Spatiotemporal Conversion of Auditory Information for Cochleotopic Mapping

Osamu Hoshino
hoshino@mx.ibaraki.ac.jp
Department of Intelligent Systems Engineering, Ibaraki University, Hitachi, Ibaraki, 316-8511 Japan

Auditory communication signals such as monkey calls are complex FM vocal sounds and in general induce action potentials in different timing in the primary auditory cortex. Delay line scheme is one of the effective ways for detecting such neuronal timing. However, the scheme is not straightforwardly applicable if the time intervals of signals are beyond the latency time of delay lines. In fact, monkey calls are often expressed in longer time intervals (hundreds of milliseconds to seconds) and are beyond the latency times observed in the brain (less than several hundreds of milliseconds). Here, we propose a cochleotopic map similar to that in vision known as a retinotopic map. We show that information about monkey calls could be mapped on a cochleotopic cortical network as spatiotemporal firing patterns of neurons, which can then be decomposed into simple (linearly sweeping) FM components and integrated into unified percepts by higher cortical networks. We suggest that the spatiotemporal conversion of auditory information may be essential for developing the cochleotopic map, which could serve as the foundation for later processing, or monkey call identification by higher cortical areas.

1 Introduction

Frequency modulation (FM) is a critical parameter for constructing auditory communication signals in both humans and monkeys. For example, human speech contains FM sounds called formant transitions that are critical for encoding consonant and vowel combinations such as “ga,” “da,” and “ba” (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Suga, 1995). Monkeys use complex FM sounds—the so-called monkey calls—that are considered to be a precursor of human speech (Poremba et al., 2004; Holden, 2004), in order to make social interactions with other members of the species (Symmes, Newman, Talmage-Riggs, & Lieblich, 1979; Janik, 2000; Tyack, 2000).

Auditory FM signals differ in informational structure from other sensory signals in that they are processed in a time-dependent manner (or characterized by time-varying spectral frequencies), while color and orientation...
in vision, odorants in olfaction, and tastants in gustation are characterized in most cases in a time-independent manner. Understanding how the auditory cortex encodes and detects the streams of spectral information arising from the temporal structure of FM sounds is one of the most challenging problems in auditory cognitive neuroscience.

Auditory signals being sent from receptor neurons enter the primary auditory area (AI). Neurons of the AI are orderly arranged and form the so-called tonotopic map, where tonotopic representation of the cochlea is well preserved (Yost, 1994). Each neuron on the map tends to respond best to its characteristic frequency. Because of the tonotopic organization, these neurons could be activated in a sequential manner and generate action potentials in different timing when stimulated with an FM sound. It is interesting to know how higher cortical areas, to which the AI project, detect the timing of action potentials so that the brain can identify the applied FM sound.

One of the effective ways for detecting the timing of action potentials is to use delay lines. Jeffress (1948) proposed a theory for sound localization. The theory is that an interaural time difference, which expresses the azimuthal location of a sound source, could be detected by neurons that receive signals from both ears through distinct delay lines. Suga (1995) proposed a multiple delay line scheme for echo sound detection in bats. Bi and Poo (1999) demonstrated in a cultured hippocampal neuronal network that the timing of action potentials could be detected when relevant delay lines are properly chosen.

The timing of action potentials in these processes was relatively short, ranging from submilliseconds to tens of milliseconds, for which the delay line scheme worked well. However, the information about FM sounds contained in monkey calls is in general expressed in longer time intervals of hundreds of milliseconds to seconds. The delay line scheme is unlikely to be applicable for them, because the delay lines observed in the brain are at most of several hundreds of milliseconds (Miller, 1987). This implies that the brain might employ another strategy for identifying individual calls.

A speculative cochleotopic map was proposed in relation to a retinotopic map in vision (Rauschecker, 1998). The retinotopic map expresses information about the location of a moving bar in a two-dimensional visual space that is projected onto the retina. As the bar moves in the visual space, the map shows a spatiotemporal firing pattern of neurons. The axes of the map indicate the location of the bar in the two-dimensional visual space. In the cochleotopic mapping scheme, the location of neuronal activation moves along the frequency axis as the frequency of a FM sound sweeps. However, the exact auditory variable for the other axis is still unknown.

We propose here a hypothetical two-dimensional (frequency axis, propagation axis) cochleotopic neural network \(N_{\text{CO}}\) network) for the AI on which information about FM sounds is mapped in a spatiotemporal manner. The propagation axis is assumed for the unknown axis. Stimulation
with a pure (single-frequency) tone activates a peripheral neuron, whose activity propagates along the propagation axis, or along its isofrequency band. When the $N_{CO}$ network is stimulated with an FM sound, peripheral neurons are sequentially activated, which then propagates along their isofrequency bands. As a consequence, the information about the applied FM sound is expressed as a specific spatiotemporal firing pattern in the $N_{CO}$ network dynamics.

Such activity propagation has been reported in the AI. Hess and Scheich (1996) stimulated Mongolian gerbils with pure tones (1 kHz to 16 kHz) and recorded the activity of AI neurons. The researchers found that neuronal activation propagated along isofrequency bands at all frequencies. Taniguchi, Horikawa, Moriyama, and Nasu (1992) stimulated guinea pigs with pure tones (1 kHz to 30 kHz) and recorded the activity of the anterior field (field A) in which tonotopic organization was well preserved. The researchers found that the focal activation beginning in field A propagated in two directions: along isofrequency bands and toward field DC.

There has been neurophysiological evidence that FM sounds are precisely detected by auditory neurons. Neurons of the lateral belt (LB) (Rauschecker, 1998) that receives signals from the AI responded to FM sounds. These neurons were relatively organized in an orderly fashion depending on the sweeping rate (between slow and fast) and direction (upward or downward) of the FMs. Based on these experimental findings, we construct a neural network model ($N_{FM}$) for the LB to which the $N_{CO}$ network projects. A given FM sound evokes a specific spatiotemporal firing pattern in the $N_{CO}$ network, to which a certain group of $N_{FM}$ neurons ($N_{FM}$ column) responds and identifies the applied FM.

It is also well known that LB neurons respond to monkey calls (Rauschecker, 1997). Monkey calls as vocal signatures are complex FM sounds and play an important role in identifying individuals, especially when their visual systems are unavailable, as in a luxuriant forest (Symmes et al., 1979). To detect such complex FM sounds, we construct a higher neural network ($N_{ID}$ network) model for the STGr (rostral portion of the superior temporal gyrus) to which the LB projects. The $N_{ID}$ network receives selective projections from the $N_{FM}$ network. When a monkey call is presented to the $N_{CO}$ network, multiple $N_{FM}$ columns are sequentially activated in a specific order. The $N_{ID}$ network integrates the sequence of the dynamic $N_{FM}$ columns, thereby identifying that call.

Based on the proposed cochleotopic mapping scheme, we investigate how FM sound information is encoded and detected. Applying to the $N_{CO}$ network simple (linearly sweeping) and complex (monkey call) FM sounds, we record the activities of neurons. By statistically analyzing them, we try to understand the neuronal mechanisms that underlie FM sound information processing in the auditory cortex.
2 Neural Network Model

2.1 Outline of the Model. The $N_{CO}$ network, modeling the AI, is organized in a tonotopic fashion as shown in Figure 1A. When the frequency of an applied FM sound sweeps upward, the neuronal activation of the periphery (p1) sweeps from $f_1$ to $f_{40}$, which then moves along the propagate axis (isofrequency bands). The filled circles schematically indicate a neuronal firing pattern induced by a simple (linearly sweeping) upward FM sound at a certain time after the stimulus onset. The gray circles indicate a neuronal firing pattern induced by a downward FM sound. The spatiotemporal firing pattern of the $N_{CO}$ network expresses combinatorial information about the direction and the sweep rate of the applied FM sound. Neurons of the $N_{FM}$ network, modeling the LB, receive convergent projections from the $N_{CO}$ network (solid and dashed lines), and detect the upward (black circles) and downward (gray circles) FM sounds. We made sets of selective convergent projections from the $N_{CO}$ to $N_{FM}$ network in order to detect specific linearly sweeping FM sounds.

Neurons within isofrequency bands are connected by excitatory and inhibitory delay lines, as shown in Figure 1B. The excitatory connections (solid lines) from the periphery to the center are the major driving force for the neuronal activation to move along the propagation axis, and the inhibitory connections (dashed lines) were employed to sharpen the spatiotemporal firing patterns. This specific circuitry was used for functionally expressing the propagation axis whose evidence in the AI is being accumulated (Hess & Scheich, 1996; Taniguchi et al., 1992). For simplicity, there is no connection between isofrequency bands. Neurons within $N_{FM}$ columns are connected with each other via excitatory synapses, and neurons between $N_{FM}$ columns are connected via inhibitory synapses. This circuitry enables each $N_{FM}$ column to respond selectively to a specific linearly sweeping FM sound.

Figures 1C and 1D are schematic drawings of neuronal responses to a linearly sweeping FM sound. When the $N_{CO}$ network is stimulated with an upward FM sound sweeping at a slow (Figure 1C, left), intermediate (Figure 1C, center), or fast rate (Figure 1C, right), the activation area moves from the lower left to upper right (arrows). When a downward FM sound is presented, the activation area moves from the lower right to upper left (see Figure 1D). When the activation area reaches a certain position (gray ellipses), the neurons send action potentials to the $N_{FM}$ network via the selective feedforward projections, and activate the $N_{FM}$ column corresponding to the applied FM (black ellipses). The neuronal activation of the other $N_{CO}$ regions (dashed ellipses) can also be used for the FM detection. Nevertheless, we chose the firing patterns (gray ellipses), because these patterns appear first in the time courses with maximal neurons simultaneously activated. This enables the $N_{FM}$ network to respond reliably and rapidly to the applied FM sounds.
Figure 1: Neural network model. (A) The $N_{CO}$ network is organized in a tonotopic manner. The $N_{FM}$ network receives selective projections from the $N_{CO}$ network. Among $N_{FM}$ columns, $c1$–$c10$ and $c31$–$c40$ detect FM sounds that sweep linearly in downward and upward directions, respectively. For clarity, only two sets of firing patterns (black and gray circles) and projections (solid and dashed lines) for an upward and a downward FM sound are depicted. (B) Neuronal connections within isofrequency bands of the $N_{CO}$ network. Neurons are connected via excitatory (solid lines) and inhibitory (dashed lines) delay lines, where $\Delta t$ denotes a signal transmission delay time. (C) Schematic drawings of the spatiotemporal neuronal responses of the $N_{CO}$ network and those of the $N_{FM}$ network to simple (linearly sweeping) upward FM sounds. Activity patterns for FM sounds that sweep at a slow (left), intermediate (center), and fast (right) rate are shown. Arrows indicate the directions of movements of active areas. (D) Schematic drawings as in C for downward FM sounds.
2.2 Model Description. Dynamic evolutions of the membrane potentials of neurons of the $N_{CO}$ and $N_{FM}$ networks are defined, respectively, by

$$
\tau_{CO} \frac{d u_{k,j}^{CO}(t)}{dt} = -u_{k,j}^{CO}(t) + \sum_{j=1}^{M_{D}} \left[ w_{k,j,k,-j}^{EX} s_{k,-j}^{CO}(t - j \Delta t) + w_{k,j,k,j+1}^{IH} s_{k,j+1}^{CO}(t - j \Delta t) \right],
$$

(2.1)

$$
\tau_{FM} \frac{d u_{i}^{FM}(t)}{dt} = -u_{i}^{FM}(t) + \sum_{k=f1}^{f40} \sum_{j=p1}^{p40} L_{i,k-j} s_{k,j}^{CO}(t - \Delta t_{FM}) + \sum_{j=1(j \neq i)}^{M_{FM}} w_{ij}^{FM} s_{j}^{FM}(t),
$$

(2.2)

where

$$
Prob[S_{i}^{Y}(t) = 1] = f_{Y}[u_{i}^{Y}(t)] \quad (Y = CO, FM),
$$

$$
f_{Y}[u] = \frac{1}{1 + e^{-\eta_{Y}(u - \theta_{Y})}}.
$$

(2.3)

$u_{k,j}^{CO}(t)$ and $u_{i}^{FM}(t)$ are the membrane potential of the $i$th $N_{CO}$ neuron of the $k$th ($k = f1–f40$) isofrequency band and that of the $i$th $N_{FM}$ neuron at time $t$, respectively. $\tau_{Y}$ ($Y = CO, FM$) is a decay time of the membrane potential of the network $N_{Y}$. $M_{D}$ is the number of excitatory or inhibitory input delay lines that a single $N_{CO}$ neuron receives from other neurons within isofrequency bands. $w_{k,j,k,-j}^{EX}$ and $w_{k,j,k,j+1}^{IH}$ are, respectively, excitatory and inhibitory synaptic connection strengths from neuron $(i - j)$ to $i$ and from $(i + j)$ to $i$ of the $k$th isofrequency band. $s_{k,j}^{CO}(t - j \Delta t) = 1$ expresses an action potential of the $j$th $N_{CO}$ neuron of the $k$th isofrequency band, where $j \Delta t$ denotes a signal transmission delay time (see Figure 1B). $(f1–f40, p1–p40)$ denotes the locations of neurons on the two-dimensional $N_{CO}$ map (see Figure 1A). $L_{i,k-j}$ is the strength of synaptic connection from the $j$th $N_{CO}$ neuron of the $k$th isofrequency band to the $i$th $N_{FM}$ neuron. $\Delta t_{FM}$ is a signal transmission delay time from the $N_{CO}$ to $N_{FM}$ network. $M_{FM}$ is the number of $N_{FM}$ neurons. $w_{ij}^{FM}$ is the synaptic connection strength from the $j$th to the $i$th $N_{FM}$ neuron, and $s_{j}^{FM}(t) = 1$ expresses an action potential of the $j$th $N_{FM}$ neuron. $\{w_{ij}^{FM}\}$ was set for the neurons within $N_{FM}$ columns ($c1–c40$; see Figure 1A) to be mutually excited and for the neurons between $N_{FM}$ columns to be laterally inhibited. $\eta_{Y}$ and $\theta_{Y}$ are the steepness and the threshold of the sigmoid function $f_{Y}$, respectively, for $Y$ neuron. Equation 2.3 defines the probability of firing of neuron $i$; that is, the
probability of \( SY_i(t) = 1 \) is given by function \( f_Y \). After firing, its membrane potential is reset to 0.

FM sound stimuli are applied to the peripheral neurons of the \( N_{CO} \) network. Dynamic evolutions of membrane potentials of these neurons are defined by

\[
\tau_{CO} \frac{du_{CO}^{k-1}(t)}{dt} = -u_{CO}^{k-1}(t) + \sum_{j=1}^{M_D} w_{k-1,k,(1+j)}^{jh}(t) S_{CO}^{k,(1+j)}(t - j \Delta t) + \alpha I_{k-1}(t),
\]

(2.4)

where \( I_{k-1}(t) \) is the input stimulus to the peripheral neuron of the \( k \)th isofrequency band, or the neuron located at \((k, p_1; k = f_1–f_{40}; \text{see Figure 1A})\). \( \alpha \) is the intensity of the input. Note that the peripheral neurons receive only an inhibitory input from the \((1 + j)\)th neuron with a delay of \( j \Delta t \) and do not receive any delayed excitatory input.

Network parameter values are as follows. The number of neurons are 40 \((f_1–f_{40}) \times 40 (p_1–p_{40}) \) and 40 \((c_1–c_{40}) \times 12 \) for the \( N_{CO} \) and \( N_{FM} \) networks, respectively: \( \tau_{CO} = 10 \text{ ms}, \tau_{FM} = 10 \text{ ms}, \theta_Y = 0.7, \eta_Y = 10.0, \) and \( M_D = 3. \)

\( u_{k,i,k,(i-1)}^{CO} = 5.0 \) and \( w_{k,j,k,(i-1)}^{ih} = -0.5 \). \( \Delta t = 10 \text{ ms} \) and \( \Delta t_{FM} = 20 \text{ ms} \). \( L_{i,k,j} \) was selectively set at either 0.1 or 0, as addressed in section 2.1, by which the specific firing patterns induced in the \( N_{CO} \) network dynamics can activate their corresponding \( N_{FM} \) columns (see Figures 1C and 1D).

\( \alpha = 8.0, \) and \( I_{k-1}(t) = 1 \) for an input and 0 for no input.

3 Results

3.1 Tuning Property to Simple FM Sounds. We show here how the information about simple (linearly sweeping) FM sounds could be expressed as spatiotemporal firing patterns in the \( N_{CO} \) network dynamics. We also show how the auditory information could be transferred to and detected by specific neuronal columns of the \( N_{FM} \) network. Response properties (action potential generation) of \( N_{FM} \) neurons are compared with those observed experimentally.

An upward FM sound sweeping linearly at 20 Hz per ms induces a specific spatiotemporal firing pattern in the \( N_{CO} \) network, in which the neuronal activation moves from the lower left toward the upper right (see Figure 2A). When the activity reaches a certain point (time = 200 ms), the active neurons send action potentials to the \( N_{FM} \) network and stimulate the corresponding \( N_{FM} \) column (arrow) at time = ∼220 ms. The difference in activation time between the \( N_{CO} \) (200 ms) and \( N_{FM} \) (220 ms) networks arises from a difference in signal transmission delay between the two networks, or \( \Delta t_{FM} = 20 \text{ ms} \) (see equation 2.2).
We assumed $f_1 = 8$ kHz to $f_{40} = 11.9$ kHz (see Figure 1A), where the isofrequency bands were placed at an even interval, or 100 Hz per band. These frequencies are within the range observed in squirrel monkeys (Symmes et al., 1979) and employed for investigating how complex FM sounds such as monkey calls could be identified, as will be shown in sections 3.2 and 3.3. Figure 2B presents the total spike counts (top) and raster plots (middle) for the neurons of a given $N_{FM}$ column when stimulated with different upward FM sounds (bottom). The columnar neurons show specific sensitivity to the upward FM sound (20 Hz per ms) but less to downward FM sounds (see Figure 2C). This tendency is almost consistent with that observed in macaque monkeys (Rauschecker, 1998).

Although the tuning characteristic of the $N_{FM}$ columnar neurons to the applied FM sound (sweep rate = 20 Hz per ms; upward) is evident, these neurons also show weak responses to the other upward FM sounds with different sweep rates (see the arrows of Figure 2B). Such weak responsiveness to the “irrelevant” FM sounds is due to the overlapping of $N_{CO}$ firing patterns, as schematically shown in Figure 3. The set of $N_{CO}$ neurons within the solid ellipse, which are to be simultaneously activated by the FM stimulus (sweep rate = 20 Hz per ms; upward), send action potentials to the relevant $N_{FM}$ column and maximally activate the column (top left). However, the subsets of $N_{CO}$ neurons within the overlapping regions (gray and black) could also be activated, respectively, by the irrelevant upward FM sounds (26.7 and 13.3 Hz per ms). These neurons send a relatively small number of action potentials to the same $N_{FM}$ column, which results in weaker neural responses in the column (bottom left and top right).

3.2 Tuning Property to Complex FM Sounds. We show here how complex FM sounds such as monkey calls could be expressed as specific spatiotemporal firing patterns in the $N_{CO}$ network dynamics. We also show
how the information about individual monkey calls can be decomposed into simple (linearly sweeping) FM components.

We used artificial isolation peep (IP) as monkey calls (Symmes et al., 1979), whose pitch profiles are shown in Figure 4A. A specific spatiotemporal firing pattern is induced in the \( N_{CO} \) network when stimulated with the IP of monkey X, which then activates multiple \( N_{FM} \) columns in a sequential manner (see Figure 4B). We have observed distinct sequential orders of columnar activation for monkey call X, Y, and Z. Figure 4C presents the details of the sequences of columnar activation for monkey X (the thick solid line), Y (the thin solid line), and Z (dashed line), indicating that the IPs are decomposed into specific sequences of simple (linearly sweeping) FM components.

In the model, the neurons of a currently active \( N_{FM} \)-column continue firing, even without any excitatory input, until another dynamic \( N_{FM} \) column emerges, or its neurons begin to fire. For example, \( N_{FM} \) column c31 is activated at time \( = 0.22 \) s (upward arrow) and continues firing (downward arrow) until the \( N_{FM} \) column c1 begins to fire (at 0.34 s; rightward arrow).
Figure 4: Spatiotemporal conversion of information about isolation-peeps (IPs).
(A) Profiles of artificial IPs for monkey calls (X, Y, Z) used in the simulations.
(B) Time courses of neuronal activation of the NCO and NFM networks induced by the IP of monkey X. (C) Sequences of dynamic NFM columns induced by the IPs of monkey X (thick solid line), Y (thin solid line), and Z (dashed line). Each NFM column, c31–c40 and c1–10, responds to a specific linearly sweeping FM component involved in the IPs. NFM columns (c11–c30) were not assigned to detect FM sounds in this simulation.
This self-generative continuous firing could be mediated by mutual excitation within \( N_{FM} \) columns. In the next section, we try to identify these monkey calls by an integrative process, to which the persistent neuronal firing effectively contributes.

### 3.3 Monkey Call Identification

We showed in the previous section that the information about complex FM sounds, or the IPs of monkeys, could be decomposed into simple (linearly sweeping) FM components. It reminds us that early visual systems decompose complex visual objects into simple features such as edge, orientation, and color. Higher visual areas integrate these features so that the visual images of the objects can be reconstructed as unified percepts.

To identify the sequences of the dynamic \( N_{FM} \) columns, we extended the model, adding a higher (integrative) neural network \( N_{ID} \) (see Figure 5A). We made selective feedforward projections from the \( N_{FM} \) to \( N_{ID} \) network (solid lines) in order to integrate the series of FM components. The specific \( N_{ID} \) column (filled circles; \( N_{ID} \)) assigned to detect a certain IP receives convergent inputs from the \( N_{FM} \) columns (filled circles; \( N_{FM} \)) that are to be sequentially activated by the FM components constituting the IP (filled circles; \( N_{CO} \)). The \( N_{ID} \) network may correspond to a lateral belt (LB) area that is known to respond to monkey calls or the rostral portion of the superior temporal gyrus (STGr) to which the LB projects (Rauschecker, 1998), as we assumed here.

Dynamic evolutions of membrane potentials of \( N_{ID} \) neurons are defined by

\[
\tau_{ID} \frac{d u_{id}^{ID}(t)}{dt} = -u_{id}^{ID}(t) + \sum_{j=1}^{M_{FM}} L_{ij}^{ID} S_{id}^{FM}(t - \Delta t_{ID}) + \sum_{j=1(j\neq i)}^{M_{ID}} w_{ij}^{ID} S_{jd}^{ID}(t),
\]

where action potentials are generated according to equation 2.3 and the other parameter values were the same as those for the \( N_{FM} \) network (see equation 2.2).

As shown in Figure 5B, when the \( N_{CO} \) network is stimulated with the IP of monkey X, multiple \( N_{FM} \) columns are sequentially activated, which then activate the call-relevant \( N_{ID} \) column at 367 ms. In this integrative process, the \( N_{ID} \) column receive consecutive action potentials from the \( N_{FM} \) columns, as addressed in the previous section (see Figure 4C). Such continuous activation of the \( N_{ID} \) column gradually depolarizes its neurons and allows them to fire when the membrane potentials reach a threshold, thereby identifying the applied IP. Figure 5C presents the identification processing for monkey Z.
Figure 5: Monkey call identification. (A) A higher (integrative) neural network $N_{ID}$ is added to the original model (see Figure 1A), which integrates the specific sequences of dynamic $N_{FM}$ columns induced by individual IPs. (B,C) Time courses of neuronal activation in the $N_{CO}$, $N_{FM}$, and $N_{ID}$ networks when stimulated with the IPs of monkeys X (B) and Z (C).

Figure 6 presents how similar monkey calls could be distinguished from each other. The IP of monkey V (see Figure 6A, left) has a similar spectrogram in the first part (time = 0–0.1 s; solid line) to that of monkey A (dashed line). When the $N_{CO}$ network is stimulated with the IP (see Figure 6B), multiple $N_{FM}$ columns are sequentially activated, which then activate the two
Figure 6: Identification of monkey calls that have similar auditory spectrograms. (A) Profiles of artificial IPs for monkeys V and W (solid lines). The dashed lines denote that of monkey A. (B,C) Time courses of neuronal activation in the NCO, NFM, and NID networks when stimulated with the IPs of monkeys V (B) and W (C).

NID columns corresponding to the IPs of monkeys V and A (at 330 ms). The two dynamic NID columns compete for a while (at 330–446 ms), and the NID column corresponding to monkey V finally prevails (at 500 ms). The neuronal competition between the two dynamic NID columns arises from the lateral inhibition between NID columns.

In contrast, when the IP of monkey W, which has a similar spectrogram in the last part (0.3–0.4 s; see Figure 6A, right), is presented, the NID column corresponding to the IP of monkey W is selectively activated (at 338 ms) without competition, as shown in Figure 6C. Note that the time required
to identify monkey W (338 ms; see Figure 6C) is shorter than that for monkey V (500 ms; see Figure 6B), which arises presumably because there is less neuronal competition between dynamic N1D columns. These results indicate that the spectrogram of the last part might be useful for further analyses if circumstances require, although it is more time consuming.

4 Discussion

We have proposed a hypothesis that (1) information about monkey calls could be mapped on a cochleotopic cortical network as spatiotemporal firing patterns, (2) which can then be decomposed into simple (linearly sweeping) FM components and (3) integrated into unified percepts by higher cortical networks.

For the cochleotopic two-dimensional map (hypothesis 1), we assumed activity propagation along isofrequency axis (bands) in order to make a distinct spatiotemporal firing pattern for a given monkey call. Imaging studies (Taniguchi et al., 1992; Hess & Scheich, 1996; Song et al., 2005) evidenced such activity propagation in the primary auditory cortex (AI). When presented with alternating pure tones or alternation between 1 and 8 kHz (Hess & Scheich, 1996), the activity propagation was confined to the low and high isofrequency bands. To our knowledge, the exact formation of spatiotemporal firing patterns for FM sound stimulation has not been identified yet, but we simply extended this scheme. Namely, the neuronal activity propagates along multiple isofrequency bands corresponding to the tone frequencies constituting the applied FM sound. Actual spatiotemporal firing patterns induced by monkey calls might be rather complex because of the interaction between isofrequency bands or the influences of other brain regions. Accordingly, the information about individual monkey calls could be encoded more precisely in the auditory cortex. Nevertheless, the proposed simple cochleotopic map was sufficient to generate distinct spatiotemporal firing patterns and worked well as the foundation for later sound processing, or monkey call identification.

The delay line proposed for the propagation axis (see Figure 1A) was our speculation prompted by a recent experimental study (Song et al., 2005). The study demonstrated that an electrical pulse, applied focally within an isofrequency band, triggered activity propagation along the isofrequency band that was similar to tone-evoked activation. When the auditory thalamus was chemically lesioned, the electrically evoked activity in the AI was not affected, but the tone-evoked activity was abolished. Based on these results, it was suggested that intracortical connectivity in the AI enables neuronal activity to propagate along isofrequency bands. The underlying neuronal mechanisms of activity propagation in the AI has not fully been understood yet, but we assumed the intracortical connectivity via delay lines (see Figure 1B) for developing the activity propagation. Note that the intracortical delay lines are relatively short (less than tens of milliseconds)
that could be neurophysiologically plausible in the brain as addressed in section 1.

For expressing the information about simple (linearly sweeping) FM components (hypothesis 2), we assumed neurons respond selectively to the sweeping rates and directions of FMs. Neurophysiological studies (Rauschecker, 1997, 1998; Tian & Rauschecker, 1998) demonstrated that many neurons of lateral belt areas to which the primary auditory cortex (AI) projects responded better to more complex stimuli, such as FMs and band passed noises, than to pure tones. These neurons were highly selective to the rates and directions of FMs. Neurons of the anterolateral (AL) and caudolateral (CL) belt areas responded better to slower and faster FM sweep rates, respectively. Neurons of the posterior auditory field were highly selective for one direction. The detailed organization of these (rate- and direction-selective) neurons has not clearly been identified yet, but we represented them in a simple and functional manner (see NFM; Figure 1A).

To integrate the sequence of linearly sweeping FMs, expressing a specific monkey call, into a unified percept (hypothesis 3), we assumed that neurons respond selectively to the call. Neurophysiological studies (Rauschecker, 1997, 1998) demonstrated that neurons of the lateral belt responded to a certain class of monkey calls. Very few neurons responded to a single call, and most neurons responded to a number of calls. These results imply that the lateral belt is not yet the end state processing monkey calls. Higher cortical areas such as the rostral portion of the superior temporal gyrus (STGr) or the prefrontal cortex, or both, might be responsible for monkey call identification (Rauschecker, 1998). The exact cortical areas whose neurons have selective responsiveness to individual monkey calls have not clearly been identified yet, but we assumed such “call-selective” neurons in the STGr (see NID; Figure 5A). The selective projections from the NFM to NID (or the lateral belt to STGr) were our speculation in order for the model to perform the identification of individual monkey calls or detect the sequence of FM components.

Coincidence detection based on a delay line scheme, as addressed in section 1, cannot be applicable for longer signals such as monkey calls (more than 500 ms), because delay lines observed in the brain are at most of several hundred milliseconds. The delay lines proposed here for the propagation axis (NCO; Figure 1B) are shorter ones (tens of milliseconds) that were speculative but neurophysiologically plausible in the brain. The idea of this study is on temporal-to-spatiotemporal conversion of auditory information mediated by shorter (plausible) delay lines (~tens of milliseconds) but not on a coincidence detection scheme.

In the propagation axis, we assumed delay lines between neighboring neurons ranging from 10 to 30 ms (see section 2.2). This architecture allowed the cochleotopic neural network (NCO) to propagate along isofrequency bands as observed in the auditory cortex (Taniguchi et al., 1992; Hess & Scheich, 1996). To our knowledge, such delay lines have not been reported
in the auditory cortex. However, Hess and Scheich (1996) pointed out that activity propagation along isofrequency bands might be closely related to the distribution of response latency in the AI. Tian and Rauschecker (1998) found a response-latency distribution (23 ± 12 ms) in the AI. The range of delay lines used for the $N_{\text{CO}}$ network as an AI area is within the range observed.

We assumed activity propagation along isofrequency bands. However, activity propagation across isofrequency bands has also been reported (Taniguchi et al., 1992; Hess & Scheich, 1996), for which the interaction between different isofrequency bands might be responsible. The neuronal activation propagated toward the two (isofrequency and tonotopic-gradient) directions, where the peak activity was shifted along isofrequency bands (Hess & Scheich, 1996). Although the spatiotemporal firing pattern propagating toward the two directions might contribute to encoding auditory information, presumably in a more precise manner, we modeled the propagation only along the isofrequency bands.

Panchev and Wermter (2004) proposed a neural network model that can detect temporal sequences in timescale from hundreds of milliseconds to several seconds. The network consisted of integrate-and-fire neurons with active dendrites and dynamic synapses. The researchers applied the model to recognizing words such as *bat*, *tab*, *cab*, and *cat*. Each word was expressed as a sequence of phonemes, (for example, *c* → *a* → *t* for *cat*. A spike train of a single input neuron encoded each phoneme of a word, with a 100 ms delay between the onset times of successive phonemes. After training based on spike-timing-dependent plasticity, a single output neuron was able to detect a particular sequence of phonemes or identify a specific word. Their model could be an alternative, especially for the integration network ($N_{\text{ID}}$; Figure 5A), that detects the sequences of simple (linearly sweeping) FMs.

Sargolini and colleagues (2006) found evidence of conjunctive representation of position, direction, and velocity in the entorhinal cortex of rats that explored two-dimensional environments. In the medial entorhinal cortex (MEC), the network of grid cells constituted a spatial coordinate system in which positional information was represented. Head direction cells were responsible for head-directional information representation. Grid cells were co-localized with head direction cells and conjunctive (grid and head direction) cells, and the running speed of the rat modulated these cells. The researchers suggested that the conjunctive cells might update the representation of spatial location by integrating positional and directional and velocity information in the grid cell network during navigation. Such a conjunctive representation may be an alternative for the spatiotemporal representation of monkey calls, in which the information about spectral components, sweep rates and sweep directions and their combinatorial information may be represented by distinct types of cells in the primary auditory cortex.
In humans, it has been suggested that a voice contains information about not only a speech but also an “auditory face” which allows us to identify individuals (Belin, Fecteau, & Bedard, 2004). This is called auditory face perception and is processed based on a neurocognitive scheme similar to that proposed for visual face perception. Among vocal components for human speech processing, formants (Fitch, 1997) and syllables (Belin & Zatorre, 2003) might be candidate components used for identifying individuals. We suggest that monkey call identification may also be a kind of auditory face perception making use of FM components.

There has been evidence that the inferior colliculus (IC) encodes spectrotemporal acoustic patterns of species-specific calls. For example, Suta, Kvasnak, Popelar, and Syka (2003) investigated the neuronal representation of specific calls in the IC of guinea pigs. Responses of individual IC neurons of anesthetized guinea pigs to four typical calls (purr, chatter, chirp, and whistle) were recorded. A majority of neurons (55% of 124 units) responded to all calls. A small portion of neurons (3%) responded to only one call or did not respond to any of the calls. A time-reversed version of calls elicited on average a weaker response. The researchers concluded that the IC neurons do not respond selectively to specific calls but encode spectrotemporal acoustic patterns of the calls.

Maki and Riquimaroux (2002) recorded responses of IC neurons of gerbils to two distinct FM sounds that have the same spectral components (5–12 kHz) with an opposite sweep direction, or an upward sweep and a downward sweep. The upward FM generated much stronger responses than the downward FM. The researchers suggested that the directional selectivity to the FM sweeps implies that the IC may encode spectrotemporal acoustic patterns of species-specific calls.

These experiments imply that the encoding of spectrotemporal acoustic images of specific calls takes place, in part, in the IC, which presumably makes the spatiotemporal pattern of neuronal activation more complex in the present cochleotopic map. Hopefully, we will know details of it in the near future.

5 Conclusion

In this study, we have proposed a cochleotopic map similar to the retinotopic map in vision. When the cochleotopic ($N_{CO}$) network was stimulated with a monkey call, the peripheral neurons located on the frequency axis were sequentially activated. The active area moved along the propagation axis, by which the information about the call was mapped as a spatiotemporal firing pattern in the cochleotopic network dynamics. This spatiotemporal conversion was quite effective for the $N_{FM}$ network to decompose the call information into simple (linearly sweeping) FM components, by which the higher network ($N_{ID}$) was able to integrate these components into a unified percept, or to identify the call.
We suggest that the information about monkey calls could be mapped on a cochleotopic cortical network as spatiotemporal firing patterns, which can then be decomposed into simple (linearly sweeping) FM components and integrated into unified percepts by higher cortical networks. The spatiotemporal conversion of auditory information may be essential for developing the cochleotopic map, which could subserve as the foundation for later processing, or monkey call identification by higher cortical areas.

Acknowledgments

I am grateful to Yuishi Iwasaki for productive discussions and to Hiromi Ohta for her encouragement throughout the study. I am also grateful to the reviewers for giving me valuable comments and suggestions on the earlier draft.

References


Received January 19, 2006; accepted June 9, 2006.