not been described for AI. This contrasts with primary visual cortex, where orientation selectivity and binularity first emerge in the thalamic recipient layer 4, and therefore experiments in that modality can be tailored to understand the development, plasticity, and circuitry governing the emergence of functional parameters. Since our findings in the ICC implicate the auditory forebrain as the site of emergent processing, experiments may now be performed to understand how this processing may be enhanced, degraded, or modified in accordance with approaches including network manipulations, development, learning, and/or attention. Therefore, the function of the auditory cortical circuit may be examined, and changes to the circuit itself may be correlated with specific aspects of receptive field processing and with the implementation of task-specific algorithms.

Previous results. Spike-triggered analysis techniques have been successfully applied to ICC neurons. Nonlinear receptive field aspects have been revealed with uncorrelated noise and DMR stimulation (Escabí and Schreiner 2002). STRFs have also been moderately successful in linearly predicting midbrain responses (Andoni et al. 2007; Lesica and Grothe 2008). In the bat ICC, the STRF from spike-triggered averaging can predict frequency modulation preferences (Andoni et al. 2007). The structure of the STRF was found to be highly correlated with frequency modulation direction selectivity. In the gerbil ICC, a complex rain sound has been used to calculate the STRF (Lesica and Grothe 2008). Here, the STRF was estimated with ridge regression, which is a modified extension of the STA (Lesica and Grothe 2008). The filter estimated from this approach was relatively successful in predicting the temporal spiking behavior of gerbil ICC cells. Finally, in the awake rhesus monkey ICC, STRFs were used to predict responses to noise and vocalizations (Versnel et al. 2009). Here STRFs were estimated from static ripple stimuli. Like the DMR, static ripple stimuli are characterized by two parameters, a temporal and a spectral modulation frequency. Unlike the DMR, the two parameters do not vary with time (Klein et al. 2000). The predictive power of the rhesus monkey STRFs was probably underestimated for two reasons. First, the predictions were not compared to the inherent variance in the neural response (Hsu et al. 2004; Sahani and Linden 2003). Second, the study did not use a static nonlinearity, which is known to increase predictive power (Sharpee et al. 2008). Therefore, our study provides a context in which to view these successes, since our findings indicate that, for some stimuli, ICC cells can be adequately described with a single filter followed by a customized static input/output nonlinearity.

Implications for receptive field analysis. We were able to apply spike-triggered techniques to ICC neurons because neural systems satisfy fundamental systems analysis requirements. Theoretically, a nonlinear dynamic system may be described by a Volterra series if the system is nonchaotic, nonoscillatory, and time-invariant and has finite memory (Boyd and Chua 1985; Marmarelis 2004; Westwick and Kearney 1998). Importantly, the Volterra series can be formulated in the linear-nonlinear (LN) model configuration (Palm 1979). The LN model, which is also termed a Wiener-Bose or parallel-cascade model, is a bank of linear filters followed by a static nonlinearity (Korenberg 1991; Marmarelis 1997, 2004). The STA/MID approach, with accompanying nonlinearities, falls under the general framework of the LN model (Schwartz et al. 2006). The only stimulus requirement for these models is that the statistics of the stimulus be broad enough to effectively cover the stimulus preferences of the neuron (Marmarelis and Marmarelis 1978; Marmarelis 2004). Appropriate techniques may then be used to correct for stimulus correlations and to estimate the proper number of filters (Sharpee et al. 2004). Hence, for auditory neurons, we can calculate STRFs if the stimulus
modulates the firing rate. Thus spike-triggered analyses can also be used to characterize tonic neural responses. For a dynamic stimulus, only when the firing rate does not modulate in response to changing stimulus features will the STRF procedure not work. However, in this case, the neuron’s response is not informative. In short, if the response varies when stimulus features vary, then judiciously applied spike-triggered approaches may be useful.

Further considerations. Like every approach, the scope of our results are necessarily constrained by the details and assumptions of our methodology. In this regard there are important issues to consider. First, we employed the DMR stimulus because it contains temporal and spectral modulations, which are important components of natural sounds. Natural sounds may have other features that are not modeled by the DMR, and thus calculating MIDs with naturalistic stimuli may lead to moderately different results (David et al. 2009; Singh and Theunissen 2003). Second, how the DMR is represented may be pertinent. Although the neurons themselves responded to the entire stimulus waveform, we estimated filters with the spectrotemporal envelope. The stimulus waveform fine structure is a richer representation than the envelope. Previous work has shown that auditory cortical neurons may indeed be sensitive to stimulus fine structure (Elhilali et al. 2004). Sensitivity to fine temporal features may be even more pronounced in the ICC, which has better temporal resolution than auditory cortex (Joris et al. 2004). Thus using a specific form of the spectrotemporal envelope may not allow us to capture the complete processing of a neuron. Third, the state of the animal may be a factor, since anesthesia, for example, may decrease the number of tonic responses and change inhibitory mechanisms (Astell et al. 1996; Lu et al. 2001; Versnel et al. 2009). This would affect the nature of the STRFs and nonlinearities, though not our ability to calculate them, since in the present study we were able to calculate filters for neurons that had tonic responses. Fourth, we presented the DMR at one mean intensity, although the responses to the DMR may change at different intensities. Earlier work showed varying STRFs for differing intensities (Lesica and Grothe 2008). This effect may be partially mitigated in the case of the DMR, since it covers a 40-dB dynamic range. Finally, further work needs to ascertain the predictive power of a single MID. Our present study cannot address this, since we did not present multiple repetitions of a single stimulus segment (Sharpee 2007). The repeats are required since predictions need to be verified that spike-triggered analyses were appropriate. The information from the STA was similar to the MID1 information, and the MID1 information was similar to the joint MID1 and MID2 information. These results generalized across all temporal response types; thus, in the ICC, a single spectrotemporal filter may adequately describe acoustic processing. In contrast, the auditory cortex requires a multifilter description. Therefore, receptive field dimensionality increases from midbrain to cortex.

ACKNOWLEDGMENTS

We thank Andrew Tan, Marc Heiser, Kazuo Imaizumi, and Beneditte Philibert for experimental assistance and Mark Kvale for the use of his SpikeSort 1.3 Bayesian spike-sorting software.

GRANTS

C. A. Atencio and C. E. Schreiner were supported by National Institutes of Health Grants DC-02260 and MH-077970, the Coleman Memorial Fund, and Hearing Research Inc. T. O. Sharpee was supported by an Alfred P. Sloan Fellowship, a Searle Scholarship, NIH Grants R01 EY-019493 and K25 MH-068904, the Ray Thomas Edwards Career Development Award in Biomedical Sciences, and the McKnight Scholarship. Computing resources were provided by the National Science Foundation (NSF) through TeraGrid resources provided by supercomputer resources at the San Diego Supercomputer Center, Argonne National Laboratory, University of Illinois National Center for Supercomputing Applications, and Texas Advanced Computing Center. Additional resources were provided by the Center for Theoretical Biological Physics (NSF PHY-0022283).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: C.A.A., T.O.S., and C.E.S. conception and design of research; C.A.A. and C.E.S. execution and analysis of research; C.A.A. and T.O.S. interpretation of results of experiments; C.A.A. and T.O.S. drafted manuscript; C.A.A., T.O.S., and C.E.S. final version of manuscript.

REFERENCES


J Neurophysiol doi:10.1152/jn.01025.2011 www.jn.org