Vowel Representations in the Ventral Cochlear Nucleus of the Cat: Effects of Level, Background Noise, and Behavioral State

BRADFORD J. MAY, GLENN S. LE PRELL, AND MURRAY B. SACHS
Departments of Otolaryngology–Head and Neck Surgery and Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, Maryland 21205

May, Bradford J., Glenn S. Le Prell, and Murray B. Sachs. Vowel representations in the ventral cochlear nucleus of the cat: effects of level, background noise, and behavioral state. J. Neurophysiol. 79: 1755–1767, 1998. Single-unit responses were studied in the ventral cochlear nucleus (VCN) of cats as formant and trough features of the vowel /ɛ/ were shifted in the frequency domain to each unit’s best frequency (BF; the frequency of greatest sensitivity). Discharge rates sampled with this spectrum manipulation procedure (SMP) were used to estimate vowel representations provided by populations of VCN neurons. In traditional population measures, a good representation of a vowel’s formant structure is based on relatively high discharge rates among units with BFs near high-energy formant features and low rates for units with BFs near low-energy spectral troughs. At most vowel levels and in the presence of background noise, chorus units exhibited formant-to-trough rate differences that were larger than VCN primary-like units and auditory-nerve fibers. By contrast, vowel encoding by primary-like units resembled auditory nerve representations for most stimulus conditions. As is seen in the auditory nerve, primary-like units with low spontaneous rates (SR <18 spikes/s) produced better representations than high SR primary-like units at all but the lowest vowel levels. Awake cats exhibited the same general response properties as anesthetized cats but larger between-subject differences in vowel driven rates. The vowel encoding properties of VCN chopper units support previous interpretations that patterns of auditory nerve convergence on cochlear nucleus neurons compensate for limitations in the dynamic range of peripheral neurons.

INTRODUCTION

Representations of speech-like stimuli in the auditory nerve and cochlear nucleus have been the subject of numerous studies during the past 20 yr (Blackburn and Sachs 1990; Geisler 1988; Kiang and Moxon 1974; Palmer et al. 1986; Sachs and Young 1979; Sinex and Geisler 1983; Young and Sachs 1979). In both the auditory nerve and cochlear nucleus, the representation of the spectra of vowel sounds has been analyzed in terms of average discharge rate (Blackburn and Sachs 1990; Sachs and Young 1979), the temporal patterns of firing (phase-locking) (Blackburn and Sachs 1990; Young and Sachs 1979), or a combination of the two (Young and Sachs 1979). This report focuses on average rate representations.

Previous population studies (Blackburn and Sachs 1990; Sachs and Young 1979) have evaluated discharge rate representations of vowel stimuli by plotting vowel-driven rates of individual units as a function of best frequency (BF; the frequency of greatest sensitivity for each unit). The resulting rate profiles are assumed to provide a good representation of the vowel’s formant structure if there are peaks in driven rates at BFs near formant frequencies and troughs in the discharge rates at BFs near spectral troughs. A difficulty with the population method is that a good estimate of vowel representation requires a large number of units with BFs spanning the frequency range of the stimulus spectrum.

Le Prell et al. (1996) introduced a method for estimating population rate profiles from the responses of individual auditory-nerve fibers. With their spectrum manipulation procedure (SMP), response profiles are created by recording the discharge rates of individual fibers as important spectral features of complex sounds are shifted to the fiber’s BF by changing the playback rate of digitized stimuli. This sampling technique is especially useful for studies of the central auditory system, where it is difficult to obtain large populations of units within the same animal. Our present study applies SMP methods to units in the ventral cochlear nucleus (VCN) of the cat to investigate the effects of stimulus level and background noise on neural representations of the vowel /ɛ/ (as in ‘get’). To evaluate the effects of anesthetic state, experiments were performed in both barbiturate-anesthetized and awake cats. Signal detection analyses of the responses evoked by vowel sounds suggest that the dynamic range of neural representations is enhanced even at the lowest levels of central auditory processing by patterns of auditory nerve convergence. Under most stimulus conditions, the best vowel representations were provided by VCN chopper units and primary-like units (Pri) with low spontaneous rate.

METHODS

The following methods are in accordance with guidelines for the use of animals in biomedical research that have been established by the Society for Neuroscience and the American Veterinary Medical Association. All procedures have been reviewed and approved by the Animal Care and Use Committee of the Johns Hopkins School of Medicine.

Acute recording procedures and response classification

General techniques for recording vowel representations in the VCN of anesthetized cats have been described in detail in a previous publication (Blackburn and Sachs 1990). Cats initially were anesthetized with intramuscular injections of xylazine (0.5 mg/kg) and ketamine (25 mg/kg). Atropine (0.05 mg/kg) was given to reduce secretions. Body temperature was monitored by a rectal probe and regulated with a heating pad. An in-dwelling catheter was placed in the cephalic vein to allow intravenous injections of barbiturate anesthesia (pentobarbital sodium) as needed to keep the cat areflexic. The cat was tracheotomized, and the scalp incised along the midline; temporalis muscles were dissected to expose...
the dorsal skull, ear canals, and tympanic bullae. The bulla of the experimental ear was ventilated with a length of PE tubing. Both ear canals were transected and an otoscopic examination performed on the tympanic membranes to confirm that they were clear and intact. The cat was placed in a stereotaxic apparatus by inserting hollow ear bars into each ear canal. The ear bars also served as conducting tubes for delivering sounds to the experimental ear. A metal plate on the stereotaxic apparatus was attached to the anterior cranium with screws to hold the head in a 45° downward position; this orientation facilitated the surgical approach to the auditory brain stem. A fenestration was made in the parietal cranium. Dura was reflected, and the lateral cerebellum partially aspirated to reveal the cochlear nucleus. Platinum-iridium microelectrodes with platinized tips were positioned visually above the brain stem using a stereotaxic electrode holder and then advanced into the cochlear nucleus with a hydraulic micromanipulator.

Noise bursts were used as search stimuli to reveal the presence of auditory units along the electrode track. When units were encountered, the microelectrode was positioned to maximize the magnitude of action potentials from a single unit. If action potentials from a single unit were sufficiently well isolated to allow precise triggering of a spike discriminator, the unit’s BF and threshold were determined audiovisually. Units with BFs <5 kHz were classified according to the shape of the peristimulus time histogram (PSTH) and their degree of regularity using responses to 50-ms BF tone bursts. Detailed descriptions of these classification methods can be found in Blackburn and Sachs (1989).

This report is limited to the following four VCN unit types: sustained choppers (ChS) had a multimodal PSTH and constant interspike interval (ISI) during a tone burst or could show a linearly increasing ISI if the coefficient of variation (CV) was <0.3. CV is defined as the ratio of the standard deviation to the mean of the ISI (Young et al. 1988); therefore units with low CVs have a more regular discharge pattern. Transient choppers (ChT) were characterized as having a multimodal PSTH, an increasing ISI, and a CV that was >0.3. Pris had a PSTH that resembled the more irregular response pattern of primary afferents (i.e., auditory-nerve fibers). Some primary-like neurons show a deep notch (PN, primary-like with notch) after a precisely timed onset peak at high presentation levels, and this PSTH pattern has been associated with morphological characteristics of globular bushy cells (Bourk 1976; Rouiller and Ryugo 1984; Smith and Rhode 1989). When PSTHs are collected with low-frequency tones (i.e., frequencies of interest in the present study), the auditory nerve input to the b Bushy cell region of the VCN creates a large neurophonic response that makes it difficult to maintain good unit isolation near stimulus onset. Consequently, a notch feature cannot be obtained in many low-BF PN units. As an alternative classification system, a separate analysis is provided for Pri units with low (<18 spikes/s) versus high spontaneous rate (SR; >18 spikes/s). Previous studies (Blackburn and Sachs 1989; Bourke 1976) have suggested that most low-BF PN units have low SR, whereas most Pri units do not. Issues that are related to this interpretation will be presented in greater detail in Discussion.

Measuring vowel representations with the SMP

When response classification was completed, the unit’s rate representation of the vowel /e/ was measured at different stimulus levels and in the presence of continuous broadband noise. Noise levels were varied with vowel levels to maintain a constant signal-to-noise ratio (S/N) of 3 dB. S/Ns were calculated as the ratio of total power in the vowel to total power in the noise at frequencies <3 kHz. Vowel levels are specified as the sum of power in the first 30 harmonics (with spacing of 100 Hz) because negligible power is contributed by harmonics at higher frequencies; noise levels were computed by summing power across 30 contiguous 100-Hz bands at frequencies <3.0 kHz. Modifications of the stimulus spectra that were produced by frequency response characteristics of the acoustic system were included in these calculations.

The amplitude spectrum of a vowel with total power of 43 dB sound pressure level (SPL) is shown in Fig. 1A. The vowel was synthesized digitally using a cascade formant synthesizer with a sampling rate of 12.8 kHz (Klatt 1980). The fundamental frequency of the stimulus was 100 Hz, and formant frequencies were placed at the 5th, 17th, and 25th harmonics with bandwidths of 60, 90, and 200 Hz to approximate the formant structure of the adult male voice (Peterson and Barney 1952). Acute (closed-field) experiments in anesthetized cats used vowel stimuli that were fil-
tered digitally to replicate sound pressure transformation by the cat’s head and pinna (Wiener et al. 1966) so that results could be compared with data obtained from awake cats using free-field stimuli. The digitized waveforms were output from the D/A converter via antialiasing filters.

The SMP estimates the population representation of vowel stimuli by measuring discharge rates as key spectral features are shifted to the BF of individual units. Harmonics marked with symbols in Fig. 1A identify the three features of interest in the present study. These features have been selected because they can be used to measure formant-to-trough rate differences and thereby characterize the quality of the vowel’s auditory representation.

The amplitude spectrum of digitized stimuli was translated in the frequency domain by changing the playback sampling rate relative to a standard synthesis rate of 12.8 kHz. Figure 1B illustrates this procedure for a hypothetical unit with BF of 2.1 kHz. With the playback rate is set to 15.8 kHz, the second formant (F2) falls to 2.1 kHz (°). When the playback rate is increased further to 24.4 and then to 53.8 kHz, the trough feature (T1) and the first formant (F1) shift to the unit’s BF. Thus these three spectral manipulations simulate the response patterns of three different units, each tuned to a different spectral feature of the natural vowel.

Figure 1C shows driven rates (total discharge rates elicited during stimulus presentations minus SR) that were recorded from a ChS unit (BF = 2.1 kHz) at a vowel level of 43 dB; the PSTH of this unit is shown in Fig. 1D. Driven rates are plotted as a function of the “effective” BF of the unit for particular test features, as defined by Eq. 1

\[ BF_{\text{effective}} = BF_{\text{unit}} \times \frac{\text{Rate}_{\text{synthesis}}}{\text{Rate}_{\text{playback}}} \]  

Effective BF is the frequency of the test feature that is shifted to BF by changing playback rate; that is, the rate to the shifted stimulus feature (BF_{unit}) estimates the rate of a unit with BF at the frequency of the test feature in the unshifted vowel (BF_{effective}). This format creates a rate profile that resembles traditional population representations of vowel stimuli (Blackburn and Sachs 1990; Sachs and Young 1979). The degree to which these plots approximate actual rate profiles depends on the assumption that discharge rates do not change significantly when the fundamental frequency of the vowel is altered to shift formant features to unit BF. This interpretation is supported by our previous SMP studies, which have shown that auditory nerve representations of vowel stimuli do not show a BF dependency in the frequency range of vowel sounds except at very high sound levels (Le Prell et al. 1996; Wong et al. 1998). Consequently, the sample of VCN units in the present study is restricted to low-frequency BFs.

SMP measures in Fig. 1C provide an accurate indication of the driven rates elicited by test features but not by other components of the stimulus. Figure 1E illustrates a method for estimating responses at other stimulus frequencies. Here, the ChS unit’s driven rates are plotted as a function of the absolute level of each test feature. A line has been drawn through the F1 and T1 data points (---) and through F2 and T1 (-----); the equations for the two lines are shown in the figure. Assuming that driven rates for components at frequencies below T1 fall along the line specified by the F1/T1 equation, responses to these components can be estimated from the vowel’s amplitude spectrum and the linear relationship between driven rate and feature level; responses to components above T1 are estimated from the F2/T1 equation. Le Prell et al. (1996) show that these assumptions are consistent with data from auditory-nerve fibers. The pseudopopulation response derived from these estimates is shown in Fig. 1C (-----). Because estimated rate responses are based on a linear transformation, the predicted population response is a scaled version of the vowel’s amplitude spectrum. This linear transformation is only measured at frequencies that fall between F1 and F2, and actual SMP results are accompanied by estimated rate profiles only for purposes of visualization when driven rates are plotted as a function of effective BF in Figs. 2B, 8B, and 10, C–F.

**Signal detection methods**

Signal detection methods offer quantitative measures of sound detection and discrimination (Green and Swets 1974) that can be applied to stimulus-based differences in single-unit discharge rates (Young and Barta 1986). In the present study, \( d' \) values were calculated for F1/T1 and F2/T1 formant-to-trough rate differences according to Eq. 2

\[ d' = \frac{\text{DR}_{\text{F1}} - \text{DR}_{\text{T0}}}{\sqrt{\text{SD}_{\text{F1}}^2 + \text{SD}_{\text{T0}}^2}} \]  

\( \text{DR}_{\text{F1}} \) is the mean and \( \text{SD}_{\text{F1}} \) is the standard deviation of driven rates for effective BFs at F1 or F2, which were estimated from responses to 20–50 stimulus presentations; \( \text{DR}_{\text{T0}} \) and \( \text{SD}_{\text{T0}} \) refer to the same rate statistics for effective BFs at T1. In this context, \( d' \) is a \( z \) transformation reflecting the statistical significance of the formant-to-trough depth in the SMP rate profile.

**Chronic recording procedures**

Detailed descriptions of our methods for chronic single-unit recording in the VCN of behaving cats can be found in previous publications (May and Sachs 1992; May et al. 1991). Cats were prepared for chronic recording by implanting the dorsal surface of the skull with a titanium connector and recording chamber. The connector was used to restrain the subject’s head during recording, and the chamber directed microelectrodes through the intact cerebellum to the VCN. Testing was performed inside an operant cage that was located near the center of a sound-attenuating booth (Industrial Acoustics). The subject’s body was restrained partially in a canvas bag, and the head was immobilized completely by at-
taching the head-restraint device to the framework of the cage. Restraint procedures protected the subject from injury while the electrode was in place and also ensured a standardized head orientation relative to the free-field speaker that served as the source of acoustic stimuli. The speaker was placed above and in front of the head where the cat’s head-related transfer function (HRTF; the transfer function describing the transformation of free-field sounds to energy at the tympanic membrane) is not marked by sharp peaks and notches (Rice et al. 1992). Vowel levels were measured by placing a microphone at the standardized location of the head without the subject in the testing apparatus. This method of calibration does not compensate for pinna-based amplification of low-frequency free-field sounds. Consequently, when data for awake cats are presented (Figs. 8 and 9), effective vowel levels at the tympanic membrane were likely to have been 10–15 dB higher than specified stimulus values because of the low-frequency gain of the HRTF (Wiener et al. 1966).

Cats initiated behavioral trials by holding down on a lever that was located on the floor of the cage near the right front paw. Testing began with the presentation of a random number of standard vowels (10–20 bursts) and ended with a 2-s interval of comparison vowels. Standard vowels were pulsed at a slow 1-Hz repetition rate (200 ms on, 800 ms off); comparison vowels were presented at a fast 4-Hz repetition rate (50 ms on, 200 ms off). Cats obtained food rewards by releasing the lever when they detected the change in the temporal pattern of vowel presentations. If the lever was released during the presentation of standard vowels or was not released during the comparison vowel interval, the trial was aborted and the cat was required to hold through another variable foreperiod before receiving the next food reward. This procedure was sufficient to produce near-perfect performance in all subjects.

A stereotaxic electrode holder was attached to the recording chamber to guide platinum-iridium microelectrodes into the VCN. Preferred recording sites were established by replicating the PSTH response patterns that have been attributed to VCN neurons in previous studies of anesthetized cats (Blackburn and Sachs 1989; Bourk 1976; Pfeiffer 1966; Rhode et al. 1983; Rouiller and Ryugo 1984; Smith and Rhode 1987). When auditory units were encountered, methods of unit classification and SMP testing were the same as those for acute recording. To increase the rate of data acquisition, single-unit responses were recorded with the cat sitting quietly in restraint. The behavioral task was activated between units or at regular intervals during long sampling periods to confirm that the subject remained awake and attending to auditory stimuli. Recording sessions were performed 3 days/wk and usually lasted ~2 h. Cats were studied for ~2 mo and then euthanized by barbiturate overdose followed by transcardial perfusion. Light microscopic examination of serial sections through the brain stem (cresyl violet stain) confirmed electrode penetrations in the VCN.

**RESULTS**

Results in anesthetized subjects are based on 58 units that were recorded in 11 cats. For testing in quiet (7 cats), the distribution of response types was as follows: 12 ChS units, 4 ChT units, 5 low SR Pri units, and 7 high SR Pri units. For testing in noise (4 cats), the sample consisted of 12 ChS units, 9 ChT units, 5 low SR Pri units, and 4 high SR Pri units. Results in two awake cats are based on 44 units (14 ChS/T, 14 low SR Pri, 16 high SR Pri), which were studied only under quiet conditions. Although units were very well isolated in awake cats, slight movement artifacts and subject-generated noise affected stimulus control and the precise timing of unit triggering in awake cats making ISI statistics more variable than in anesthetized cats. Under these recording conditions, classification according to rigorous criteria established by Blackburn and Sachs (1989) was not possible. For this reason, chopper responses are presented using the combined category ChS/T. BFs ranged from 0.43 to 4.28 kHz for studies of anesthetized cats and from 0.64 to 5.63 kHz for awake cats. Previous SMP studies (Le Prell et al. 1996) have revealed no obvious effects of BF on the quality of discharge rate representations across this frequency range.

**Effects of stimulus level**

The effects of stimulus level on vowel representations in the VCN of an anesthetized cat are illustrated in Fig. 2 with results from the same ChS unit (BF = 2.1 kHz) that was used in Fig. 1 to describe the SMP method. In Fig. 2A, driven rate is plotted versus feature level for vowel levels from 23 to 93 dB; the three data points for each vowel level are shown by the same symbol. Driven rates for the same test feature are connected across different vowel levels by cubic spline fits (lines). Notice how the resulting sensitivity functions diverge at feature levels between 15 and 20 dB SPL. The function for the T1 feature saturates at a driven rate of ~250 spikes/s, whereas the functions for both formant features achieve maximum rates exceeding 450 spikes/s.

Similar differences in rate saturation between formant and trough features have been demonstrated for auditory-nerve fibers with low and medium rates of spontaneous activity and are interpreted as suppression of responses to the low energy T1 feature by higher energy at surrounding formant frequencies (Le Prell et al. 1996; Sachs and Young 1980). Much of this suppression is likely due to energy in the first formant, which represents the most intense concentration of energy in the vowel. In situations where the suppressor occurs at frequencies below BF, as it does when F1 suppresses discharge rates that are elicited with the T1 feature at BF, the lower threshold boundary of the suppression area is nearly absolute in level regardless of the tuning characteristics and threshold of individual auditory-nerve fibers (Schmidt 1982). This boundary occurs well above the dynamic range of the more sensitive high SR auditory-nerve fibers, which as a result show little suppression of T1 responses (Sachs and Abbas 1979; Sachs and Kiang 1969).

Figure 2B plots data from the representative ChS unit as a function of effective BF. Lines reflect the pseudopopulation response that was estimated using the methods described in Fig. 1. These predicted responses give a fuller impression of the effects of stimulus level on the quality of rate representations. Notice how the difference in maximum driven rates between T1 and both formant features preserves the rate representation of the vowel’s formant structure at vowel levels reaching 93 dB (▲).

Figure 3 provides a statistical description of vowel-driven rates at a 43-dB stimulus level. The points show average driven rates computed across all units of a given response type; error bars indicate ±SE of mean values. The range of variability shown under this stimulus condition is typical of responses observed at other vowel levels. ChS units (Fig. 3A) exhibited the largest variation in formant-driven activity at this stimulus level, but in general, variation in discharge rates was relatively small even though data are combined...
likely that suppression effects arising in the auditory periphery contribute to the reduced trough-driven response and the term “suppression” will be used to describe these phenomena in the present study. Although ChT (Fig. 4B) and low SR Pri units (Fig. 4C) also exhibited relatively low maximum trough-driven rates, most high SR Pri units (Fig. 4D) showed little differences in trough-driven and formant-driven responses at high vowel levels (△).

Sensitivity functions in Fig. 4 point out differences in thresholds between chopper and primary-like units. Although no unit type responded to the low level T1 feature at a vowel level of 23 dB (leftmost ○), both ChS and ChT units showed strong responses to the F2 and F1 features (middle and right ○). Primary-like units, in particular the low SR Pri units, were minimally driven by formant features at this low vowel level.

The formant- and trough-driven saturation rates of chopper units were also clearly higher than those of primary-like units. ChS units displayed maximum F1-driven rates ç350 spikes/s and T1-driven rates of 200 spikes/s, whereas high SR Pri responses reached driven rates only slightly above 100 spikes/s with either formant or trough features at BF. As a result of these differences in minimum to maximum driven rates, ChS and ChT units exhibited a steep slope across all vowel levels; whereas responses of primary-like units appeared to have a more restricted dynamic range. At low vowel levels, the data of low SR Pri units displayed a shallow slope because these units were not sensitive enough to respond to the test features. At high vowel levels, responses of high SR Pri units to all three features saturated at about the same driven rate producing a flat sensitivity function. Similar differences in sensitivity and response across several subjects. In each plot, a line has been drawn through F1 and T1 data points (— —) and F2 and T1 (— —). The slope (in spikes/db) for each line is determined by the magnitude of rate differences between formant and trough driven rates and therefore provides an indication of sensitivity to amplitude differences in the vowel’s spectrum at a fixed overall level. At 43 dB, ChS units produced the steepest slopes; that is, the greatest sensitivity to amplitude differences, whereas high SR Pri units displayed relatively shallow slopes. Sensitivity functions of chopper units fall close to a single straight line because the differences in rates to F1 and F2 are not limited by rate saturation. In the case of the primary-like units, the F2/T1 slope is less than the F1/T1 slope, presumably because the F2 level is low enough to reflect the generally concave upward shape of rate-level functions near threshold.

Effects of threshold and rate saturation are more clear in Fig. 4, which shows average sensitivity functions for vowel levels from 23 to 83 dB; the slopes of these functions are given in Table 1. At 23 dB, chopper units show the threshold effects seen at 43 dB for primary-like units (less sensitivity to F2/T1 level differences than to F1/T1). At 63 and 83 dB, F1 and F2 rates have saturated, making the slope of the F1/T1 sensitivity function less than that of F2/T1. As in Fig. 2, the T1 rates appear to saturate at lower values than the F1 or F2 rates. A difference in saturation rate is apparent even when trough and formant features are of approximately equal absolute level. For example, compare the T1 response at 83 dB to the F1 response at 63 dB in Fig. 4A. Le Prell et al. (1996) noted similar patterns in the auditory nerve and attributed the reduced saturation rate at T1 to two-tone suppression. Additional processing factors probably play a role in shaping vowel representations in the VCN, but it is likely that suppression effects arising in the auditory periphery contribute to the reduced trough-driven response and the term “suppression” will be used to describe these phenomena in the present study. Although ChT (Fig. 4B) and low SR Pri units (Fig. 4C) also exhibited relatively low maximum trough-driven rates, most high SR Pri units (Fig. 4D) showed little differences in trough-driven and formant-driven responses at high vowel levels (△).

Sensitivity functions in Fig. 4 point out differences in thresholds between chopper and primary-like units. Although no unit type responded to the low level T1 feature at a vowel level of 23 dB (leftmost ○), both ChS and ChT units showed strong responses to the F2 and F1 features (middle and right ○). Primary-like units, in particular the low SR Pri units, were minimally driven by formant features at this low vowel level.

The formant- and trough-driven saturation rates of chopper units were also clearly higher than those of primary-like units. ChS units displayed maximum F1-driven rates ç350 spikes/s and T1-driven rates of 200 spikes/s, whereas high SR Pri responses reached driven rates only slightly above 100 spikes/s with either formant or trough features at BF. As a result of these differences in minimum to maximum driven rates, ChS and ChT units exhibited a steep slope across all vowel levels; whereas responses of primary-like units appeared to have a more restricted dynamic range. At low vowel levels, the data of low SR Pri units displayed a shallow slope because these units were not sensitive enough to respond to the test features. At high vowel levels, responses of high SR Pri units to all three features saturated at about the same driven rate producing a flat sensitivity function. Similar differences in sensitivity and response across several subjects. In each plot, a line has been drawn through F1 and T1 data points (— —) and F2 and T1 (— —). The slope (in spikes/db) for each line is determined by the magnitude of rate differences between formant and trough driven rates and therefore provides an indication of sensitivity to amplitude differences in the vowel’s spectrum at a fixed overall level. At 43 dB, ChS units produced the steepest slopes; that is, the greatest sensitivity to amplitude differences, whereas high SR Pri units displayed relatively shallow slopes. Sensitivity functions of chopper units fall close to a single straight line because the differences in rates to F1 and F2 are not limited by rate saturation. In the case of the primary-like units, the F2/T1 slope is less than the F1/T1 slope, presumably because the F2 level is low enough to reflect the generally concave upward shape of rate-level functions near threshold.

Effects of threshold and rate saturation are more clear in Fig. 4, which shows average sensitivity functions for vowel levels from 23 to 83 dB; the slopes of these functions are given in Table 1. At 23 dB, chopper units show the threshold effects seen at 43 dB for primary-like units (less sensitivity to F2/T1 level differences than to F1/T1). At 63 and 83 dB, F1 and F2 rates have saturated, making the slope of the F1/T1 sensitivity function less than that of F2/T1. As in Fig. 2, the T1 rates appear to saturate at lower values than the F1 or F2 rates. A difference in saturation rate is apparent even when trough and formant features are of approximately equal absolute level. For example, compare the T1 response at 83 dB to the F1 response at 63 dB in Fig. 4A. Le Prell et al. (1996) noted similar patterns in the auditory nerve and attributed the reduced saturation rate at T1 to two-tone suppression. Additional processing factors probably play a role in shaping vowel representations in the VCN, but it is
TABLE 1. Slopes (sp/dB) of sensitivity functions for units tested under quiet conditions

<table>
<thead>
<tr>
<th>Level, dB</th>
<th>F1/T1 Contrast</th>
<th>F2/T1 Contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ChS</td>
<td>ChT</td>
</tr>
<tr>
<td>23</td>
<td>7.17</td>
<td>4.62</td>
</tr>
<tr>
<td>33</td>
<td>8.34</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>9.19</td>
<td>7.45</td>
</tr>
<tr>
<td>53</td>
<td>7.93</td>
<td>5.67</td>
</tr>
<tr>
<td>63</td>
<td>5.01</td>
<td>4.10</td>
</tr>
<tr>
<td>73</td>
<td>6.40</td>
<td>6.10</td>
</tr>
<tr>
<td>83</td>
<td>5.21</td>
<td>4.67</td>
</tr>
<tr>
<td>93</td>
<td>3.83</td>
<td></td>
</tr>
</tbody>
</table>

ChS, sustained choppers; ChT, transient choppers; LSR Pri, low spontaneous rate primarylike units; HSR Pri, high spontaneous rate primarylike units; F1, first formant; T1, trough feature; F2, second formant.

magnitude were noted between low and high SR auditory-nerve fibers by Le Prell et al. (1996). To illustrate the transformation of auditory nerve rate information by VCN units, the trapezoid in each panel of Fig. 4 indicates the range of auditory nerve sensitivity functions at a vowel level of 63 dB. As shown in Fig. 4A, the upper edge of the trapezoid is determined by the discharge rates of high SR fibers, whereas the lower edge reflects discharge rates of low SR fibers. These responses were calculated by applying linear models published in May et al. (1996) to F1, T1, and F2 feature levels. Both classes of chopper units showed higher driven rates and larger formant-to-trough rate differences than auditory-nerve fibers. Vocal-driven rates of primary-like units, on the other hand, often fell within the range of auditory nerve responses.

Effects of background noise

Figure 5 shows VCN sensitivity functions that were obtained in continuous broadband noise (S/N = 3 dB); the slopes of these functions are given in Table 2. In the figure, average rate differences are shown for all sampled units within the four response types. These measures were calculated by subtracting each unit’s driven rate for noise alone from its rate to the vowel-in-noise. Because higher noise-driven rates were associated with higher levels of background noise, smaller rate differences were seen at higher vowel/noise levels. Rate differences below zero mean that responses to vowels in noise were less than responses to noise alone. These negative rate differences were observed only when the T1 feature was placed at BF and probably reflect, at least in part, suppression of the trough-in-noise response by energy at F1 and F2, which is seen in the auditory nerve inputs to these units (Le Prell et al. 1996; Sachs et al. 1983). For ChS, ChT, and, to a lesser extent, low SR Pri units, such suppression-like effects grew in magnitude at higher vowel levels and offset decreasing rate differences for formant features. As a result, F1/T1 (— — —) and F2/T1 functions (— — —) show a relatively constant slope across vowel levels from 33 to 73 dB. As was seen for vowels in quiet, high SR Pri units failed to exhibit suppression-like effects for vowels in noise and consequently showed minimal formant-to-trough rate differences at high stimulus levels.

Signal detection analysis

To this point, the quality of vowel encoding has been inferred from rate difference measures. From a signal detection perspective, the detectability of a rate difference can only be judged in reference to the standard deviation of the discharge rates on which it is based (Eq. 2). Thus large but highly variable rate differences convey little information. The SDs of ChS and ChT units are plotted as a function of total rate in Fig. 6 (A and B). Each symbol is the response at a vowel level of 63 dB. As shown in Fig. 4A, the upper edge of the trapezoid is determined by the discharge rates of one unit; several data points come from the same neuron.
of their responses to complex, periodic low-frequency sounds requires further analysis that was not performed in the present series of experiments. These functions also point out that formant features could elicit total rates approaching 500 spikes/s for ChS units, while total rates of ChT units never reached 400 spikes/s.

Formant-to-trough rate differences for chopper units (Figs. 4 and 5) were combined with SD measures using Eq. 2 to produce the \( d' \) statistics in Fig. 6 (C–F). At most vowel levels in quiet, ChS units exhibited larger \( d' \) values than ChT units for both F1/T1 (Fig. 6C) and F2/T1 contrasts (Fig. 6D) even though these units produced smaller formant-to-trough rate differences (Fig. 4, A and B). Figure 6 also shows \( d' \) values of auditory-nerve fibers (filled regions) using data from May et al. (1996). As indicated in Fig. 6C, the upper bound of each filled region was determined by \( d' \) values for low SR fibers, which typically provide the best auditory nerve representation of vowel sounds; similarly the lower bound of each filled region reflects the detectability of representations provided by high SR fibers.

Vowel-encoding by both ChS and ChT units was as good or better than in the auditory nerve at all stimulus levels. Even though chopper units showed large SDs, they produced larger \( d' \) values than auditory-nerve fibers because of their large formant-to-trough rate differences. \( d' \) values for ChS units are generally greater than those for ChT units because of the smaller variability in ChS driven rates. Both VCN response types displayed \( d' \) values for the F2/T1 contrast that were quite similar to those of F1/T1 even though the level of the F1 formant was 8.5 dB higher. This similarity is seen at low vowel levels because lower F2-driven rates are associated with lower SDs and at high levels because F1 and F2 saturation rates are nearly equal (Fig. 4, A and B). Both VCN units and auditory-nerve fibers exhibited a trend toward decreasing \( d' \) values at higher vowel levels.

Differences in the quality of vowel representations between chopper and low SR auditory-nerve fibers were more apparent for testing in background noise. Although F1/T1 rate differences (Fig. 6E) were more detectable than F2/T1 differences (Fig. 6F), both formant-to-trough contrasts elicited similar results. That is, ChS units provided the best representation of formant features across all vowel levels, whereas ChT representations were consistently better than those of auditory-nerve fibers. In signal detection theory, a \( d' \) value of 1 is the most often cited criterion for detectability. Although \( d' \) dropped sharply at high vowel levels, VCN representations for both F1/T1 and F2/T1 remained well above this criterion in the presence of background noise, as did \( d' \) values of auditory-nerve fibers for F1/T1.

SDs of low and high SR Pri units are shown in Fig. 7 (A and B). The power function indicating the effects of total rate on the average SD of low SR Pri units (——) was slightly above that of auditory-nerve fibers (——) but not substantially different from that of ChS and ChT units. No power function is shown for high SR Pri units because classification criteria preclude data at SRs <18 spikes/s; however, SDs for most units follow the auditory nerve power function. The exception to this relationship is a cluster of units that displayed higher SDs at total rates between 50 and 100 spikes/s (→). Total rates for Pri units in both SR classes were usually <200 spikes/s. 

\( d' \) values based on the F1/T1 and F2/T1 rate differences of Pri units are shown in the remaining panels of Fig. 7 for testing in quiet (C and D) and in background noise (E).
Vowel representations in anesthetized cats have been described in terms of average driven rates among units with similar PSTH response properties. Sample averages for different unit types were created by combining data across several cats. An analysis of this nature is not as straightforward for results obtained in awake cats because the two subjects (cats 1Ma and 2Mi) exhibited very different driven rates for units in the same response class. An example of this between-subject variability is shown in Fig. 9. Sensitivity functions for ChS/T units are plotted in Fig. 9 (A and B). Both sets of data display the T1 suppression effects that were associated with ChS and ChT units in anesthetized cats (Fig. 4, A and B), but cat 2Mi exhibited much lower driven rates than cat 1Ma or anesthetized subjects. Differences in responsivity between cats were even more pronounced for low SR Pri units (Fig. 9C), where virtually no overlap was seen in the vowel-driven rates of cat 1Ma (○, □, ◊) and 2Mi (●, ●, ▲, ▲). High SR Pri units in cat 2Mi (Fig. 9D) showed extended dynamic range properties at high vowel levels relative to responses in anesthetized cats (Fig. 4D); average sensitivity functions are not shown for cat 1Ma since only 1 high SR Pri unit was studied at multiple vowel levels.

DISCUSSION

Among the unit types evaluated in our study, the driven rates of chopper units provided the best overall representation of vowel formant structure at high and low stimulus levels in quiet (Fig. 4A) and in the presence of background noise (Fig. 5A). A signal detection analysis indicated that vowel encoding by VCN chopper units was superior to auditory nerve representations under most stimulus conditions (Fig. 6). It also was noted that the vowel encoding properties of Pri units could be classified according to spontaneous rate, with low SR Pri units typically providing a better representation of formant structure than high SR Pri units. Be-
sMP studies because suppression effects create a saturated T1-driven rate that is less than maximum formant-driven rates (Le Prell et al. 1996). ChT units exhibited a deeper T1 trough than ChS units and therefore may be more sensitive to suppression effects. The good agreement between the two sampling techniques supports the validity of sMP methods for studies of vowel encoding in the central auditory system at least at the level of the ventral cochlear nucleus.

Figure 10 (E and F) shows average sMP rate profiles of low and high SR Pri units at several vowel levels. In comparison with chopper units, Pri units showed good vowel representations over a narrower range of levels. The dynamic range of low SR Pri units was limited at low vowel levels (●) by their high-threshold, whereas the more sensitive high SR Pri units (with flat saturation) produced a rate profile lacking peaks and troughs at high vowel levels (△). The same response patterns are observed for low and high SR auditory-nerve fibers (May et al. 1996) (Figs. 4 and 6), and these similarities presumably arise from the powerful synaptic coupling between small numbers of auditory-nerve fibers and VCN Pri units (Ryugo and Sento 1991). Chopper units, on the other hand, receive highly convergent inputs from auditory-nerve fibers (Cant 1981) and appear to be

Auditory nerve input to stellate cells in the ventral cochlear nucleus

It has been suggested that a rate representation that is robust over a broad range of stimulus levels could be formed
by differentially weighting the discharge rates of low and high SR auditory-nerve fibers (Delgutte 1982). This hypothetical neural circuit is supported by response properties of VCN chopper units presented in Fig. 10, A–D, and by auditory nerve innervation patterns of stellate cells in the anteroventral cochlear nucleus (Cant 1981).

Chopper responses presumably are recorded from stellate cells in the VCN (Bourk 1976; Rhode et al. 1983; Rouiller and Ryugo 1984). Although details of the innervation patterns of stellate cells are not known, electron microscopic studies suggest two principle patterns of input to the soma and proximal dendrites of large multipolar cells (Cant 1981). Type I stellate cells show few synaptic contacts on their soma and many contacts on their proximal dendrites, whereas type II stellate cells receive a heavier innervation covering 70% of the somatic surface. By combining intracellular labeling with electrophysiological recording, Smith and Rhode (1989) have shown that type I stellate cells are chopper units with sustained responses (ChS/ChT), whereas type II stellate cells produce a brief chopping pattern at stimulus onset then cease responding (onset C). No onset C neurons were recorded in the present study.

Given the anatomic arrangement just described, the response patterns of ChS and ChT units (i.e., the type I stellate cells) are likely to be governed by the balance of their somatic and dendritic inputs. Although no single explanation can account for the rich complexity of excitatory and inhibitory influences on chopper units, aspects of their vowel encoding properties can be explained by differences in the location of excitatory inputs from low and high SR auditory-nerve fibers. The selective listening model introduced by Banks and Sachs (1991) and later elaborated on by Lai et al. (1994a,b) proposes that distal inputs of stellate cells are dominated by high SR auditory-nerve fibers and proximal inputs by low SR fibers. In support of this interpretation, Liberman (1991) traced physiologically characterized auditory-nerve fibers into the anteroventral cochlear nucleus and found that stellate cell bodies lie in close apposition with projections from low SR auditory-nerve fibers (see also, Fekete et al. 1984; Rouiller et al. 1986; Ryugo and Rouiller 1988).

The influence of high SR auditory nerve inputs to the distal dendrites of ChS units may explain why these neurons have excellent sensitivity at the lowest vowel levels but also exhibit less suppression of T1-driven rates than ChT units at moderate vowel levels (Fig. 4, A and B). It is also likely that ChS responses are shaped strongly by proximal inputs from low SR auditory-nerve fibers because these units maintain a clear representation of formant-to-trough amplitude differences at high stimulus levels (Fig. 4A) and in background noise (Fig. 5A), which is not the case for high SR auditory-nerve fibers (May et al. 1996). The selective listening model proposes that a shunting inhibition is applied to the proximal dendrites of chopper units to eliminate high SR inputs at high stimulus levels. Along this same line of reasoning, ChT units may exhibit higher thresholds and greater peak-to-trough rate differences than ChS units (Fig. 4B) because low SR inputs play a more dominant role in their responses to vowel sounds. If auditory nerve innervation patterns are similar for ChS and ChT units, the ChT bias toward low SR inputs could be achieved by a stronger inhibitory input, as previously observed by Blackburn and Sachs (1992).

Auditory nerve inputs to bushy cells in the ventral cochlear nucleus

Previous descriptions of vowel encoding by VCN primary-like neurons (Blackburn and Sachs 1990; Palmer et al. 1986) have emphasized temporal representations that can be derived from phase-locked responses to harmonic components of the vowel’s amplitude spectrum (Young and Sachs 1979). Temporal responses of primary-like neurons resemble those of auditory-nerve fibers. This similarity is not surprising given the powerful synaptic coupling between the auditory nerve and VCN bushy cells, which are the putative anatomic source of primary-like activity in the cochlear nucleus (Bourk 1976; Rhode et al. 1983; Rouiller and Ryugo 1984; Smith and Rhode 1987).

The creation of interpretable rate profiles was made difficult in earlier population studies of primary-like units by the heterogeneity of their discharge rates (Blackburn and Sachs 1990). Normalizing vowel-driven rates by maximum rates for BF tones can reduce this inherent variability in populations of auditory-nerve fibers (Sachs and Young 1979), but such methods of normalization do not work well in the cochlear nucleus where primary-like units differ greatly in their relative responsivity to vowels and tones (see Blackburn and Sachs 1990, Fig. 7). The SMP methods used in the present study demonstrate that primary-like units are capable of providing very good rate representations of vowel-like stimuli if the neural representation is viewed in terms of how individual neurons change their discharge rates when tested with different formant structures. Similar rate difference measures have proven to be a sensitive indicator of the quality of auditory nerve representations for vowels (Conley and Keilson 1995; Le Prell et al. 1996) as well as the complex high-frequency spectra of pinna-filtered noise bursts (Rice et al. 1995).

Low versus high SR primarylike units showed substantial differences in the quality of their vowel representations in quiet (Fig. 4, C and D) and in noise (Fig. 5, C and D). These differences also suggest functional interpretations that are based on patterns of auditory nerve inputs to the cochlear nucleus. Although auditory-nerve fibers from all SR groups converge on spherical bushy cell regions (Liberman 1991), morphological features of endbulbs of Held suggest that individual spherical cell soma are contacted by a small number of fibers, all within the same SR group (Ryugo and Sento 1991). If VCN primary-like units reflect these peripheral inputs in their own spontaneous activity, high SR Pri units may show sensitive vowel representations at lower stimulus levels because their responses are dictated by high SR auditory-nerve fibers. Low SR Pri units, on the other hand, may show their best vowel representations at high stimulus levels because of their high-threshold inputs from low SR fibers.

An alternative explanation for the response properties of low SR Pri units can be drawn from previous results of Blackburn and Sachs (1990). Among the units with BF <2 kHz in their database, 77% of PN units versus only 30% of Pri units exhibited SRs that were <18 spikes/s. Differences in the SR distributions of PN and Pri units may
be even larger than their numbers indicate because it is likely that some low SR PN units were classified incorrectly as Pri units because of PSTH sampling limitations at low frequencies. From this perspective, a majority of the low SR Pri units in the present study actually may be PN units.

The PN response type has been associated with the anatomic classification of globular bushy cells in the VCN (Bourk 1976; Smith and Rhode 1987). Relative to spherical bushy cells, the soma and proximal dendrites of globular bushy cells receive a larger number of auditory nerve terminals via modified (smaller) endbulb synapses (Rouiller et al. 1986); in addition, globular bushy cells may be innervated preferentially by high SR auditory-nerve fibers (Liberman 1991). Thus the enhanced dynamic range of PN units may arise from highly convergent inputs of high SR fibers with heterogeneous thresholds or different BFs.

The frequent occurrence of low SR among PN units has been noted by both Bourk (1976) and Blackburn and Sachs (1989); however, the conclusion that low SR Pri units are mainly PN units lacks complete consensus. Liberman (1991), for example, asserts that the correlation of low SRs among PN units may arise from sampling errors. Among our own observations, the relatively high vowel thresholds of low SR Pri units (note the lack of response to the 23 dB vowel in Fig. 4C) does not imply strong innervation by the more sensitive high SR auditory-nerve fibers, which are presumed to be the principal input to globular bushy cells (Liberman 1991). Furthermore, a subset of high SR Pri units in the present sample had PSTHs with clear PN characteristics. Whatever the source of the physiological differences among VCN primary-like units, the results presented here indicate that SR is an appropriate alternative classification method for low BF primary-like units.

**Limitations for SMP assessment of vowel representations in awake cats**

Awake cats displayed the same general patterns of vowel encoding as anesthetized cats, but substantial interanimal differences were noted in the magnitude of vowel-driven responses (Fig. 9). These results suggest that our chronic recording procedures may not provide an equally accurate assessment of the quality of vowel representations in all subjects. Differences in vowel-driven rates may arise from methodological problems inherent in the delivery of well-calibrated free-field sounds. The pinna of the cat is highly mobile and pinna movements are known to change the head-related transfer function of the cat (Young et al. 1996). Although it is possible that subtle pinna movements may have altered vowel spectra as stimuli propagated to the ear drum, it is not likely that acoustic phenomena were the sole source of major differences in overall response rates. Awake cats were video monitored during electrophysiological recording and gross differences in pinna orientation were not observed. Furthermore, pinna-filtering effects do not exhibit much directional sensitivity over frequencies <5 kHz at the frontal elevation where the free-field speaker was located in the present study (May and Huang 1997; Rice et al. 1992).

A more intriguing interpretation of the variability observed among awake cats is that vowel-driven rates may reflect differences in attentional state. Behavioral methods used in the present study did not explicitly focus the subject’s attention on spectral properties of acoustic stimuli because the task was based solely on the discrimination of temporal patterns that could be conveyed by pure tones, bursts of noise, or vowels (May et al. 1991). Our previous psychophysical experiments (Hienz et al. 1996a,b) have shown that the cat’s ability to discriminate formant frequency changes approaches that of human listeners (Kewley-Port and Watson 1994). Across a wide range of vowel levels and in the presence of background noise, these behavioral thresholds are slightly better than performances predicted by rate representations in the auditory nerve (May et al. 1996) but quite close to those predicted by the response properties of VCN chopper units in anesthetized cats (May et al. 1997). Therefore, although it is possible that vowel representations could be made less variable in awake cats by improving stimulus control and by focusing the subject’s attention on important spectral features, such representations are not likely to be fundamentally better than the response patterns described in the present study.

**Applying SMP methods to other structures and stimulus conditions**

The reliability of the spectrum manipulation procedure as a sampling method for vowel studies in the auditory nerve and cochlear nucleus has been confirmed by comparing SMP rate profiles with results obtained in previous population experiments. This validation in the auditory periphery notwithstanding, widespread application of SMP methods to other auditory structures and stimulus conditions should proceed with some caution in the absence of similar confirmatory measures.

Of particular concern is the sensitivity of higher order neurons to amplitude modulated (AM) stimuli. Synthetic vowels can create such AM effects when multiple harmonics interact within the auditory filter of individual neurons; manipulations of playback rate during SMP testing alters the pattern of AM by changing the frequency and spacing of harmonic components. Although auditory-nerve fibers and cochlear nucleus neurons show strong phase locking to AM (Møller 1972; Wang and Sachs 1993, 1994), their average discharge rates are relatively unaffected by changes in AM frequency (Joris and Yin 1992; Rhode 1994). By contrast, most neurons in the central nucleus of the inferior colliculus (ICC) achieve maximum firing rates at one best modulation frequency (BMF) (Langner and Schreiner 1988). Given the sharp band-pass characteristics of ICC temporal modulation transfer functions, SMP-based changes in AM temporal properties are likely to introduce rate fluctuations that will confound the patterns of activity created by shifting different formant features to BF. These stimulus artifacts are amplified by the perfectly periodic harmonic nature of synthetic vowels and presumably can be eliminated by performing SMP tests with digital recordings of natural speech sounds.

Another important question that remains to be answered is whether the SMP can be used to investigate auditory processing of high-frequency complex sounds. For example, the cat’s HRTF produces important spectral cues for sound localization at frequencies between 5 and 20 kHz (Huang and May 1996; Rice et al. 1992), which is well above the
high-frequency cutoff for the inclusion of unit BFs in the present study. It is not known how frequency-dependent changes in the bandwidth of neural tuning curves, absolute thresholds, and two-tone suppression will influence the reliability of the SMP testing method at these high frequencies. May and Huang (1997) recently have demonstrated linear rate-level relationships between HRTF-based spectral features and auditory nerve discharge rates. Their results suggest a general applicability of our current methods but more detailed parametric analyses of the assumptions underlying the SMP are needed for high-frequency stimuli, especially the assumption that responses are generated by the frequency component at BF alone (Calhoun et al. 1998).

The authors thank C. M. Aleszczyk and H. Jain for assistance with surgical procedures and behavioral training. E. Young and D. Ryugo provided constructive criticism during manuscript preparation.

This research was supported by National Institute on Deafness and Other Communication Disorders Grant 2 R01 DC-00109 to M. B. Sachs.

Address for reprint requests: B. J. May, Dept. of Otolaryngology-HNS, Johns Hopkins School of Medicine, 720 Rutland Ave., Traylor Building Room 505, Baltimore, MD 21205.

Received 1 October 1997; accepted in final form 2 January 1998.

REFERENCES


Ryugo, D. K. and Sento, S. Synaptic connections of the auditory nerve.