

# Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey

Mitchell Steinschneider<sup>a)</sup>

*Departments of Neurology and Neuroscience, Albert Einstein College of Medicine, Rose F. Kennedy Center, Bronx, New York 10461*

Yonatan I. Fishman

*Department of Neurology, Albert Einstein College of Medicine, Rose F. Kennedy Center, Bronx, New York 10461*

Joseph C. Arezzo

*Departments of Neurology and Neuroscience, Albert Einstein College of Medicine, Rose F. Kennedy Center, Bronx, New York 10461*

(Received 13 December 2002; revised 20 April 2003; accepted 25 April 2003)

Voice onset time (VOT) signifies the interval between consonant onset and the start of rhythmic vocal-cord vibrations. Differential perception of consonants such as /d/ and /t/ is categorical in American English, with the boundary generally lying at a VOT of 20–40 ms. This study tests whether previously identified response patterns that differentially reflect VOT are maintained in large-scale population activity within primary auditory cortex (A1) of the awake monkey. Multiunit activity and current source density patterns evoked by the syllables /da/ and /ta/ with variable VOTs are examined. Neural representation is determined by the tonotopic organization. Differential response patterns are restricted to lower best-frequency regions. Response peaks time-locked to both consonant and voicing onsets are observed for syllables with a 40- and 60-ms VOT, whereas syllables with a 0- and 20-ms VOT evoke a single response time-locked only to consonant onset. Duration of aspiration noise is represented in higher best-frequency regions. Representation of VOT and aspiration noise in discrete tonotopic areas of A1 suggest that integration of these phonetic cues occurs in secondary areas of auditory cortex. Findings are consistent with the evolving concept that complex stimuli are encoded by synchronized activity in large-scale neuronal ensembles. © 2003 Acoustical Society of America. [DOI: 10.1121/1.1582449]

PACS numbers: 43.64.Qh, 43.64.Sj, 43.71.Es [WPS]

## I. INTRODUCTION

Understanding how information relevant to speech processing is extracted from the neural signal is a cardinal goal of modern auditory neurophysiology. In this study, we examine neural representation of voice onset time (VOT). This important speech parameter is utilized by most of the world languages and signifies the interval between consonant release (onset) and the onset of periodic vocal-cord vibrations (voicing) (Lisker and Abramson, 1964). In American English, a short VOT promotes the perception of the voiced stop consonants /b/, /d/, and /g/, whereas a long VOT facilitates the perception of the unvoiced stop consonants /p/, /t/, and /k/. Differential perception of these phonemes is categorical, with the boundary for discriminating a voiced from an unvoiced stop consonant in syllable-initial position generally lying at a VOT of 20–40 ms.

Various features of speech perception are similar in humans and animals, indicating that phonetic processing relies in part on basic auditory system mechanisms (e.g., Kuhl and Miller, 1975, 1978; Kuhl and Padden, 1982; Sinnott and Adams, 1987; Kluender and Lotto, 1994; Dent *et al.*, 1997; Sin-

nott and Brown, 1997; Kluender *et al.*, 1998; Ohlemiller *et al.*, 1999; Holt *et al.*, 2001; Le Prell *et al.*, 2001). Animals show categorical perception of voiced versus unvoiced stop consonants with VOT boundary values comparable to those in humans, suggesting that neurophysiological experiments in animals may be directly relevant for investigating mechanisms involved in VOT encoding. Earlier work examining cortical responses to speech sounds in monkey primary auditory cortex (A1) found that syllables with a short VOT often evoked a single transient response at consonant onset, whereas syllables with a long VOT typically evoked two transient responses, one at consonant onset and another at voicing onset (Steinschneider *et al.*, 1994, 1995). These data led to the hypothesis that the perception of voiced from unvoiced stop consonants is partly determined by the difference between these two response patterns.

In this study, we examine the neural representation of stop consonant–vowel syllables with variable VOT in A1, using neuronal population measures in awake, naive old-world monkeys. The goal is to test the hypothesis that previously identified temporal response patterns reflecting the perceptual boundary in humans are maintained in large-scale population activity. Alternatively, these response patterns could be obscured in the overall population activity by other,

<sup>a)</sup>Electronic mail: steinsch@aecom.yu.edu

less specific response components evoked by the syllables, or could be restricted to specific regions of A1. Growing evidence that complex stimuli are encoded by synchronized activity in large-scale neuronal ensembles emphasizes the need to examine population responses sampled across a wide extent of auditory cortex (Creutzfeldt *et al.*, 1980; Gehr *et al.*, 2000; Rotman *et al.*, 2001; Nagarajan *et al.*, 2002; Petersen *et al.*, 2002). Maintenance of temporal response patterns differentially representing VOT in population activity could facilitate the rapid representation of stop consonants that occurs during perception of real-time streams of speech. Studies in awake animals are mandated by the marked effects of anesthesia on auditory cortical and medial geniculate responses to sound (Aitkin and Dunlap, 1968; Mäkelä *et al.*, 1990; Madler *et al.*, 1991; Zurita *et al.*, 1994; Gaese and Ostwald, 2001; Talwar and Gerstein, 2001). Results obtained in naive animals can serve as a benchmark for evaluating physiological changes that occur with learning (e.g., Recanzone *et al.*, 1993; Tremblay *et al.*, 1997; Weinberger, 1997; Ahissar *et al.*, 1998; Merzenich *et al.*, 1999; Blake *et al.*, 2002), a consideration especially relevant for modeling speech-evoked activity in young infants with limited language exposure.

Adequate assessment of speech-evoked activity requires that temporal response patterns be related to organizational features of A1. Thus, the relationship between patterns of activity and the tonotopic organization is examined to assess the degree to which specific response profiles are restricted to discrete best-frequency regions. Lamina-specific responses are also evaluated to determine whether temporal patterns are merely reflections of transmitted subcortical activity or represent newly generated cortical transformations, occur in laminae that project to secondary auditory cortical fields, and are preserved in later polysynaptic activity.

## II. METHODS

Six male macaque monkeys (*Macaca fascicularis*) weighing between 2.5 and 3.5 kg were studied following approval by our institutional Animal Care and Use Committee. Animals were housed in our AAALAC-accredited Animal Institute, and their health was monitored daily by the investigators and veterinary staff. Experiments were conducted in accordance with institutional and federal guidelines governing the use of primates. Animals were initially trained to sit comfortably with hands restrained in customized primate chairs. Surgery was then performed using sterile techniques and general anesthesia (sodium pentobarbital). Holes were drilled into the skull to accommodate epidural matrices that allowed access to the brain. Matrices consisted of 18-gauge stainless-steel tubes glued together into a honeycomb form. They were shaped to approximate the contour of the cortical convexity, covered with a protective layer of sterile silastic, and stereotaxically positioned to target A1 at an angle 30 deg from normal to approximate the anterior–posterior tilt of the superior temporal gyrus. This angle permitted electrode penetrations nearly orthogonal to the surface of A1, a requirement for performing one-dimensional current source density (CSD) analysis (e.g., Vaughan and Arezzo, 1988). Matrices and Plexiglas bars permitting painless head

fixation were embedded in a mound of dental acrylic secured to the skull with inverted bolts keyed into the bone. Peri- and postoperative anti-inflammatory agents were given to reduce potential discomfort. Recordings began 2 weeks after surgery.

Recordings were performed with multicontact electrodes constructed in our laboratory (Barna *et al.*, 1981). They contained 14 recording contacts arranged in a linear array and evenly spaced at 150- $\mu\text{m}$  intervals ( $<10\%$  error), permitting simultaneous recording across A1 laminae. Contacts were 25- $\mu\text{m}$  stainless-steel wires insulated except at the tip, which were fixed in place within the sharpened distal portion of a 30-gauge tube, and were maintained at an impedance of 0.1–0.4 M $\Omega$  at 1 kHz. The reference was an occipital epidural electrode. Headstage preamplification was followed by amplification ( $\times 5000$ ) with differential amplifiers (down 3 dB at 3 and 3 kHz). Signals were digitized at a rate of 3400 Hz and averaged by NEUROSCAN software to generate auditory evoked potentials (AEPs). Multiunit activity (MUA) was extracted by high-pass filtering the raw input at 500 Hz (roll-off 24 dB/octave), further amplifying ( $\times 8$ ) and full-wave rectifying the derived signal, and computer averaging the resultant activity.

MUA measures the envelope of action potential activity generated by neuronal aggregates, weighted by neuronal location, size, and electrode impedance (see Vaughan and Arezzo, 1988). MUA is similar to cluster activity but has greater response stability than either cluster or single-unit responses (Nelken *et al.*, 1994). We observe sharply differentiated MUA at recording contact spacings of 75  $\mu\text{m}$  (e.g., Schroeder *et al.*, 1990), and other investigators have demonstrated a similar sphere of recording (Brosch *et al.*, 1997). For some recording sessions, data were stored on either an analog (bandpass 0–5 kHz) or a digital tape recorder (sample rate 6 kHz). Due to limitations of the acquisition computer, sampling rates were less than the Nyquist frequency of the low-pass filter setting of the amplifiers. Empirical testing revealed negligible signal distortion, as almost all energy in the neural signals was less than 1 kHz. Samples of off-line data from the digital tape recorder were redigitized at 6 kHz, and resultant MUA had waveshapes and amplitudes nearly identical to those of data sampled at the lower rate (distortion  $<1\%$ ). MUA acquired from the digitally taped data was also low-pass filtered below 800 Hz using newer digital filters (96 dB/octave, RP2 modules, Tucker Davis Technologies) and then averaged at a sampling rate of 2 kHz to further test the accuracy of the initial measurements. Differences between these and initial measurements were negligible (Fishman *et al.*, 2001). To further validate MUA measures, peristimulus-time-histograms (PSTHs) were constructed from high-pass-filtered (500 Hz) data sampled at 65 kHz with a binwidth of 1 ms (BRAINWARE32, Tucker Davis, Tech.) for all electrode penetrations whose responses were stored on tape. Triggers for spike acquisition were set at 2.5 times the amplitude of the high-frequency background activity. Neuronal cluster responses were acquired from lower lamina 3, the same depth at which MUA was analyzed.

One-dimensional CSD analysis was used to characterize the laminar pattern of net current sources and sinks within

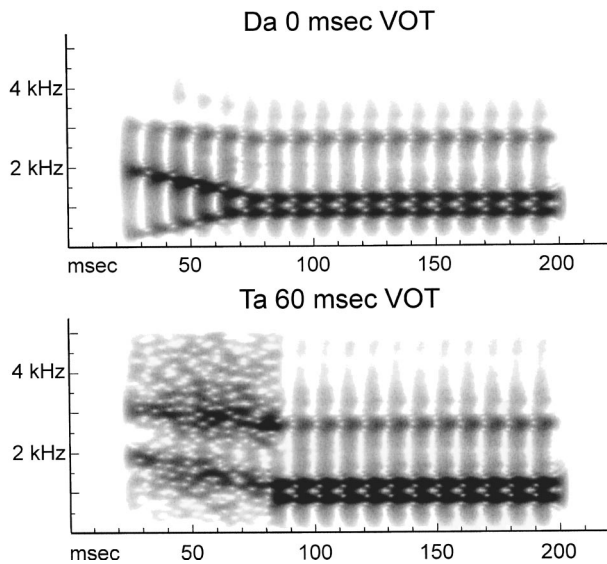


FIG. 1. Sound spectrographs of the synthetic syllables /da/ and /ta/ used in this study. Syllables with a VOT of 0 and 60 ms are shown. Syllables with a VOT of 20 and 40 ms were also presented. See the text for details.

A1. CSD was calculated from AEP laminar profiles using an algorithm that approximated the second spatial derivative of the field potentials across three adjacent depths (Freeman and Nicholson, 1975). Sinks generally index regions of net synaptic depolarization, although they can also represent current return for hyperpolarization occurring at an adjacent site. Sources denote either sites of net current return for nearby depolarizations or locations of hyperpolarization. These possibilities were distinguished by using the concurrently recorded MUA as a measure of net neuronal excitation and inhibition. A sink associated with increased MUA reflects excitatory post-synaptic potentials (EPSPs), while a source associated with MUA reduction indicates hyperpolarization.

Speech sounds were the syllables /da/ and /ta/ with VOTs of 0, 20, 40, and 60 ms. They were initially synthesized at the Haskins Laboratories (New Haven, CT), and later digitized at a sampling frequency of 44.1 kHz and edited. Syllables with a VOT of 0 and 20 ms were reliably perceived by human listeners as /da/, while those with a VOT of 40 and 60 ms were perceived as /ta/. Sound spectrographs of /da/ with a 0-ms VOT and /ta/ with a 60-ms VOT are shown in Fig. 1. Syllables were composed of three formants with steady-state center values of 817 Hz for the first formant ( $F_1$ ), 1181 Hz for the second formant ( $F_2$ ), and 2632 Hz for the third formant ( $F_3$ ). Starting formant frequencies were 200 Hz for  $F_1$ , 1835 Hz for  $F_2$ , and 3439 Hz for  $F_3$ . Formant transitions were 30 ms for  $F_1$ , and 40 ms for  $F_2$  and  $F_3$ . The fundamental frequency ( $F_0$ ) was 100 Hz. Syllables were presented at 80 dB SPL. The intensity of aspiration noise at onset was 17 dB less than the sound-pressure peaks in the vowels, and it rose nearly linearly by 5 dB over the 60-ms maximum time span of the noise. Isointensity frequency response functions based on pure-tone responses at 60 dB SPL were used to characterize the frequency sensitivity of the cortical sites and derive an estimate of the best frequency (BF). BF was defined as the frequency that elicited the largest response to tone onset. Pure tones ranged

from 0.2 to 12 kHz, with 10-ms linear rise–decay ramps. They were digitally constructed, edited, and delivered at a sampling frequency of 44.1 kHz using DIGIDESIGN TURBO-SYNTH and SOUND DESIGNER II software and hardware. All stimuli were 175 ms in duration and were presented once every 658 ms via a dynamic headphone (Sony, MDR-7502) coupled to a 60-cc plastic tube that was placed against the ear contralateral to the recording site. Sound intensity was measured with a Bruel & Kjaer sound-level meter (type 2236) positioned at the opening of the plastic tube attached to the headphone.

Recordings were performed in a sound-attenuated chamber with the animals painlessly restrained. Monkeys maintained a relaxed, but alert state, facilitated by frequent contact and delivery of juice reinforcements. Positioning of the electrode was guided by on-line inspection of AEPs and MUA evoked by 80-dB clicks. Tone bursts and speech sounds were presented when the recording contacts of the linear-array electrode straddled the inversion of early cortical AEP components. Response averages were generated from 50–100 stimulus presentations.

After completion of a recording series, animals were deeply anesthetized with sodium pentobarbital and perfused through the heart with physiological saline and 10% buffered formalin. A1 was physiologically delineated by its typically large amplitude responses and by a BF map that was organized with low BFs located anterolaterally and higher BFs posteromedially (e.g., Merzenich and Brugge, 1973; Morel *et al.*, 1993). Electrode tracks were reconstructed from coronal sections stained with cresyl violet, and A1 was anatomically identified using published criteria (e.g., Morel *et al.*, 1993). Depths of the earliest click-evoked current sinks were used to locate lamina 4 and lower lamina 3 (Steinschneider *et al.*, 1992). This method is compatible with findings of other studies (Müller-Preuss and Mitzdorf, 1984; Metherate and Cruikshank, 1999; Rose and Metherate, 2001; Cruikshank *et al.*, 2002). A later current sink in upper lamina 3 and a concurrent source located more superficially were almost always identified in the recordings. The sink served as an additional marker of laminar depth, and has been repeatedly observed (e.g., Müller-Preuss and Mitzdorf, 1984; Steinschneider *et al.*, 1992, 1994, 1998; Metherate and Cruikshank, 1999; Fishman *et al.*, 2000b; Cruikshank *et al.*, 2002). This physiological procedure for laminar identification was anatomically checked by correlation with measured widths of A1 and its laminae at select electrode penetrations. MUA recorded from infragranular laminae and subjacent white matter, and with onset and peak latencies earlier than intracortical activity, was ascribed to activity in thalamocortical (TC) afferents (Steinschneider *et al.*, 1992, 1994).

Representation of VOT in temporal patterns of population responses was evaluated in several ways. We first examined relative MUA amplitudes within lower lamina 3, segregated according to the BF of the recording sites. Response ratios were derived by dividing the amplitude of the MUA at 10 ms increments after the peak “on” response by the maximum amplitude of the initial “on” response. Mean peak latencies of the initial “on” responses ranged from 12 to 13 ms for the four syllables and were not significantly different

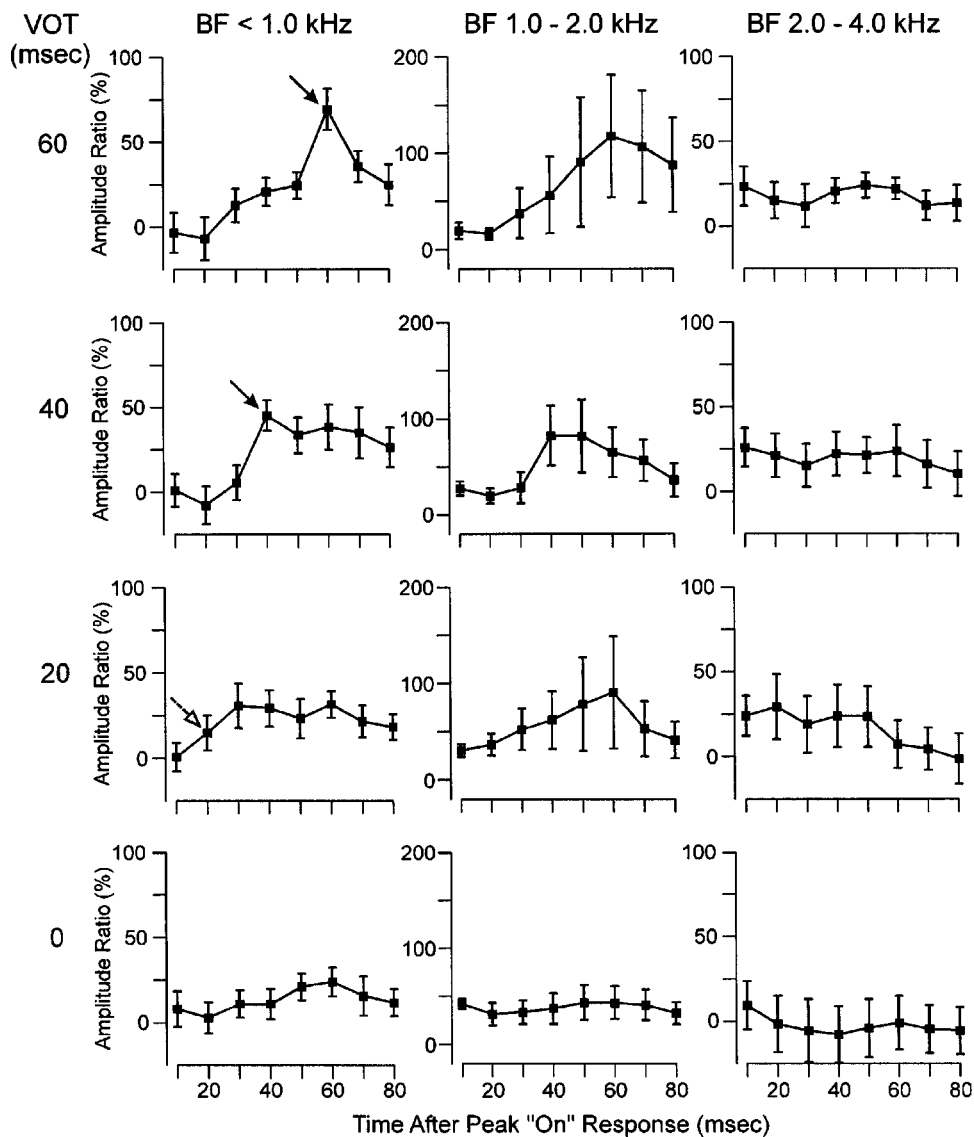


FIG. 2. Response ratios depicting amplitude of lower lamina 3 MUA at 10-ms increments relative to the onset response for each syllable. Ratios are segregated by the BF of the recording site. Syllable VOT is shown at the left. See the text for details.

from each other (ANOVA,  $p=0.32$ ). Data were subjected to a two-way ANOVA, and *post hoc* analysis was performed using a Tukey–Kramer adjustment for multiple comparisons. We also examined the absolute mean amplitude of lower lamina 3 MUA segregated according to BF. Statistical analysis was performed at key time points using one-way ANOVA and *post hoc* Newman–Keuls multiple comparisons test. Significance level for *post hoc* tests was  $p<0.05$ . Similar analyses were performed for MUA measures of TC fiber activity, PSTH indices, and amplitudes of CSD sinks and sources. Only  $p$  values of the ANOVA will be stated.

### III. RESULTS

Results are based on 78 electrode penetrations into A1 where lower lamina 3 MUA exhibited a short latency ( $<25$  ms), transient response elicited by syllable onset. MUA had a BF of  $<1$  kHz, 1–2 kHz, 2–4 kHz, and  $>4$  kHz in 33, 15, 12, and 18 penetrations, respectively. Excitation without an “on” response was evoked in five additional penetrations, sustained MUA suppression occurred in one penetration, and no responses were obtained in two penetrations.

#### A. Representation of VOT in A1

Neural representation of VOT is determined by the tonotopic organization of A1. Results are graphically depicted in Fig. 2, which illustrates response ratios obtained in low, middle, and higher BF regions for each stimulus VOT and at each 10-ms incremental time point following the peak of the initial response to stimulus onset. Response patterns differentially reflecting VOT are restricted to lower BF recording sites. At sites with a BF of  $<1$  kHz, response peaks time-locked to voicing onset are observed for syllables with 40- and 60-ms VOTs (solid arrows). This low-frequency region would be predicted to respond to voicing onset, which initiates the increase in  $F1$  sound energy. However, the stimulus with a 20-ms VOT fails to evoke a significant peak of activity time-locked to voicing onset (dotted, unfilled arrow). The key statistical result is a significant stimulus $\times$ time interaction ( $p=0.0001$ ). *Post hoc* analysis reveals significant increases in activity at voicing onset for the syllables with a VOT of 40 and 60 ms when compared against all preceding time points. This effect is absent in the response evoked by the 20-ms VOT syllable. MUA at locations with BFs between 2–4 kHz does not contain response peaks to voicing

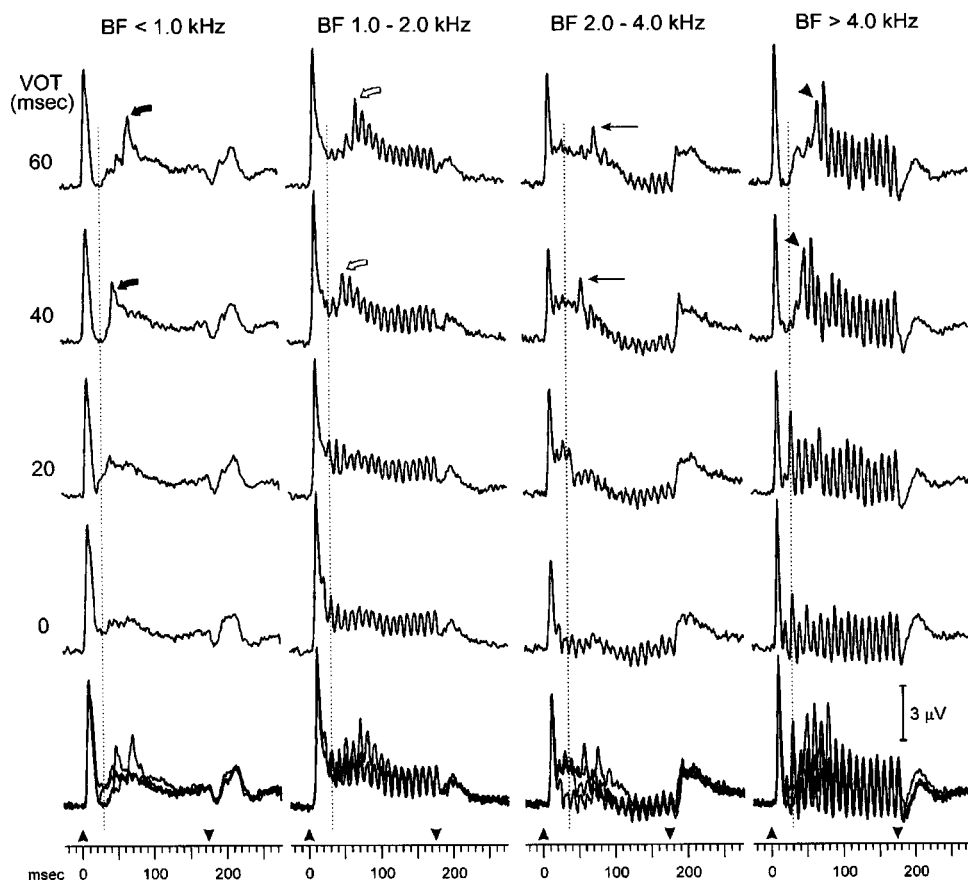


FIG. 3. Averaged absolute amplitude responses in lower lamina 3 evoked by the syllables and segregated by recording site BF. Syllable VOT is shown at the left and superimposed waveforms at the bottom. Arrowheads above the timelines indicate stimulus onset and offset. Vertical dotted lines are aligned with predicted times for responses to voicing onset for the 20-ms VOT syllable. See the text for details.

onset (right-hand column). While the BF region of 1–2 kHz has response peaks evoked by voicing onset for the longer VOT stimuli, there is a marked increase in response variability. Examination of the individual penetrations in this BF range suggests that a differential VOT effect is observed when large-amplitude responses are evoked by lower frequency tones that approximate  $F1$ .

Similar to results obtained using relative amplitude measures, differential representation of VOT is present in the absolute mean amplitude of synchronously active A1 populations (Fig. 3). Vertical dotted lines are aligned with the predicted location of the peak in the response to voicing onset for the 20-ms VOT stimulus to facilitate waveform comparisons. Overlaid responses to all four syllables are depicted at the figure bottom. Sites with BFs <1.0 kHz show peaks in MUA time-locked to both consonant release and voicing onset only in response to syllables with VOTs of 40 and 60 ms (solid curved arrows). MUA elicited by these syllables contains a brief period of suppression that follows the responses to consonant release. This suppression decays in the response to the 60-ms VOT stimulus, and is followed by a plateau of activity upon which the response to voicing onset occurs. The expected response to voicing onset for the 20-ms VOT stimulus occurs at the peak of the MUA suppression, truncating the suppression but restricting the development of a response peak time-locked to voicing onset. ANOVAs performed on response amplitudes at the times of response peaks evoked by voicing onset for the 60- and 40-ms VOT syllables, and at the time of the predicted response to voicing onset for the 20-ms VOT syllable, show

significant main effects of stimulus VOT ( $p < 0.0001$ ,  $< 0.0029$ , and  $< 0.0025$ , respectively). *Post hoc* analyses reveal that peak responses to voicing onset evoked by the 40- and 60-ms VOT syllables are larger than those evoked by the other syllables at the same time points. In contrast, at the time of the predicted response to voicing onset for the 20-ms VOT syllable, activity is similar in amplitude to that elicited by the 0-ms VOT syllable, and both responses are larger than the responses evoked by the prolonged VOT syllables. Somewhat less pronounced changes are observed in the 1.0–2.0-kHz BF region. MUA increases time-locked to voicing onset are still only observed for the prolonged VOT stimuli (unfilled, curved arrows), and ANOVAs reveal main effects of stimulus VOT at the 60-, 40-, and 20-ms VOT time points ( $p < 0.0001$ , 0.0135, and 0.0008, respectively). *Post hoc* tests reveal differences identical to those observed in the <1-kHz BF region, with the exception that the response to voicing onset for the 40-ms VOT syllable is not significantly larger at that time point than the activity evoked by the 20-ms VOT stimulus.

A different pattern of activity is seen in the BF region of 2.0–4.0 kHz. Following an initial burst of MUA time-locked to consonant release, there is a sustained plateau of activity that persists throughout the higher frequency aspiration noise. Sustained activity is absent in the response evoked by the 0-ms VOT stimulus, which has no aspiration noise. Another phasic burst of activity that is present for the two longer VOT stimuli (Fig. 3, straight arrows) occurs at the termination of the sustained increase, and is absent in the response to the 20-ms VOT stimulus (vertical, dotted line).

Differences among these latter responses, however, fail to reach statistical significance.

Regions with BFs greater than the spectral content of the syllables (>4.0 kHz) are also responsive to the sounds presented at 80 dB. Response bursts are evoked by consonant onset, and are followed by variable responses occurring during the aspiration noise, and modest increases in activity to voicing onset in the longer VOT stimuli (arrowheads). As for all BF regions greater than 1.0 kHz, periodic activity phase-locked to the 100-Hz  $F_0$  is present. These modulations are in-phase across the neuronal populations of the three higher BF regions.

Similar response patterns reflecting VOT are observed in PSTHs derived from lower lamina 3 cell cluster activity (data not shown). Responses are drawn from BF regions where MUA contained components time-locked to consonant release and voicing onset: BF was less than 1 kHz in 18 penetrations, and 1–2 kHz in 10 others. PSTHs contain response components time-locked to both consonant and voicing onset for the 40- and 60-ms VOT stimuli, while the 0- and 20-ms VOT syllables evoke a response only to consonant release followed by a plateau of sustained activity. ANOVAs reveal significant main effects of stimulus VOT at the 60-, 40-, and 20-ms VOT time points ( $p < 0.0128$ , 0.05, and 0.0006, respectively). *Post hoc* analyses indicate that responses to the 40- and 60-ms VOT syllables are larger at the time of their voicing onsets than activity evoked by the shorter VOT syllables at the same time points. The response to the 20-ms VOT syllable at the expected peak to voicing onset is not different from that evoked by the 0-ms VOT sound, and both are larger than responses to the 40- and 60-ms VOT syllables.

## B. Thalamocortical fiber (TC) representation of VOT

Cortical responses could simply reflect transmission of a preexisting pattern present in the activity of TC fibers, or alternatively, they could represent a transformation of the input activity. This issue was addressed by examining MUA from infragranular laminae and subjacent white matter recorded simultaneously with that in lower lamina 3. MUA was ascribed to activity in TC afferents based on its depth and earlier response latency. TC fiber MUA was identified in 22, 7, and 7 electrode penetrations that entered low, middle, and higher BF regions of A1, respectively. The average separation between MUA recorded from lower lamina 3 and deeper activity in TC afferents was 0.9, 0.8, and 1.2 mm in the low, middle, and high BF groups, respectively. Peak latencies of the initial “on” responses of TC fibers were 3, 2, and 2 ms earlier than those of lower lamina 3 MUA for the three BF groups, respectively.

TC fiber MUA recorded from low BF regions of A1 also exhibits accentuated response peaks evoked by voicing onset in the 40- and 60-ms VOT stimuli (Fig. 4, solid arrows). Responses evoked by the 0- and 20-ms VOT stimuli are very similar in morphology. The absent peak in the response to the 20-ms VOT stimulus is indicated by the unfilled arrow. While this finding indicates that the cortical MUA is partly based on subcortical response patterns, intracortical processes appear to amplify differences between responses

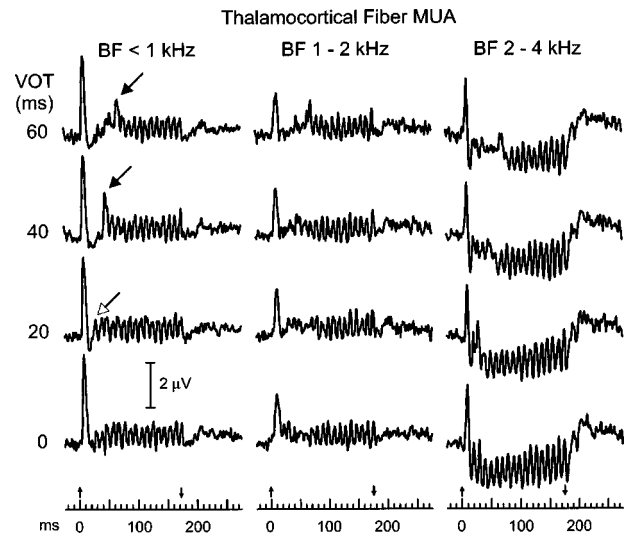


FIG. 4. TC fiber MUA recorded from infragranular laminae and white matter immediately beneath A1. Response peaks to voicing onset are evoked by the prolonged VOT stimuli but not by the syllable with the 20-ms VOT in the low-BF region (arrows). Low amplitude response peaks to voicing onset, especially to the 60-ms VOT syllables, are also observed in the TC fiber MUA recorded in higher BF regions. Additional speech features are represented in TC fiber responses. In all BF regions, the 100-Hz  $F_0$  is represented by phase-locked responses. This pattern is absent from the cortical MUA within low BF regions (Fig. 3), paralleling findings using click train and complex tone stimuli (Steinschneider *et al.*, 1998; Fishman *et al.*, 2000a).  $F_1$  duration is tracked by sustained increases in MUA within BF regions less than 1 kHz and by sustained suppression in the highest BF region, patterns also observed in the cortical MUA (Fig. 3).

evoked by short and long VOT syllables. The TC fiber response peak evoked by voicing onset in the 60-ms VOT syllable fails to reach statistical significance when compared against responses evoked by the other syllables at the same time point ( $p < 0.15$ ). While this failure could be due to the decreased sample size relative to the lower lamina 3 MUA ( $N = 22$  vs 33), the same analysis performed on lower lamina 3 MUA from these 22 penetrations demonstrates a significant response increase at voicing onset ( $p < 0.0001$ ). Similarly, while TC fiber response amplitudes at the time of voicing onset for the 40-ms VOT syllable are significantly different from each other ( $p < 0.03$ ), no *post hoc* test is significant. In contrast, *post hoc* analysis performed on the more limited cortical sample does demonstrate a larger response to voicing onset for the 40-ms VOT syllable when compared against activity evoked by the other syllables at the same time point (associated ANOVA,  $p = 0.0007$ ).

CSD profiles recorded in lower lamina 3 partly reflect the initial synaptic activity in A1 evoked by its thalamic input, and display a differential pattern of VOT representation that is also less pronounced than the MUA in lower lamina 3 (Fig. 5). Following large amplitude sinks evoked by consonant release, low amplitude sinks are generated by the syllables with a 40- and 60-ms VOT in lower BF regions (left-hand column, solid arrows). Their low amplitude may be partly based on net current summation with current sources evoked by other concurrent synaptic events. Despite their small size, activity evoked by voicing onset in the four syllables is significantly different ( $p < 0.032$  and  $p < 0.0001$  for the 60- and 40-ms VOT stimuli, respectively), and *post*

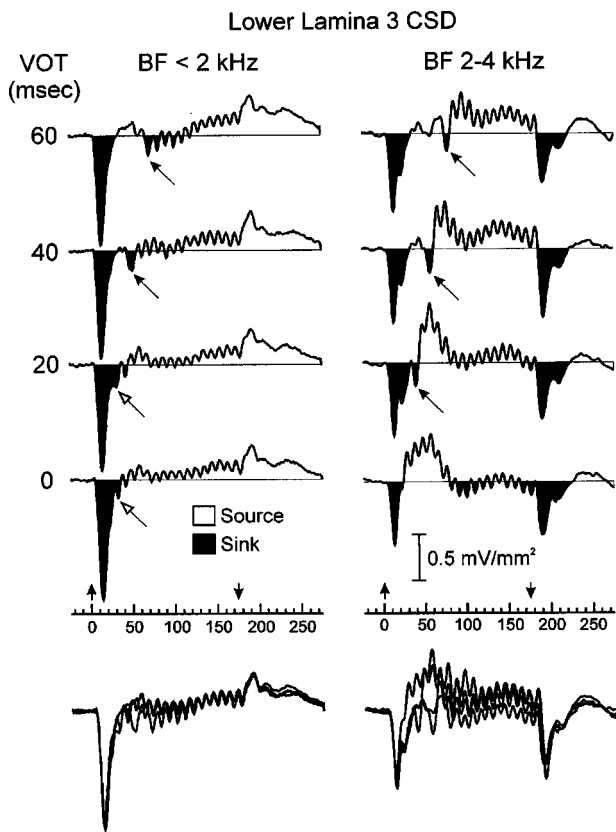


FIG. 5. Averaged CSD profiles recorded from lower lamina 3 segregated by BF of the recording sites. Superimposed waveforms are shown at the bottom. See the text for details.

*hoc* analyses indicate that the sinks evoked by the 40- and 60-ms VOT syllables are significantly larger than the activity evoked by the other two syllables at the same time points. While activity at the time point for the expected peak in the response to voicing onset for the 20-ms VOT sound is different among the syllables ( $p < 0.005$ ), the sinks evoked by the 0- and 20-ms VOT syllables are equivalent (unfilled arrows), and are greater than the responses to the longer duration VOT sounds. A different CSD pattern is observed in the higher BF region (right-hand column). Here, sinks occurring after the termination of the aspiration noise are seen for the syllables with 20-, 40-, and 60-ms VOTs (arrows), followed by sources and sinks evoked by syllable offsets.

Cortical accentuation of responses reflecting VOT is exemplified by the laminar profile of MUA simultaneously recorded across middle and lower laminae during an electrode penetration into a lower BF area (Fig. 6). Latencies of MUA peaks at depths A–F are more than 2 ms later than those recorded at depths G–I. The very short onset latency of the deeper responses ( $\sim 5$  ms) is indicative of their TC fiber origin. Tone-evoked responses are sharply tuned and are maximal to frequencies below 1 kHz at all sites except depth H, which has a broad tuning curve spanning 0.2 to 8 kHz. MUA at the three deeper sites (G–I) contains prominent bursts evoked by syllable onset followed by activity phase-locked to the  $F_0$  and variable increases time-locked to voicing onset. In contrast, MUA at the more superficial depths (A–F) contains prominent responses to both syllable and voicing onsets for the longer VOT stimuli (solid arrows) and an absence of

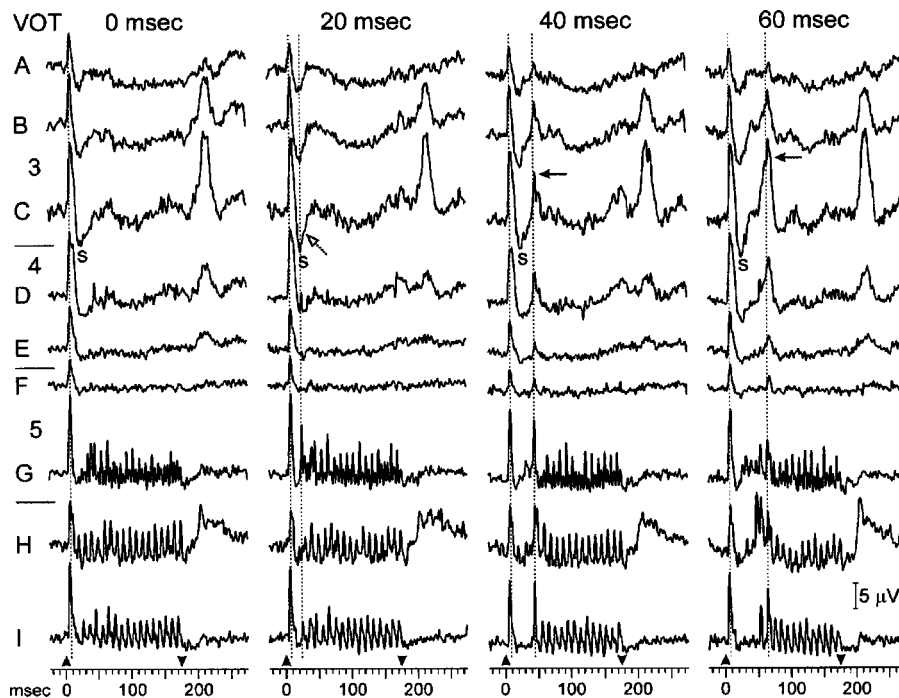


FIG. 6. Laminar profile of MUA simultaneously recorded at 150- $\mu$ m intervals during an electrode penetration into a low BF site in A1. Each recording depth is labeled at the far left, where approximate laminar boundaries are also shown. Note the difference in response patterns from the cortical MUA recorded at depths A through F, and those from TC fibers recorded at depths G through I. Dotted lines facilitate comparisons across responses at consonant and voicing onsets, and help illustrate the greater than 2-ms latency shift in the TC fiber responses and those from cortex. Following the response burst to consonant onset, there is a period of MUA suppression denoted by the symbol S. The predicted response to voicing onset for the 20-ms VOT syllable occurs during this suppression (dotted arrow), and the response is similar to that evoked by the 0-ms VOT stimulus. In contrast, both consonant and voicing onset elicit phasic responses for the two longer VOT stimuli (solid arrows).

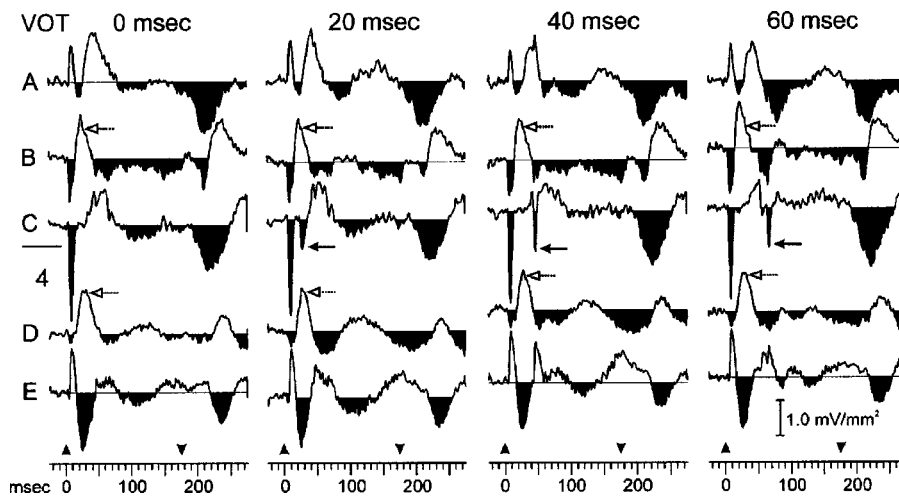


FIG. 7. Laminar CSD profile recorded simultaneously with the MUA shown in Fig. 6. Depths shown at the left correspond to the same depths in the previous figure. Note that responses to voicing onset at depth C in the CSD are present to syllables with VOTs of 20, 40, and 60 ms (solid arrows), while the cortical MUA shown in Fig. 6 has responses to voicing onset for only the two more prolonged VOT stimuli. MUA suppression in the previous figure is coincident with current sources at depths B and D (unfilled arrows).

phase-locked activity to the  $F_0$ . The location of the predicted response to voicing onset for the 20-ms VOT syllable is shown by the unfilled dotted arrow. At the time when the response to voicing onset should occur, MUA suppression below baseline levels (denoted by S) is present throughout depths A–F. This suggests that MUA patterns differentially reflecting VOT are accentuated by transient neuronal suppression.

Simultaneously recorded CSD supports this conclusion (Fig. 7). Sinks located at depths B–D mark the initial cortical depolarizations. At the center of the initial depolarization (depth C), additional sinks evoked by voicing onset are observed for the syllables with VOTs of 20, 40, and 60 ms (solid arrows). This pattern is less restrictive than the simultaneously recorded MUA (Fig. 6), which fails to reveal a cortical response to voicing onset for the 20-ms VOT syllable. However, the CSD does resemble the pattern of TC fiber MUA recorded at depth G, suggesting that the CSD represents activity derived from TC fiber input. This change in temporal response patterns between TC fiber MUA and subsequent synaptic activity, and the intracortical MUA, supports the conclusion that intracortical processes sharpen responses reflecting VOT. Sources concurrent with MUA suppression (Fig. 7, unfilled dotted arrows) likely reflect currents induced by inhibitory events.

### C. Interlaminar processing of VOT

Early cortical response patterns reflecting syllable VOT are maintained in subsequent interlaminar processing within regions with  $BFs < 2$  kHz (Fig. 8). The left-hand column depicts the averaged CSD recorded in upper lamina 3, while the right-hand column illustrates CSD recorded from even more superficial laminae. The averaged CSD recorded from these upper lamina 3 sites consists of an initial current sink evoked by consonant release with onset and peak latencies 5–6 ms later than those in lower lamina 3. Sinks evoked by voicing onset (arrows) show a marked amplitude reduction

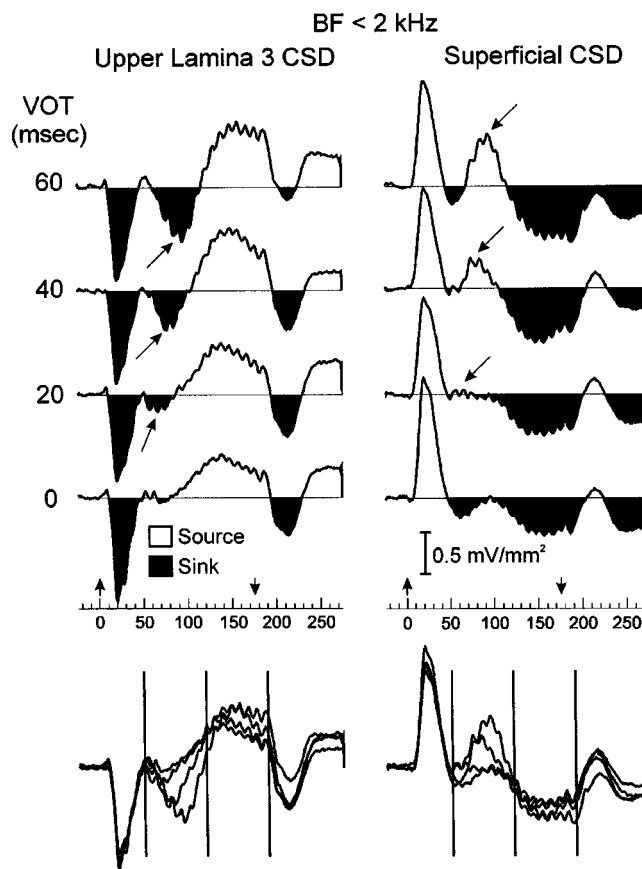


FIG. 8. Averaged upper lamina 3 and more superficial CSD recorded from penetrations into A1 areas with  $BFs < 2$  kHz. Average separation between the earlier lower lamina 3 sink and the more superficial and later sink in upper lamina 3 is  $380 \mu\text{m}$ , while the even more superficial activity was recorded an average of  $333 \mu\text{m}$  above the upper lamina 3 CSD. Superficial sources are coincident with the upper lamina 3 sinks, consistent with a current dipole within apical dendrites of pyramidal cells. Response evoked by voicing onset to the longer VOT syllables is greater than that elicited by the 20-ms VOT syllable (arrows). Effects are highlighted by the superimposed waveforms shown at the bottom of the figure, which also illustrates boundary demarcations used to quantify the data.



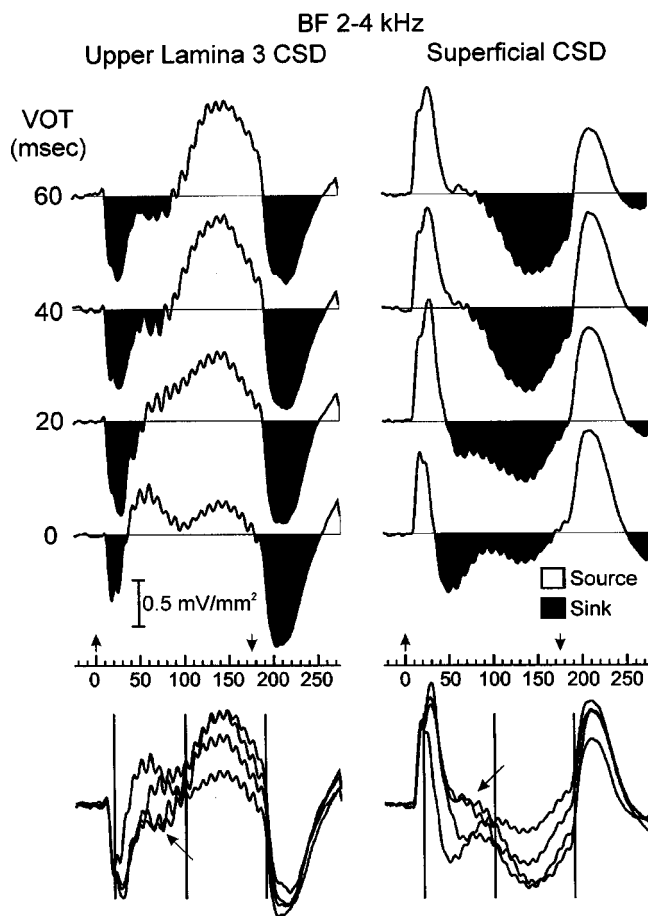


FIG. 9. Averaged upper lamina 3 and more superficial CSD recorded from penetrations into A1 areas with BFs of 2–4 kHz. Separation between the lower lamina 3 and upper lamina 3 recording sites averages  $355\ \mu\text{m}$ , while separation between the latter depth and the superficial CSD is  $368\ \mu\text{m}$ . Arrows overlying the superimposed waveforms at the bottom of the figure highlight the extended sources and sinks evoked by the 40- and 60-ms VOT syllables. Vertical lines denote boundaries for the areas used to quantify the data.

when elicited by the 20-ms VOT syllable. More superficial CSD patterns mirror these findings (arrows). This dipole configuration is consistent with synaptic activation of pyramidal cell apical dendrites in lamina 3 with passive current return occurring in more superficial laminae. Effects are quantified by examining the area of sources and sinks evoked by the various stimuli. Two portions of the waveforms are examined. The first encompasses the time interval from 50 to 120 ms, and the second incorporates the interval from 120 to 190 ms. Interval boundaries correspond to points of common divergence in upper lamina 3 between the ends of the initial sink, the period ending the second sink evoked by voicing onset and start of the following source, and the CSD response elicited by stimulus offset (Fig. 8, drop lines in superimposed waveforms). In the 50–120-ms time interval, responses evoked by the syllables are significantly different from one another in both upper lamina 3 and the more superficial CSD (both  $p < 0.0001$ ). In both cases, the syllable with the 60-ms VOT produces a larger response than the syllables with a VOT of 0 and 20 ms, while the response to the 40-ms VOT syllable is greater than that to the 0-ms VOT sound. Additionally, the source in the superficial CSD

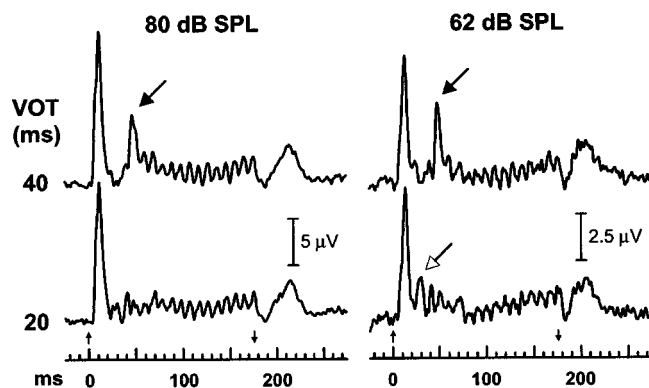


FIG. 10. Averaged lower lamina 3 MUA evoked by syllables presented at the standard 80-dB SPL intensity and at 62 dB SPL. Activity represents the mean of the responses recorded from six A1 sites, three located in areas with BFs  $< 1$  kHz and three with BFs between 1 and 2 kHz. Solid arrows mark the accentuated responses evoked by voicing onset of the 40-ms VOT syllable. There is an increase in the ratio of responses evoked by voicing onset relative to the activity evoked by consonant release for the less intense syllables, including that elicited by the 20-ms VOT stimulus (unfilled arrow).

evoked by the 40-ms VOT sound is larger than that evoked by the 20-ms VOT syllable. These effects are best illustrated by the superimposed waveforms shown at the bottom of the figure.

Aspiration noise modulates interlaminar processing in the 2–4-kHz BF region (Fig. 9). The prolonged VOT syllables elicit extended sinks in upper lamina 3 and sources in the more superficial CSD (arrows overlying superimposed waveforms at figure bottom). Sources and sinks evoked by the 20-ms VOT sound are intermediate in amplitude between those elicited by the /ta/ stimuli and those generated by the 0-ms VOT syllable. Areas of sources and sinks were computed from the intervals of 20 to 100 ms, and 100 to 190 ms, boundaries representing common points of diversion in the waveforms. The magnitude of the sink from 20 to 100 ms in upper lamina 3 is the only difference reaching statistical significance ( $p = 0.02$ ), with *post hoc* analysis indicating that the responses to the longer VOT syllables are larger than the activity evoked by the 0-ms VOT syllable.

#### D. Effects of stimulus intensity

Comparisons between syllable-evoked activity elicited by stimuli presented at 80 and 62 dB SPL were made at a limited number of A1 sites. Three sites had BFs  $< 1$  kHz, while 3 had BFs between 1 and 2 kHz. The averaged MUA responses elicited by the syllables with a 20- and 40-ms VOT are shown in Fig. 10. Accentuated responses evoked by voicing onset of the more prolonged 40-ms VOT stimulus persist despite attenuation of syllable intensity (filled arrows) and a twofold decrease in overall response amplitudes. There is for the softer stimuli, however, an increase in the ratio of responses evoked by voicing onset relative to that elicited by consonant onset that occurs for both the 40-ms and the 20-ms VOT stimuli (unfilled arrow).

## IV. DISCUSSION

### A. Summary of findings and relationship to psychoacoustic studies

The present study demonstrates that features of speech important for discriminating voiced (e.g., /d/) from unvoiced (e.g., /t/) stop consonants are represented by synchronized responses in A1 neuronal populations. VOT is the principal determinant of this phonetic perception (Lisker and Abramson, 1964; Faulkner and Rosen, 1999). Syllables which are generally perceived as /ta/ evoke statistically reliable responses to both consonant release and voicing onset in A1, whereas syllables which are usually perceived as /da/ elicit statistically significant activity only to consonant release. Syllables generating a “double on” response pattern could be rapidly placed into the category of unvoiced stop consonants, whereas those eliciting a “single on” response could be identified as a voiced consonant. Discrimination between syllables that straddle these response patterns would require only that the brain differentiate between “single on” and “double on” activity profiles, whereas more subtle timing discriminations would be required for differentiating syllables residing on the same side of the perceptual boundary (Pisoni and Lazarus, 1974; Carney *et al.*, 1977; Pisoni *et al.*, 1982; Kewley-Port *et al.*, 1988). This scheme is consistent with the idea that phonetic encoding is partly based on a “warped” representation of acoustical properties, such that some acoustic differences located along a specific portion of a physical continuum are readily distinguished from one another, whereas similar magnitude differences located along a different portion of that continuum are difficult to discriminate (Stevens, 1981).

Three main categories of VOT relationships occur in speech (Lisker and Abramson, 1964). Voicing either begins before (lead), after (lag), or near the time of consonant release, and there is a nonoverlapping trimodal distribution of VOTs for almost all the world’s languages. Voicing lead falls in the range of  $-125$  to  $-75$  ms, voicing near the time of consonant release at  $0$  to  $+25$  ms, and voicing following release at  $+60$  to  $+100$  ms. Averages for these three categories are  $-100$ ,  $+10$ , and  $+70$  ms, respectively. An A1 physiological boundary of between  $+20$  and  $+40$  ms falls well within the range of the latter two VOT categories. This temporal processing scheme is not negated by changes in speaking rate modulating VOT (Summerfield, 1981; Utman, 1998; Allen and Miller, 1999; Boucher, 2002). While faster rates shorten VOTs of unvoiced, aspirated stops, values remain greater than 40 ms. Temporal patterns alone, however, cannot distinguish between voicing lead and voicing lag categories, as both would yield “double on” responses. Representation of spectral cues such as the presence of low-frequency voicing energy prior to consonant release, or high-frequency aspiration noise after consonant release, could facilitate this discrimination.

The  $+20$ - to  $+40$ -ms VOT interval likely represents a natural psychoacoustical boundary in mammalian hearing. Developmentally, young infants with little language exposure can discriminate a VOT contrast of  $+20$  with  $+40$  ms, even when this contrast is not phonetically relevant for the

native language of the child (Eimas *et al.*, 1971; Lasky *et al.*, 1975; Eilers *et al.*, 1979; Jusczyk *et al.*, 1989). An interval of  $20$ – $40$  ms is generally required to perceive the temporal order of two acoustic events (Hirsh, 1959; Kewley-Port *et al.*, 1988; Jusczyk *et al.*, 1989), supporting the notion that VOT perception is partly based on the ability to determine whether consonant release and voicing onset are sequential in time (Pisoni, 1977). Many animal species respond to VOT discriminations in a categorical manner with perceptual boundaries of  $+20$  to  $+40$  ms (e.g., Kuhl and Miller, 1975, 1978; Kuhl, 1981; Kuhl and Padden, 1982; Sinnott and Adams, 1987; Dooling *et al.*, 1989). “Single on” vs “double on” temporal response patterns could serve as neurophysiological cues for all these perceptual phenomena.

Temporal response patterns do not directly account for VOT boundary shifts that occur with changes in consonant place of articulation. Boundaries are shortest for the distinction between /b/ and /p/ ( $\sim 20$  ms), longest for /g/ and /k/ ( $\sim 40$  ms), and intermediate for /d/ and /t/ ( $\sim 30$  ms) (Lisker and Abramson, 1964). Boundary shifts similar to those in humans are observed in animals, suggesting that the phenomenon is not based on language-specific processes (Kuhl and Miller, 1978).  $F1$  frequency at voicing onset is the key acoustic feature responsible for this boundary shift (Lisker, 1975; Summerfield and Haggard, 1977; Soli, 1983). It is highest for labial stops, lowest for velar stops, and intermediate for alveolar stops. As  $F1$  frequency is lowered, a longer VOT is required for perceiving an unvoiced stop consonant. Greater frequency differences between two consecutive tones increase the likelihood that they will be perceived as occurring simultaneously (Parker, 1988). Japanese quail are more likely to label stop consonants as unvoiced when they incorporate a high  $F1$  frequency (Kluender, 1991; Kluender and Lotto, 1994). These studies implicate auditory system mechanisms and stress the importance of  $F1$  spectral cues interacting with VOT in modulating discrimination of voiced from unvoiced consonants. Preliminary data in both humans and monkeys suggest that the capacity of A1 to generate one vs two onset responses is determined by the frequency disparity between  $F1$  and higher frequency components (Stein-schneider *et al.*, 2000), supporting the capability of these temporal response patterns to account for the perceptual VOT boundary shifts that occur with changes in consonant place of articulation.

Aspiration noise is an additional cue important for discrimination of voiced from unvoiced stop consonants that is represented in patterns of A1 activity (Repp, 1979; Sinnott and Adams, 1987; Kluender *et al.*, 1995; Lotto and Kluender, 2002). The likelihood of perceiving an unvoiced stop parallels increases in aspiration noise intensity. This acoustic cue is represented by sustained activity extending throughout its duration in higher BF regions.

Syllable intensity is an important variable that was not thoroughly examined in the present study. Perception of unvoiced stop consonants diminishes with decreases in syllable intensity (Kluender *et al.*, 1995; Lotto *et al.*, 2002). This effect cannot be satisfactorily predicted on the basis of whether low BF areas in A1 produce a “single” or “double on” neural response. Our limited data obtained at lower stimulus

intensities suggest that as intensity is lowered there is an increased relative response to voicing onset. This would suggest more unvoiced stop perceptions at softer intensities (i.e., a greater number of “double on” neural responses), and runs contrary to the perceptual data. A different interpretation offered by a reviewer suggests that as response amplitude evoked by voicing onset increases relative to that evoked by consonant onset, a bias towards a percept of a voiced stop consonant would ensue. This mechanism could be especially important at stimulus intensities where the aspiration noise is near or below threshold. In either case, a key component of the intensity effect is the relative loudness of the aspiration noise (Repp, 1979; Kluender *et al.*, 1995; Lotto *et al.*, 2002). One can predict that as intensity decreases, responses to this lower amplitude component in higher BF regions of A1 would also diminish, thereby decreasing aspiration noise as a potential cue for unvoiced stops and shifting VOT perception towards voiced stops. This physiological result could in turn counterbalance any opposite effect on perception based on activity within lower BF regions. Given that VOT and aspiration noise are represented in different tonotopic areas of A1, it follows that integration of these cues does not occur at this level of the auditory system. Secondary areas of auditory cortex, however, do integrate activity from multiple regions of A1 (Rauschecker *et al.*, 1995), implicating these downstream cortical fields as possible sites that will ultimately determine VOT perception.

## B. Relationship to other auditory physiological studies

Many response features similar to those seen in A1 are also observed in auditory-nerve fibers and cells of the auditory brainstem, indicating that the auditory periphery initiates a physiological representation of VOT that is maintained through the auditory pathways. In low BF auditory-nerve fibers, syllables evoke a transient response to consonant release, a period of suppression, and a response to voicing onset for syllables with VOTs greater than 20 ms (Sinex and McDonald, 1988). Response increases evoked by voicing onset are abrupt for syllables with VOTs of 30–40 ms, and more gradual for longer VOT values. Activity in low BF regions of A1 parallel these peripheral patterns (see Figs. 2 and 3). Within the inferior colliculus, responses evoked by consonant release and voicing onset usually merge into a single burst at a VOT less than 20 ms (Chen *et al.*, 1996), though response patterns are also modulated by interaural time differences for binaurally presented syllables (Chen and Sinex, 1999).

We indirectly examine activity in medial geniculate by investigating TC fiber response patterns, and find population responses that differentially reflect VOT. These response patterns, while consistent with continued transmission of VOT-related information from inferior colliculus, are not as robust as those seen in A1. Findings parallel our earlier observations on the transformation of activity between TC fibers and neuronal populations in A1 (Steinschneider *et al.*, 1994), and indicate that one effect of A1 processing is to accentuate responses to lower rate acoustical transients. This accentua-

tion has been previously observed (e.g., Creutzfeldt *et al.*, 1980; Miller *et al.*, 2001), and may facilitate the representation of VOT.

Present findings are compatible with those of other A1 investigations. Spatio-temporal profiles of activity in cat A1 evoked by the onsets of /be/ and /pe/ are similar, whereas a second burst of activity is evoked by the 60-ms VOT of /pe/ in low BF regions (Schreiner, 1998). A boundary of 15–20 ms is seen in the ability to generate a “double on” response to /pa/ in epidural recordings overlying guinea pig auditory cortex (McGee *et al.*, 1996). Reliable “double on” profiles are observed for syllables with VOTs of 30 ms or longer at stimulus levels of 45–65 dB SPL (Eggermont, 1995a). For syllables presented at 45–75 dB, the ratio of the response evoked by voicing relative to that evoked by consonant onset markedly increases when the VOT is greater than 25 ms (Eggermont, 1995b). These temporal response patterns are not restricted to syllables varying in their VOT, and similar activity occurs for gap-in-noise stimuli (Eggermont, 1995b), repetitive frequency-modulated tones (Lu *et al.*, 2001), and two-tone complexes varying in their relative onset times (Sugimoto *et al.*, 2002).

The spatially distributed activity representing VOT and aspiration noise supports the emerging concept that complex sounds are encoded by temporally precise responses within a network of neuronal populations distributed across tonotopically organized A1 (e.g., Creutzfeldt *et al.*, 1980; Wang *et al.*, 1995; Gehr *et al.*, 2000; Nagarajan *et al.*, 2002; see also Petersen *et al.*, 2002). These distributed responses are time-locked to transient elements embedded within the sounds, and synchronized with each other (Creutzfeldt *et al.*, 1980; Wang *et al.*, 1995). This property may help bind together activity engaged in the processing of speech, which has multiple spectral components activating widespread areas of A1. A temporally precise population coding method is relatively resistant to the addition of background noise (Wang *et al.*, 1995), a prerequisite for speech perception in noisy environments.

Integration of the distributed activity in A1 likely occurs in secondary auditory cortex. Primate A1 projects to multiple surrounding areas, including those located on the lateral surface of the superior temporal gyrus (see Morel *et al.*, 1993; Kaas *et al.*, 1999). MUA and PSTHs recorded from lower lamina 3 are especially relevant to this process, as pyramidal neurons from this depth project to these secondary fields (Galaburda and Pandya, 1983; Jones *et al.*, 1995). Lateral belt areas preferentially respond to vocalizations and noise bursts (Rauschecker *et al.*, 1995; Tian *et al.*, 2001). This characteristic implies that their activation requires integrated activity emanating from multiple frequency-specific areas of A1. Thus, neuronal activity in these areas could unite the different attributes of speech sounds represented in A1. The synchronized responses within A1 would be well-suited for driving these integration cells of nonprimary cortex (Eggermont, 1994, 2000a; deCharms and Merzenich, 1996), leading to a more holistic phonetic representation that incorporates many of the acoustic variables modifying the differential perception of voiced and unvoiced stop consonants.

The ability to ascribe MUA and CSD sinks and sources

to specific neuronal populations and events is central to our interpretations. Response latencies obtained in the present study and attributed to responses of TC fibers and A1 neurons are similar to those previously reported (e.g., Allon *et al.*, 1981; Phillips and Hall, 1990; Heil, 1997). We found peak response latencies of 12–13 ms following stimulus onset in A1 populations within lower lamina 3, and responses in TC fibers located 0.8–1.2 mm below the cortical recordings that peaked 2–3 ms earlier. A conduction velocity of about 0.4 m/s is calculated if responses were only derived from distal segments of TC afferents. This is much slower than calculated velocities of distal TC afferent inputs into A1 (Metherate and Cruikshank, 1999), supporting the conclusion that lower lamina 3 MUA predominantly reflects activity in postsynaptic cortical cells, while deeper MUA is mainly a reflection of distal, TC fiber input. This conclusion is further supported by cross correlations that reveal an average lag time of 1–5 ms between A1 activity and that in the medial geniculate (Creutzfeldt *et al.*, 1980; Miller *et al.*, 2001; see also Usrey, 2002).

The early cortical sink in lower lamina 3 and the later, more superficial sink that we have analyzed are characteristic findings in A1, and likely reflect monosynaptic EPSPs from TC fiber input induced by glutamate and mediated by AMPA/KA receptors, and later polysynaptic EPSPs indicative of intracortical processing, respectively (Metherate and Ashe, 1994; Klink *et al.*, 1999; Metherate and Cruikshank, 1999; Rose and Metherate, 2001; Cruikshank *et al.*, 2002). TC fibers emanating from the main lemniscal (ventral) division of the medial geniculate nucleus ascend through lower cortical laminae and primarily terminate within lamina 4 and lower lamina 3 of A1 (Jones and Burton, 1976; Hashikawa *et al.*, 1995). It is the synaptic activity induced by this input that largely generates the coextensive, early cortical sink (Steinschneider *et al.*, 1992). In contrast, the later, more superficial sink is likely the net result of depolarizations produced by multiple inputs, including those from lamina 4 and contralateral A1 (Mitani *et al.*, 1985; Wallace *et al.*, 1991; Pandya and Rosene, 1993). Furthermore, synaptic events in this laminar region are modulated by cholinergic activity (Metherate and Ashe, 1995; Bandrowski *et al.*, 2001). Thus, the persistence of differential responses reflecting syllable VOT and aspiration noise in the superficial sink offers an opportunity for speech-related activity to interact with learning-related processes. The early current source associated with MUA suppression is also a feature of A1 physiology (Fig. 7), and has been attributed to GABA<sub>A</sub> receptor-mediated inhibitory post-synaptic potentials from inhibitory interneurons (Metherate and Ashe, 1994; Metherate and Cruikshank, 1999; Cruikshank *et al.*, 2002) and Ca<sup>++</sup>-gated K<sup>+</sup> channel-mediated afterhyperpolarization (Eggermont, 2000b). This functional refractory period appears to amplify differential processing of VOT by suppressing responses to voicing for short VOT stimuli, suggesting that multiple factors including peripheral and subcortical processing, A1 inhibitory circuitry, and intrinsic membrane properties are responsible for shaping the temporal response patterns.

Ultimately, the utility of studying speech processing in animals must be assessed by how well the responses model

those observed in humans. Multiple studies in humans using noninvasive recording techniques have observed temporal response patterns evoked by syllables and their nonspeech analogs similar to those in monkey A1 (e.g., Kaukoranta *et al.*, 1987; Mäkelä *et al.*, 1988; Kuriki *et al.*, 1995; Sharma and Dorman, 1999). One study of note has reported that the physiological boundary for the N1 component of the scalp-recorded AEP did not reliably predict the voiced/voiceless distinction between stop consonants varying in both VOT and place of articulation (Sharma *et al.*, 2000). While there was concordance between the perceptual and physiological boundaries for the /ba/–/pa/ contrast, the perceptual boundary for /ga/–/ka/ increased to greater than 40 ms while the physiological boundary remained between 20 and 40 ms. This report, however, needs to be interpreted with caution. N1 is a later, composite wave that reflects activation of multiple auditory cortical fields, each with its own capacity to follow temporal features of complex sounds (Steinschneider *et al.*, 1999; Fishman *et al.*, 2001). It is therefore difficult to extrapolate the relevance of short latency, A1 temporal response patterns for phonetic processing from N1 properties. Finally, the two studies examining speech-evoked activity recorded directly from Heschl's gyrus identified response patterns similar to those seen in the monkey, with short VOT syllables eliciting a “single on” response and those with a prolonged VOT eliciting “double on” responses (Liégeois-Chauvel *et al.*, 1999; Steinschneider *et al.*, 1999). The latter study, using the same synthetic syllables as those in the present study, showed differential responses with a boundary of between 20- and 40-ms VOT that paralleled subject perception. Clearly, additional research must address this problem in greater detail. At the very least, the primate data serve as a valuable model for human physiological activity, whose specific relevance to phonetic perception remains to be determined by future investigations.

## V. SUMMARY AND CONCLUSIONS

Representation of stop consonant–vowel syllables with variable VOT is examined in A1 of awake monkeys to test the hypothesis that temporal response patterns differentially reflecting onsets of consonant release and voicing are maintained in large-scale population activity. In low BF areas, syllables with VOTs of 0 and 20 ms evoke a peak of activity time-locked to consonant release, while those with VOTs of 40 and 60 ms elicit an additional peak time-locked to voicing onset. Aspiration noise is represented by sustained increases in neural activity within higher BF areas. While similar patterns are seen in responses emanating from thalamocortical afferents, there is an accentuation of differential activity induced by intracortical mechanisms. Neural suppression following responses to consonant release may serve as a functional refractory period that diminishes responses to voicing onset for syllables with a short VOT. Physiological responses are compatible with the perceptual boundary of +20 to +40 ms that occurs in many languages. These findings support the hypothesis that the 20- to 40-ms interval represents a natural psychoacoustical boundary utilized for VOT perception, based in turn on determining whether the onsets of consonant release and voicing onset occur sequentially in time. Syn-

chronized responses offer an economical way in which stop consonants can be rapidly represented by auditory cortex using the well-described propensity of A1 for responding to low-frequency acoustic transients. These signature patterns are poised to provide potent inputs to higher order auditory fields involved in further phonetic encoding. Secondary fields likely integrate the activity patterns occurring in disparate tonotopic regions of A1, thereby uniting responses representing the various acoustic cues used for the differential perception of voiced from unvoiced stop consonants.

## ACKNOWLEDGMENTS

This research was supported by Grant Nos. DC00657, HD01799, and NS07098. We thank Dr. Charles Schroeder, Dr. David Reser, and Dr. Haftan Eckholdt, and Ms. Shirley Seto and Ms. May Huang for providing excellent intellectual, technical, statistical, and histological assistance.

- Ahissar, E., Abeles, M., Ahissar, M., Haidarliu, S., and Vaadia, E. (1998). "Hebbian-like functional plasticity in the auditory cortex of the behaving monkey," *Neuropharmacology* **37**, 633–655.
- Aitkin, L. M., and Dunlap, C. W. (1968). "Interplay of excitation and inhibition in the cat medial geniculate body," *J. Neurophysiol.* **31**, 44–61.
- Allen, J. S., and Miller, J. L. (1999). "Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words," *J. Acoust. Soc. Am.* **106**, 2031–2039.
- Allon, N., Yeshurun, Y., and Wollberg, Z. (1981). "Responses of single cells in the medial geniculate body of awake squirrel monkeys," *Exp. Brain Res.* **41**, 222–232.
- Bandrowski, A. E., Moore, S. L., and Ashe, J. H. (2001). "Cholinergic synaptic potentials in the supragranular layers of auditory cortex," *Synapse* **41**, 118–130.
- Barna, J., Arezzo, J. C., and Vaughan, Jr., H. G. (1981). "A new multicontact array for the simultaneous recording of field potentials and unit activity," *Electroencephalogr. Clin. Neurophysiol.* **52**, 494–496.
- Blake, D. T., Strata, F., Churchland, A. K., and Merzenich, M. M. (2002). "Neural correlates of instrumental learning in primary auditory cortex," *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10114–10119.
- Boucher, V. J. (2002). "Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates," *Percept. Psychophys.* **64**, 121–130.
- Brosch, M., Bauer, R., and Eckhorn, R. (1997). "Stimulus-dependent modulations of correlated high-frequency oscillations in cat visual cortex," *Cereb. Cortex* **7**, 70–76.
- Carney, A. E., Widin, G. P., and Viemeister, N. F. (1977). "Noncategorical perception of stop consonants differing in VOT," *J. Acoust. Soc. Am.* **62**, 961–970.
- Chen, G.-D., Nuding, S. C., Narayan, S. S., and Sinex, D. G. (1996). "Responses of single neurons in the chinchilla inferior colliculus to consonant-vowel syllables differing in voice onset time," *Aud. Neurosci.* **3**, 179–198.
- Chen, G.-D., and Sinex, D. G. (1999). "Effects of interaural time differences on the responses of chinchilla inferior colliculus neurons to consonant-vowel syllables," *Hear. Res.* **138**, 29–44.
- Creutzfeldt, O., Hellweg, F.-C., and Schreiner, C. (1980). "Thalamocortical transformation of responses to complex auditory stimuli," *Exp. Brain Res.* **39**, 87–104.
- Cruikshank, S. J., Rose, H. J., and Metherate, R. (2002). "Auditory thalamocortical synaptic transmission *in vitro*," *J. Neurophysiol.* **87**, 361–384.
- deCharms, R., and Merzenich, M. M. (1996). "Primary cortical representation of sounds by the coordination of action-potential timing," *Nature (London)* **381**, 610–613.
- Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., and Pierce, A. (1997). "Perception of synthetic /ba/-/wa/ speech continuum by budgerigars (*Melopsittacus undulatus*)," *J. Acoust. Soc. Am.* **102**, 1891–1897.
- Dooling, R. J., Okanoya, K., and Brown, S. D. (1989). "Speech perception by budgerigars (*Melopsittacus undulatus*): The voiced-voiceless distinction," *Percept. Psychophys.* **46**, 65–71.
- Eggermont, J. J. (1994). "Neural interaction in cat primary auditory cortex. II. Effects of sound stimulation," *J. Neurophysiol.* **71**, 246–270.
- Eggermont, J. J. (1995a). "Representation of a voice onset time continuum in primary auditory cortex of the cat," *J. Acoust. Soc. Am.* **98**, 911–920.
- Eggermont, J. J. (1995b). "Neural correlates of gap detection and auditory fusion in cat auditory cortex," *NeuroReport* **6**, 1645–1648.
- Eggermont, J. J. (2000a). "Sound-induced synchronization of neural activity between and within three auditory cortical areas," *J. Neurophysiol.* **83**, 2708–2722.
- Eggermont, J. J. (2000b). "Neural responses in primary auditory cortex mimic psychophysical, across-frequency-channel, gap-detection thresholds," *J. Neurophysiol.* **84**, 1453–1463.
- Eilers, R. E., Gavin, W., and Wilson, W. R. (1979). "Linguistic experience and phonetic perception in infancy: A crosslinguistic study," *Child Dev.* **50**, 14–18.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). "Speech perception in infants," *Science* **171**, 303–306.
- Faulkner, A., and Rosen, S. (1999). "Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception," *J. Acoust. Soc. Am.* **106**, 2063–2073.
- Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (2000a). "Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness," *J. Acoust. Soc. Am.* **108**, 235–246.
- Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (2000b). "Complex tone processing in primary auditory cortex of the awake monkey. II. Pitch versus critical band representation," *J. Acoust. Soc. Am.* **108**, 247–262.
- Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., Howard, M. A., and Steinschneider, M. (2001). "Consonance and dissonance of musical chords: Neural correlates in auditory cortex of monkeys and humans," *J. Neurophysiol.* **86**, 2761–2788.
- Freeman, J. A., and Nicholson, C. (1975). "Experimental optimization of current source density techniques for anuran cerebellum," *J. Neurophysiol.* **38**, 369–382.
- Gaese, B. H., and Ostwald, J. (2001). "Anesthesia changes frequency tuning of neurons in the rat primary auditory cortex," *J. Neurophysiol.* **86**, 1062–1066.
- Galaburda, A. M., and Pandya, D. N. (1983). "The intrinsic architectonic and connective organization of the superior temporal region of the rhesus monkey," *J. Comp. Neurol.* **221**, 169–184.
- Gehr, D. D., Komiya, H., and Eggermont, J. J. (2000). "Neuronal responses in cat primary auditory cortex to natural and altered species-specific call," *Hear. Res.* **150**, 27–42.
- Hashikawa, T., Molinari, M., Rausell, E., and Jones, E. G. (1995). "Patchy and laminar termination of medial geniculate axons in monkey auditory cortex," *J. Comp. Neurol.* **362**, 195–208.
- Heil, P. (1997). "Auditory cortical onset responses revisited. I. First-spike timing," *J. Neurophysiol.* **77**, 2616–2641.
- Hirsh, I. J. (1959). "Auditory perception of temporal order," *J. Acoust. Soc. Am.* **31**, 759–767.
- Holt, L. L., Lotto, A. J., and Kluender, K. R. (2001). "Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?" *J. Acoust. Soc. Am.* **109**, 764–774.
- Jones, E. G., and Burton, H. (1976). "Areal differences in the laminar distribution of thalamic afferents in cortical fields of the insular, parietal and temporal regions of primates," *J. Comp. Neurol.* **168**, 197–248.
- Jones, E. G., Dell'Anna, M. E., Molinari, M., Rausell, E., and Hashikawa, T. (1995). "Subdivisions of macaque monkey auditory cortex revealed by calcium-binding protein immunoreactivity," *J. Comp. Neurol.* **362**, 153–170.
- Jusczyk, P. W., Rosner, B. S., Reed, M. A., and Kennedy, L. J. (1989). "Could temporal order differences underlie 2-month-olds' discrimination of English voicing contrasts?" *J. Acoust. Soc. Am.* **85**, 1741–1749.
- Kaas, J. H., Hackett, T. A., and Tramo, M. J. (1999). "Auditory processing in primate cerebral cortex," *Curr. Opin. Neurobiol.* **9**, 164–170.
- Kaukoranta, E., Hari, R., and Lounasmaa, O. V. (1987). "Responses of the human auditory cortex to vowel onset after fricative consonants," *Exp. Brain Res.* **69**, 19–23.
- Kewley-Port, D., Watson, C. S., and Foyle, D. C. (1988). "Auditory temporal acuity in relation to category boundaries; speech and nonspeech stimuli," *J. Acoust. Soc. Am.* **83**, 1133–1145.
- Klinke, R., Kral, A., Heid, S., Tillein, J., and Hartmann, R. (1999). "Recruitment of the auditory cortex in congenitally deaf cats by long-term cochlear electrostimulation," *Science* **285**, 1729–1733.

- Kluender, K. R. (1991). "Effects of first formant onset properties on voicing judgments result from processes not specific to humans," *J. Acoust. Soc. Am.* **90**, 83–96.
- Kluender, K. R., and Lotto, A. J. (1994). "Effects of first formant onset frequency on [-voice] judgments result from auditory processes not specific to humans," *J. Acoust. Soc. Am.* **95**, 1044–1052.
- Kluender, K. R., Lotto, A. J., and Jenison, R. L. (1995). "Perception of voicing for syllable-initial stops at different intensities: Does synchrony capture signal voiceless stop consonants?" *J. Acoust. Soc. Am.* **97**, 2552–2567.
- Kluender, K. R., Lotto, A. J., Holt, L. L., and Bloedel, S. L. (1998). "Role of experience for language-specific functional mappings of vowel sounds," *J. Acoust. Soc. Am.* **104**, 3568–3582.
- Kuhl, P. K. (1981). "Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories," *J. Acoust. Soc. Am.* **70**, 340–349.
- Kuhl, P. K., and Miller, J. D. (1975). "Speech perception by the chinchilla: Voiced–voiceless distinction in alveolar plosive consonants," *Science* **190**, 69–72.
- Kuhl, P. K., and Miller, J. D. (1978). "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli," *J. Acoust. Soc. Am.* **63**, 905–917.
- Kuhl, P. K., and Padden, D. M. (1982). "Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques," *Percept. Psychophys.* **32**, 542–550.
- Kuriki, S., Okita, Y., and Hirata, Y. (1995). "Source analysis of magnetic field responses from the human auditory cortex elicited by short speech sounds," *Exp. Brain Res.* **104**, 144–152.
- Lasky, R. E., Syrdal-Lasky, A., and Klein, R. E. (1975). "VOT discrimination by four- to six- and-a-half-month-old infants from Spanish environments," *J. Exp. Child Psychol.* **20**, 215–225.
- Le Prell, C. G., Niemiec, A. J., and Moody, D. B. (2001). "Macaque thresholds for detecting increases in intensity: Effects of formant structure," *Hear. Res.* **162**, 29–42.
- Liégeois-Chauvel, C., de Graaf, J. B., Laguitton, V., and Chauvel, P. (1999). "Specialization of left auditory cortex for speech perception in man depends on temporal coding," *Cereb. Cortex* **9**, 484–496.
- Lisker, L. (1975). "Is it VOT or a first-formant transition detector?" *J. Acoust. Soc. Am.* **57**, 1547–1551.
- Lisker, L., and Abramson, A. S. (1964). "A cross-language study of voicing in initial stops: Acoustical measurements," *Word* **20**, 384–422.
- Lotto, A. J., and Kluender, K. R. (2002). "Synchrony capture hypothesis fails to account for effects of amplitude on voicing perception," *J. Acoust. Soc. Am.* **111**, 1056–1062.
- Lu, T., Liang, L., and Wang, X. (2001). "Temporal and rate representations of time-varying signals in the auditory cortex of awake monkeys," *Nat. Neurosci.* **4**, 1131–1138.
- Madler, C., Keller, I., Schwender, D., and Pöppel, E. (1991). "Sensory information processing during general anaesthesia: Effect of isoflurane on auditory evoked neuronal oscillations," *Br. J. Anaesth.* **66**, 81–87.
- Mäkelä, J. P., Hari, R., and Leinonen, L. (1988). "Magnetic responses of the human auditory cortex to noise/square wave transitions," *Electroencephalogr. Clin. Neurophysiol.* **69**, 423–430.
- Mäkelä, J. P., Karmos, G., Molnár, M., Csépe, V., and Winkler, I. (1990). "Steady-state responses from the cat auditory cortex," *Hear. Res.* **45**, 41–50.
- McGee, T., Kraus, N., King, C., and Nicol, T. (1996). "Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe," *J. Acoust. Soc. Am.* **99**, 3606–3614.
- Merzenich, M. M., and Brugge, J. F. (1973). "Representation of the cochlear partition on the superior temporal plane of the macaque monkey," *Brain Res.* **50**, 275–296.
- Merzenich, M. M., Tallal, P., Peterson, B., Miller, S., and Jenkins, W. M. (1999). "Some neurological principles relevant to the origins of—and the cortical plasticity-based remediation of—developmental language impairments," in *Neuronal Plasticity: Building A Bridge From The Laboratory To The Clinic*, edited by J. Grafman and Y. Christen (Springer, Berlin), pp. 169–187.
- Metherate, R., and Ashe, J. H. (1994). "Facilitation of an NMDA receptor-mediated EPSP by paired-pulse stimulation in rat neocortex via depression of GABAergic IPSPs," *J. Physiol. (London)* **481**, 331–348.
- Metherate, R., and Ashe, J. H. (1995). "Synaptic interactions involving acetylcholine, glutamate, and GABA in rat auditory cortex," *Exp. Brain Res.* **107**, 59–72.
- Metherate, R., and Cruikshank, S. J. (1999). "Thalamocortical inputs trigger a propagating envelope of gamma-band activity in auditory cortex *in vitro*," *Exp. Brain Res.* **126**, 160–174.
- Miller, L. M., Escabi, M. A., Read, H. L., and Schreiner, C. E. (2001). "Functional convergence of response properties in the auditory thalamocortical system," *Neuron* **32**, 151–160.
- Mitani, A., Shimokouchi, M., Itoh, K., Nomura, S., Kudo, M., and Mizuno, N. (1985). "Morphology and laminar organization of electrophysiologically identified neurons in the primary auditory cortex in the cat," *J. Comp. Neurol.* **235**, 430–447.
- Morel, A., Garraghty, P. E., and Kaas, J. H. (1993). "Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys," *J. Comp. Neurol.* **335**, 437–459.
- Müller-Preuss, P., and Mitzdorf, U. (1984). "Functional anatomy of the inferior colliculus and the auditory cortex: Current source density analyses of click-evoked potentials," *Hear. Res.* **16**, 133–142.
- Nagarajan, S. S., Cheung, S. W., Bedenbaugh, P., Beitel, R. E., Schreiner, C. E., and Merzenich, M. M. (2002). "Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex," *J. Neurophysiol.* **87**, 1723–1737.
- Nelken, I., Prut, Y., Vaadia, E., and Abeles, M. (1994). "Population responses to multifrequency sounds in the cat auditory cortex: One- and two-parameter families of sounds," *Hear. Res.* **72**, 206–222.
- Ohlemiller, K. K., Jones, L. B., Heidbreder, A. F., Clark, W. W., and Miller, J. D. (1999). "Voicing judgments by chinchillas trained with a reward paradigm," *Behav. Brain Res.* **100**, 185–195.
- Pandya, D. N., and Rosene, D. L. (1993). "Laminar termination patterns of thalamic, callosal, and association afferents in the primary auditory area of the rhesus monkey," *Exp. Neurol.* **119**, 220–234.
- Parker, E. M. (1988). "Auditory constraints on the perception of voice-onset time: The influence of lower tone frequency on judgments of tone-onset simultaneity," *J. Acoust. Soc. Am.* **83**, 1597–1607.
- Petersen, R. S., Panzeri, S., and Diamond, M. E. (2002). "Population coding in somatosensory cortex," *Curr. Opin. Neurobiol.* **12**, 441–447.
- Phillips, D. P., and Hall, S. E. (1990). "Response timing constraints on the cortical representation of sound time structure," *J. Acoust. Soc. Am.* **88**, 1403–1411.
- Pisoni, D. B. (1977). "Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops," *J. Acoust. Soc. Am.* **61**, 1352–1361.
- Pisoni, D. B., and Lazarus, J. H. (1974). "Categorical and noncategorical modes of speech perception along the voicing continuum," *J. Acoust. Soc. Am.* **55**, 328–333.
- Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (1982). "Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants," *J. Exp. Psychol. Hum. Percept. Perform.* **8**, 297–314.
- Rauschecker, J. P., Tian, B., and Hauser, M. (1995). "Processing of complex sounds in the macaque nonprimary auditory cortex," *Science* **268**, 111–114.
- Recanzone, G. H., Schreiner, C. E., and Merzenich, M. M. (1993). "Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys," *J. Neurosci.* **13**, 87–103.
- Repp, B. H. (1979). "Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants," *Lang Speech* **22**, 173–189.
- Rose, H. J., and Metherate, R. (2001). "Thalamic stimulation largely elicits orthodromic, rather than antidromic, cortical activation in an auditory thalamocortical slice," *Neuroscience* **106**, 331–340.
- Rotman, Y., Bar-Yosef, O., and Nelken, I. (2001). "Relating cluster and population responses to natural sounds and tonal stimuli in cat primary auditory cortex," *Hear. Res.* **152**, 110–127.
- Schreiner, C. E. (1998). "Spatial distribution of responses to simple and complex sounds in the primary auditory cortex," *Audiol. Neuro-Otol.* **3**, 104–122.
- Schroeder, C. E., Tenke, C. E., Givre, S. J., Arezzo, J. C., and Vaughan, Jr., H. G. (1990). "Laminar analysis of bicuculline-induced epileptiform activity in area 17 of the awake macaque," *Brain Res.* **515**, 326–330.
- Sharma, A., and Dorman, M. F. (1999). "Cortical auditory evoked potential correlates of categorical perception of voice-onset time," *J. Acoust. Soc. Am.* **106**, 1078–1083.
- Sharma, A., Marsh, C. M., and Dorman, M. F. (2000). "Relationship between N1 evoked potential morphology and the perception of voicing," *J. Acoust. Soc. Am.* **108**, 3030–3035.
- Sinex, D. G., and McDonald, L. P. (1988). "Average discharge rate repre-

- sentation of voice onset time in the chinchilla auditory nerve," *J. Acoust. Soc. Am.* **83**, 1817–1827.
- Sinnott, J. M., and Adams, F. S. (1987). "Differences in human and monkey sensitivity to acoustic cues underlying voicing contrasts," *J. Acoust. Soc. Am.* **82**, 1539–1547.
- Sinnott, J. M., and Brown, C. H. (1997). "Perception of the American English liquid/ra-la/contrast by humans and monkeys," *J. Acoust. Soc. Am.* **102**, 588–602.
- Soli, S. D. (1983). "The role of spectral cues in the discrimination of voice onset time differences," *J. Acoust. Soc. Am.* **73**, 2150–2165.
- Steinschneider, M., Tenke, C., Schroeder, C., Javitt, D., Simpson, G. V., Arezzo, J. C., and Vaughan, Jr., H. G. (1992). "Cellular generators of the cortical auditory evoked potential initial component," *Electroencephalogr. Clin. Neurophysiol.* **84**, 196–200.
- Steinschneider, M., Schroeder, C., Arezzo, J. C., and Vaughan, Jr., H. G. (1994). "Speech-evoked activity in primary cortex: effects of voice onset time," *Electroencephalogr. Clin. Neurophysiol.* **92**, 30–43.
- Steinschneider, M., Schroeder, C. E., Arezzo, J. C., and Vaughan, Jr., H. G. (1995). "Physiologic correlates of the voice onset time (VOT) boundary in primary auditory cortex (A1) of the awake monkey: Temporal response patterns," *Brain Lang* **48**, 326–340.
- Steinschneider, M., Reser, D. H., Fishman, Y. I., Schroeder, C. E., and Arezzo, J. C. (1998). "Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception," *J. Acoust. Soc. Am.* **104**, 2935–2955.
- Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard, M. A., III (1999). "Temporal encoding of the voice onset time (VOT) phonetic parameter by field potentials recorded directly from human auditory cortex," *J. Neurophysiol.* **82**, 2346–2357.
- Steinschneider, M., Volkov, I., Zotova, E., Bakken, H., Fishman, Y., Damasio, H., Arezzo, J. C., and Howard, III, M. (2000). "Differential representation of the voice onset time phonetic parameter within human auditory cortex and a monkey model," *Soc. Neurosci. Abst.* **26**(2), 1969.
- Stevens, K. N. (1981). "Constraints imposed by the auditory system on the properties used to classify speech sounds: Data from phonology, acoustics, and psychoacoustics," in *The Cognitive Representation of Speech*, edited by T. Myers, J. Laver, and J. Anderson (North-Holland, Amsterdam), pp. 61–74.
- Sugimoto, S., Hosokawa, Y., Horikawa, J., Nasu, M., and Taniguchi, I. (2002). "Spatial focusing of neuronal responses induced by asynchronous two-tone stimuli in the guinea pig auditory cortex," *Cereb. Cortex* **12**, 506–514.
- Summerfield, Q. (1981). "Articulatory rate and perceptual constancy in phonetic perception," *J. Exp. Psychol. Hum. Percept. Perform.* **7**, 1074–1095.
- Summerfield, Q., and Haggard, M. (1977). "On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants," *J. Acoust. Soc. Am.* **62**, 435–448.
- Talwar, S. K., and Gerstein, G. L. (2001). "Reorganization in awake rat auditory cortex by local microstimulation and its effect on frequency-discrimination behavior," *J. Neurophysiol.* **86**, 1555–1572.
- Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). "Functional specialization in rhesus monkey auditory cortex," *Science* **292**, 290–293.
- Tremblay, K., Kraus, N., Carrell, T. D., and McGee, T. (1997). "Central auditory system plasticity: Generalization to novel stimuli following listening training," *J. Acoust. Soc. Am.* **102**, 3762–3773.
- Usrey, W. M. (2002). "The role of spike timing for thalamocortical processing," *Curr. Opin. Neurobiol.* **12**, 411–417.
- Utman, J. A. (1998). "Effects of local speaking rate context on the perception of voice-onset time in initial stop consonants," *J. Acoust. Soc. Am.* **103**, 1640–1653.
- Vaughan H. G., Jr. and Arezzo, J. C. (1988). "The neural basis of event-related potentials," in *Handbook of Electroencephalography and Clinical Neurophysiology, Vol. 3, Human Event-Related Potentials*, edited by T. W. Picton (Elsevier, Amsterdam), pp. 45–96.
- Wallace, M. N., Kitzes, L. M., and Jones, E. G. (1991). "Intrinsic inter- and intralaminar connections and their relationship to the tonotopic map in cat primary auditory cortex," *Exp. Brain Res.* **86**, 527–544.
- Wang, X., Merzenich, M. M., Beitel, R., and Schreiner, C. E. (1995). "Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics," *J. Neurophysiol.* **74**, 2685–2706.
- Weinberger, N. M. (1997). "Learning-induced receptive field plasticity in the primary auditory cortex," *Semin. Neurosci.* **9**, 59–67.
- Zurita, P., Villa, A. E. P., de Ribaupierre, Y., de Ribaupierre, F., and Rouiller, E. M. (1994). "Changes of single unit activity in the cat's auditory thalamus and cortex associated to different anesthetic conditions," *Neurosci. Res. (NY)* **19**, 303–316.