Temporal Encoding of the Voice Onset Time Phonetic Parameter by Field Potentials Recorded Directly From Human Auditory Cortex

MITCHELL STEINSCHNEIDER,[1] IGOR O. VOLKOV,[2] M. DANIEL NOH,[2] P. CHARLES GARELL,[2] AND
MATTHEW A. HOWARD, III[2]

[1]*Departments of Neurology and Neuroscience, Albert Einstein College of Medicine, Bronx, New York 10461; and*
[2]*Department of Surgery (Division of Neurosurgery), University of Iowa College of Medicine, Iowa City, Iowa 52242*

Steinschneider, Mitchell, Igor O. Volkov, M. Daniel Noh, P. Charles Garell, and Matthew A. Howard, III. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J. Neurophysiol.* 82: 2346–2357, 1999. Voice onset time (VOT) is an important parameter of speech that denotes the time interval between consonant onset and the onset of low-frequency periodicity generated by rhythmic vocal cord vibration. Voiced stop consonants (/b/, /g/, and /d/) in syllable initial position are characterized by short VOTs, whereas unvoiced stop consonants (/p/, /k/, and t/) contain prolonged VOTs. As the VOT is increased in incremental steps, perception rapidly changes from a voiced stop consonant to an unvoiced consonant at an interval of 20–40 ms. This abrupt change in consonant identification is an example of categorical speech perception and is a central feature of phonetic discrimination. This study tested the hypothesis that VOT is represented within auditory cortex by transient responses time-locked to consonant and voicing onset. Auditory evoked potentials (AEPs) elicited by stop consonant-vowel (CV) syllables were recorded directly from Heschl's gyrus, the planum temporale, and the superior temporal gyrus in three patients undergoing evaluation for surgical remediation of medically intractable epilepsy. Voiced CV syllables elicited a triphasic sequence of field potentials within Heschl's gyrus. AEPs evoked by unvoiced CV syllables contained additional response components time-locked to voicing onset. Syllables with a VOT of 40, 60, or 80 ms evoked components time-locked to consonant release and voicing onset. In contrast, the syllable with a VOT of 20 ms evoked a markedly diminished response to voicing onset and elicited an AEP very similar in morphology to that evoked by the syllable with a 0-ms VOT. Similar response features were observed in the AEPs evoked by click trains. In this case, there was a marked decrease in amplitude of the transient response to the second click in trains with interpulse intervals of 20–25 ms. Speech-evoked AEPs recorded from the posterior superior temporal gyrus lateral to Heschl's gyrus displayed comparable response features, whereas field potentials recorded from three locations in the planum temporale did not contain components time-locked to voicing onset. This study demonstrates that VOT at least partially is represented in primary and specific secondary auditory cortical fields by synchronized activity time-locked to consonant release and voicing onset. Furthermore, AEPs exhibit features that may facilitate categorical perception of stop consonants, and these response patterns appear to be based on temporal processing limitations within auditory cortex. Demonstrations of similar speech-evoked response patterns in animals support a role for these experimental models in clarifying selected features of speech encoding.

INTRODUCTION

Recent technological advances have invigorated the investigation of neural mechanisms associated with speech perception. One exciting avenue of research involves the expanded use of functional neuroimaging to define regions of brain participating in specific aspects of language processing (e.g., Binder et al. 1996; Wise et al. 1991). These investigations have implicated both primary and secondary auditory cortex as important components of a neural network subserving the initial cortical stages of phonetic processing (Zatorre et al. 1992, 1996). Event-related potentials and neuromagnetic data have complemented functional neuroimaging studies by identifying temporally dynamic features of speech sound encoding in auditory cortex and by using physiological markers to address important controversies surrounding phonetic perception (e.g., Kaukoranta et al. 1987; Kuriki et al. 1995; Poeppel et al. 1997; Sharma and Dorman 1998; Sharma et al. 1993). Methodological constraints of these techniques, however, preclude resolution of the specific synaptic events that activate auditory cortex, generate the evoked responses, or underlie the phonetic discrimination. These considerations reinforce the need for more detailed physiological investigations of phonetic encoding that can best be obtained by performing invasive studies of auditory cortex.

Important features of phonetic discrimination that mirror human results occur in monkeys and other experimental animals (e.g., Hienz et al. 1996; Kuhl and Padden 1983; Lotto et al. 1997; Sinnott et al. 1997; Sommers et al. 1992). Findings such as these suggest that physiological studies in animals can reveal basic mechanisms underlying aspects of human speech perception common to multiple species. Similarities in speech sound discrimination have been especially well documented for perception of the voice onset time (VOT) phonetic parameter (Kluender and Lotto 1994; Kuhl and Miller 1978; Kuhl and Padden 1982; Sinnott and Adams 1987). VOT denotes the time interval between consonant release and the onset of low-frequency periodicity in the speech sound generated by rhythmic glottal pulsations. Voiced stop consonants (/b/, /d/, and /g/) are characterized by short-duration VOTs, whereas unvoiced consonants (/p/, /t/, and /k/) contain longer VOTs. As the VOT is increased in incremental steps, the perception abruptly changes between 20 and 40 ms from a voiced consonant (e.g., /d/) to an unvoiced consonant (e.g., /t/). This abrupt transition in consonant identification is an example of categorical speech perception, an essential feature of phonetic discrimination. The ability of animals to discriminate consonants differing in their VOT in a manner similar to humans targets this phonetic parameter as a prime candidate for elucidating relevant neural events associated with its categorical perception.

Studies undertaken in primary auditory cortex (A1) of the awake macaque monkey have suggested a mechanism by which the VOT phonetic parameter can be encoded rapidly in a categorical manner (Steinschneider et al. 1994, 1995b). These studies found that acoustical transients associated with consonant release and voicing onset are represented in the temporal response patterns of neuronal ensembles. Consonant-vowel syllables with short VOTs evoked short-latency responses primarily time-locked to consonant release alone. In contrast, consonant-vowel syllables with longer VOTs evoked responses at the same cortical loci time-locked to both consonant release and voicing onset. These neural patterns led to the hypothesis that categorical perception of consonants varying in their VOT is based partially on temporal encoding mechanisms within A1. The occurrence of two transient response bursts time-locked to both consonant release and voicing onset would signal an unvoiced stop consonant, whereas voiced stop consonants would be represented by a single response time-locked only to consonant release. The VOT where the second response to voicing onset dissipates would denote the categorical boundary for this phonetic contrast. Of special note was the replication of this activity pattern in the auditory evoked potential (AEP), suggesting that similar responses could be recorded in humans.

While speech-evoked activity in monkeys may be relevant for phonetic encoding of VOT in humans, these observations could represent simply an epiphenomenon of acoustical transient processing in A1. Support for the significance of these responses would be a demonstration that similar activity patterns occur in the human auditory cortex. Neurons in both A1 and secondary auditory cortex can respond in select ways to behaviorally relevant species-specific vocalizations (Rauschecker 1998; Rauschecker et al. 1995; Wang et al. 1995). Learning and neuronal response plasticity induced by exposure to behaviorally relevant sounds also modify the activity of auditory cortical cells (Ohl and Scheich 1997; Recanzone et al. 1993; Weinberger 1997). Therefore response patterns generated in human auditory cortex by biologically important speech sounds may bear little resemblance to the activity profiles seen in the naïve monkey. This study investigates whether response patterns evoked by speech sounds varying in their VOT are similar in the monkey and human. Speech-evoked AEPs were recorded directly from human auditory cortex in three patients undergoing evaluation for surgical intervention of medically intractable epilepsy. Specifically, we tested the hypothesis that acoustical transients associated with consonant release and voicing onset are represented by temporally discrete responses in human auditory cortex. Further, we tested whether these responses would display categorical-like features similar to those seen in the monkey.

METHODS

Subjects

Three right-handed patients with medically intractable epilepsy were studied. Experimental protocols were approved by the University of Iowa Human Subjects Review Board and reviewed by the National Institutes of Health and informed consent was obtained from each subject before their participation. Subjects underwent placement of intracranial electrodes for acquiring diagnostic electroencephalographic (EEG) information required to plan subsequent surgical treat-

ment. Research recordings did not disrupt simultaneous gathering of clinically necessary data, and patients did not undergo any additional risk by participating in this study. Patient information is summarized in Table 1. All had a suspected epileptic focus in or near the auditory cortex of the right hemisphere, and all had normal hearing determined by standard audiometric tests. *Patient 2* complained that occasional seizures were precipitated by acoustic stimuli. Recording sessions were carried out in a quiet room in the Epilepsy Monitoring Unit of the University of Iowa Hospitals and Clinics with the patients lying comfortably in their hospital beds. Patients were awake and alert throughout the recordings but were not asked to perform any specific task beyond listening to the presented sounds. For all patients, clinical evaluations ultimately indicated that Heschl's gyrus and nearby auditory cortical tissue were not epileptogenic foci and were not targeted for surgical extirpation.

Recording methods

Auditory evoked potentials (AEPs) were recorded at a gain of 5,000 using headstage amplification followed by differential amplification (BAK Electronics, Germantown, MD). Specific methods varied among the three subjects. This limitation was partly based on time and other technical constraints imposed by the clinical needs of the patients as well as by an on-going process to optimize recording parameters. Recordings were obtained in *patient 1* from three EEG ring contacts spaced at 10 mm intervals incorporated into a depth electrode implanted in the right Heschl's gyrus and from subdural grid electrodes placed over the lateral convexity of the temporal lobe. The reference electrode was a subdural grid electrode located on the undersurface of the ipsilateral, anterior temporal lobe. AEPs were recorded at a band-pass of 2–500 Hz (3 dB down, roll-off 6 dB/octave) and digitization rate of 1,000 Hz. *Patient 2* was studied with hybrid clinical-research depth electrodes implanted in the right Heschl's gyrus and planum temporale (Howard et al. 1996a,b). Bipolar recordings at three depths were performed from closely spaced high-impedance recording contacts interspersed between the standard, clinical EEG contacts (Radionics, Burlington, MA) at a band-pass of 1–5,000 Hz and digitization rate of 2,050 Hz. In both patients, depth electrodes were placed stereotaxically along the long axis of Heschl's gyrus. AEPs in *patient 3* were obtained from a subdural grid electrode placed over the lateral convexity of the posterior temporal lobe, referenced to the deepest low-impedance contact of a depth electrode located anterior to Heschl's gyrus (band-pass, 1–3,000 Hz, digitization rate, 2,000 Hz). This subdural electrode was chosen over other surrounding contacts in the grid after preliminary mapping of AEPs demonstrated focal activity restricted to this site. The relative inactivity of the reference was determined by recording speech-evoked AEPs between this site and a subdural electrode located beneath the temporal lobe. Field potentials were computer averaged using an analysis time of 1,000 ms (300 ms prestimulus) for *patient 1* or an

TABLE 1. *Patient summary*

Subject	Age	Sex	Language Dominance*	Seizure Types
1	32	M	Left	Complex partial (20)
2	30	F	Left	Complex partial (13)
3	37	M	Bilateral	Complex partial, generalized tonic-clonic (24)

Subject	Recording Electrodes
1	Depth electrode in right Heschl's gyrus; right posterior superior temporal gyrus subdural grid electrodes
2	Depth electrodes in right Heschl's gyrus and planum temporale
3	Right posterior superior temporal gyrus subdural grid electrode

Duration of seizures in years is in parentheses. * AngioWADA test.
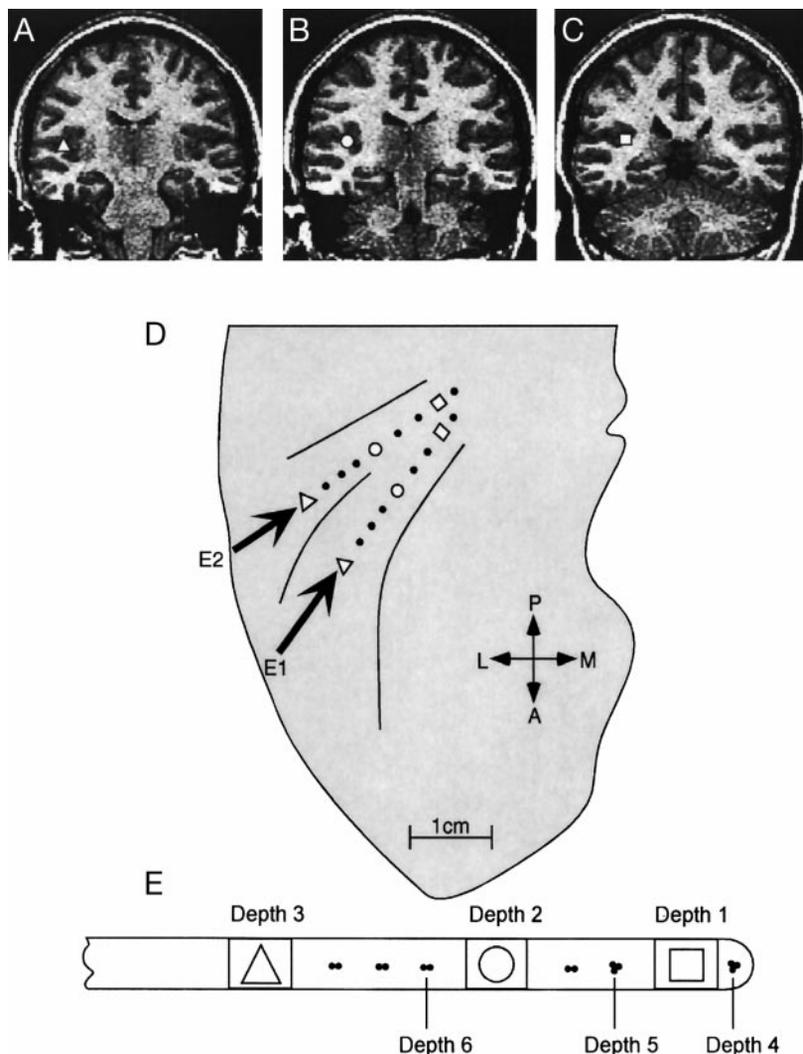
FIG. 1. Anatomic data demonstrating the locations of the intracortical electrodes in *patient 2*. Electrode configuration with 3 electroencephalographic (EEG) recording contacts (△, ○, and □) and multiple microelectrode recording sites is depicted in the schematic diagram (*E*). Coronal magnetic resonance imaging (MRI) images anterior (*A*), middle (*B*), and posterior (*C*) show locations of the 3 EEG contacts of the most anteriorly positioned electrode (E1) within right-sided Heschl's gyrus. *D*: schematic view of the tracks for the E1 electrode and an additional probe placed in the planum temporale (E2). Diagram represents a scaled tracing of a 3-dimensional MRI image of the exposed superior temporal plane. *E*: schematic of the intracortical electrode. Recordings were obtained from the low-impedance contacts in *patient 1* (*depths 1–3*), and 3 of the high-impedance sites in *patient 2* (*depths 4–6*).

analysis time of 500 ms (25 ms prestimulus) for *patients 2* and *3*. Averages were generated from 50 to 75 stimulus presentations, and high-amplitude artifacts led to automatic rejection of the presentation for inclusion into the averaged responses. Raw EEG and appropriate timing pulses were stored on a multichannel FM tape recorder (Racal, Irvine, CA) for subsequent analysis. An independent search for isolated single units and their responses to pure tones also was performed in parallel with this AEP experiment (Howard et al. 1996a).

Anatomic locations of all recording sites were identified on postimplantation MRIs. This procedure is illustrated in Fig. 1, which demonstrates implantation in *patient 2* of the depth electrodes in Heschl's gyrus and the planum temporale. The coronal MRI scans (*A–C*) show the locations of the three EEG contacts in Heschl's gyrus, denoted by the geometric shapes in the images. Figure 1*D* is a schematic diagram of the superior temporal plane, and the tracks for the electrodes in Heschl's gyrus (E1) and planum temporale (E2) in *patient 2*. A schematic of the intracortical electrode is shown in Fig. 1*E*. The three recording contacts in *patient 1* are denoted as *depths 1–3*, whereas *depths 4–6* indicate the three recording sites with high-impedance wires from which AEPs were recorded in *patient 2*.

Stimuli

Two sets of synthetic speech stimuli previously used in primate studies were presented to the subjects. The first set consisted of all six stop consonants followed by the vowel /a/. They were constructed on the parallel branch of a KLSYN88a speech synthesizer. Frequency characteristics of the voiced consonant-vowel (CV) syllables (/ba/, /ga/, and /da/) have been published previously (Steinschneider et al. 1995a). Steady-state formant frequencies were 700, 1,200, 2,500, and 3,600 Hz. Onset frequencies for the second (F2) and third (F3) formants of /ba/ were 800 and 2,000 Hz and 1,600 and 3,000 Hz for /da/. Starting frequencies for /ga/ were 1,600 and 2,000 Hz. Onset frequency of the first formant (F1) was 200 Hz for all syllables. F1 transition duration was 30 ms and 40 ms for F2 and F3 transitions. Formant structure was such that /ba/ and /da/ had diffuse onset spectra maximal at either low or high frequencies, whereas /ga/ had a compact onset spectrum maximal at intermediate values (Stevens and Blumstein 1978). The unvoiced CV syllables (/pa/, /ka/, and /ta/) were identical to their voiced counterparts except for an increase in the VOT from 5 to 40 ms. The first 5 ms of all six syllables consisted of frication. For the unvoiced CV syllables, the next 35 ms contained aspiration noise. The second set of speech stimuli were three formant syllables /da/ and /ta/ produced at the Haskins Laboratories (New Haven, CT). Five syllables with VOT varying from 0 to 80 ms in 20 ms increments were created. Specific parameters of the sounds have been published (Steinschneider et al. 1995b). Steady-state formant frequencies were 817 Hz for F1, 1,181 Hz for F2, and 2,632 Hz for F3 and onset frequencies were 200, 1,835, and 3,439 Hz. All syllables were 175 ms in duration and presented by computer. They were delivered binaurally by insert ear phones (Etmotic Research, Elk Grove Village, IL) in *patient 1* and to the ear contralateral to the

recording sites by a Koss K240DF headphone coupled to a 4-cm cushion in *patients 2* and *3.* Syllable intensity was 70 dB SPL for *patient 1* and 80 dB SPL for the other two subjects. Perceptual testing of multiple listeners in the laboratory of the lead investigator indicated that syllables with a VOT of 0 and 20 ms consistently sounded like /da/, whereas those with a longer VOT were perceived as /ta/. Informal questioning of *patient 2* during recording yielded identical perceptions, whereas *patient 3* gave variable responses only for /da/ with the 20-ms VOT, indicating that this stimulus was near the perceptual boundary. *Patient 1* was not questioned.

RESULTS

Neural representation of stop consonant-vowel syllables in Heschl's gyrus

Voiced CV syllables (/ba/, /ga/, and /da/) elicit a triphasic sequence of field potentials in Heschl's gyrus. These potentials are not specific to speech sounds and occur in response to other sounds such as tone and noise bursts. AEPs are illustrated in Fig. 2, which depicts *patient 1* responses evoked by the six voiced and unvoiced stop CV syllables recorded simultaneously from the three EEG contacts within Heschl's gyrus. Onset and peak latencies of the three AEP components are earlier at more medial recording sites. Cortical activity begins at 11–12 ms at the most medial depth recording site (*depth 1*), and 19–21 ms at more lateral sites (*depths 2* and *3*). The first component is a small wave of positive polarity (wave A) that peaks at 26–28 ms at *depth 1,* 34–37 ms at *depth 2,* and 43–47 ms at *depth 3.* An additional positive deflection peaking at 55–59 ms also can be observed at *depth 2.* These positive waves are followed by a large amplitude negativity (wave B) peaking at 56–62 ms at *depth 1,* and 88–95 ms at the other two sites. The third component is a large amplitude positivity (wave C) peaking at 140 ms at depth 1, and 144–155 ms at *depths 2* and *3.* Additional potentials time-locked to stimulus offset conclude the responses (asterisks).

Unvoiced CV syllables (/pa/, /ka/, and /ta/) evoke a different pattern of activity that reflects both stimulus onset and the onset of voicing (Fig. 2). Thus the initial activity elicited by consonant onset consists of the same positive wave A. However, the following negativity (wave B) is truncated and replaced by a positive-going (solid arrows) and a negative-going wave (unfilled arrows) time-locked to the 40 ms VOT. Additionally, wave C peaks 40 ms later than when elicited by voiced CV syllables. These findings indicate that VOT is represented in Heschl's gyrus by synchronized activity time-locked to the onset of consonant release and voicing onset.

Observations are extended by the AEPs recorded at the three locations in Heschl's gyrus of *patient 2* (Fig. 3). Because these AEPs are recorded between adjacent high-impedance contacts, amplitudes of the responses are diminished, depicted polarity of the waveforms is arbitrary, and phase shifts of the response peaks are likely. The strength of this method is the marked attenuation of far-field potentials and accentuation of local activity at the recording sites. Field potentials at *depth 4,* the most medial of the recording sites, are illustrated with a recording montage that places the electrode contact evoking the largest amplitude response as the "active" electrode (contact 2), and the contact eliciting the smallest response as the reference. With this montage, waves A–C are clearly identified. Peak latencies for wave A are 27 ms for /da/, 28 ms for /ga/,
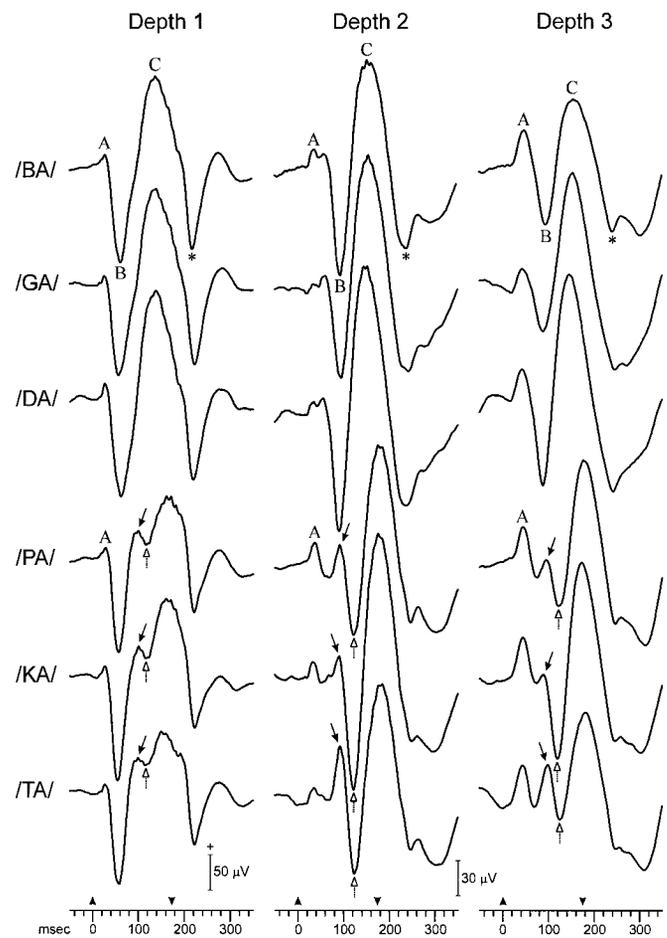
FIG. 2. *Patient 1*: auditory evoked potentials (AEPs) evoked by the 6 stop consonant-vowel (CV) syllables and recorded from 3 low-impedance electrode sites in Heschl's gyrus. *Depth 1* is the most medial site, whereas *depths 2* and *3* are progressively more anterolateral in increments of 10 mm. Response morphology is similar for the AEPs elicited by the 3 voiced CV syllables, /ba/, /ga/, and /da/, and contains 3 components (A, B, and C). A negativity evoked by syllable offset concludes the responses (asterisks). AEPs evoked by the 3 unvoiced CV syllables, /pa/, /ka/, and /ta/, contain additional components time-locked to voicing onset (solid and unfilled arrows). Separation between the initial waves and these latter components equal the 40-ms voice onset time (VOT). Arrowheads just *above* the timelines denote stimulus onset and offset. As for all figures, electrical positivity is upward.

and 37 ms for /ba/. A preceding positivity is present in the response to /da/ that peaks at 16 ms and can be seen as a positive deflection on the other responses. Wave B follows the initial activity and peaks at 54 ms for /da/, 57 ms for /ga/, and 63 ms for /ba/. Wave C peaks between 102 and 108 ms. The near-field responses also contain superimposed, low-amplitude waves phase-locked to the syllable fundamental frequency (f_0). Response components time-locked to the offset of the syllables conclude the response. No consistent response time-locked to voicing onset is present. AEPs at *depth 5,* located 6 mm anterolateral from *depth 4,* are dominated by a near-field component with peak latencies of 63–66 ms that overlap in time with wave B. Low-amplitude oscillatory activity phase-locked to the syllable f_0 is superimposed on the slower components, but responses to voicing onset are also absent.

Local field potentials evoked by the voiced CV syllables at *depth 6,* 14 mm anterolateral from *depth 4,* contain a large amplitude positive wave peaking at 42–47 ms. The positive
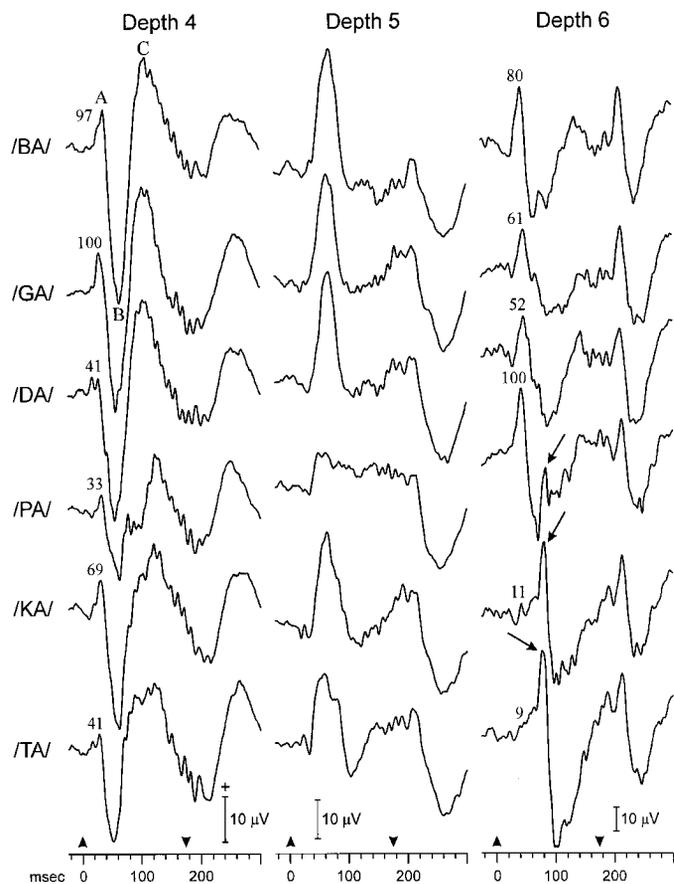
FIG. 3. *Patient 2*: AEPs evoked by the 6 stop CV syllables and recorded from 3 high-impedance electrode sites in Heschl's gyrus. Responses at the most medial location, *depth 4*, contain AEP components similar to those seen in *patient 1* (Fig. 2). Responses reflecting the extended VOT for the unvoiced CV syllables are restricted to the lateral site, *depth 6* (→). Numbers *above* the initial AEP components at *depths 4* and *6* denote the percent of maximum for these responses. In relation to the independently determined tonotopic sensitivity of these sites, the AEPs to syllables reflect the spectral characteristics of the stop consonants. See text for details.

wave is followed by a large negativity with variable peak latency (62–86 ms), which in turn is succeeded by a later positivity peaking at 134–144 ms for the voiced CV syllables. These components overlap in time with waves A–C recorded at the lateral sites in *patient 1*. The most relevant finding at this site is the additional positivity time-locked to voicing onset for the three unvoiced CV syllables (→). This finding, based on bipolar recordings between adjacent contacts, indicates that representation of VOT by synchronized activity to consonant release and voicing onset can occur at the same auditory cortical sites.

Spectral features of the CV syllables also are reflected at these sites. These findings mirror previous observations in the monkey (Steinschneider et al. 1995a). Independent analysis detected and isolated multiple units at *depths 4* and *6*. Maximum tone responses of these units were $2,125 \pm 252$ and 736 ± 91 Hz (means $\pm$ SD), respectively, and are in accord with findings that higher frequencies are encoded at more posteromedial locations in human A1 (Howard et al. 1996a). The initial component in the AEP is largest to /ga/ at *depth 4*. The positivity evoked by /ba/ is almost identical in amplitude but is reduced to only 41% in the response to /da/ (% maximum

response indicated in Fig. 3). Similarly, the largest initial response to the unvoiced CV syllables is to /ka/ followed by /ta/ and /pa/. This pattern is generally consistent with the spectral content at consonant release for the syllables and the tonotopic sensitivity of the sites. /Ga/ and /ka/ have a spectral maximum at midfrequencies that overlap the maximum tone sensitivity of *depth 4*. In contrast, the simultaneous recordings to the CV syllables at *depth 6* are maximal to the labial consonants /b/ and /p/. These consonants have energy concentrated in lower frequencies that overlap the tonal sensitivity of the site. Note that the additional responses at *depth 6* reflecting the extended VOT for the unvoiced CV syllables are present regardless of the preceding consonant place of articulation. These findings suggest parallel representation for the spectral feature of place of articulation and the temporal feature of voicing onset in human auditory cortex. Furthermore, the tonotopic sensitivity at *depth 6* suggests that voicing onset may be preferentially represented in lower best frequency regions of Heschl's gyrus.

Neural representation of VOT in Heschl's gyrus

AEPs at *depth 6* also exhibit marked changes in response morphology that suggest a differential representation of VOT evoked by the voiced consonant /d/ and the unvoiced consonant /t/ (Fig. 4). Figure 4, *left*, depicts the responses to the syllables /da/ and /ta/ with VOTs of 0–80 ms in 20 ms
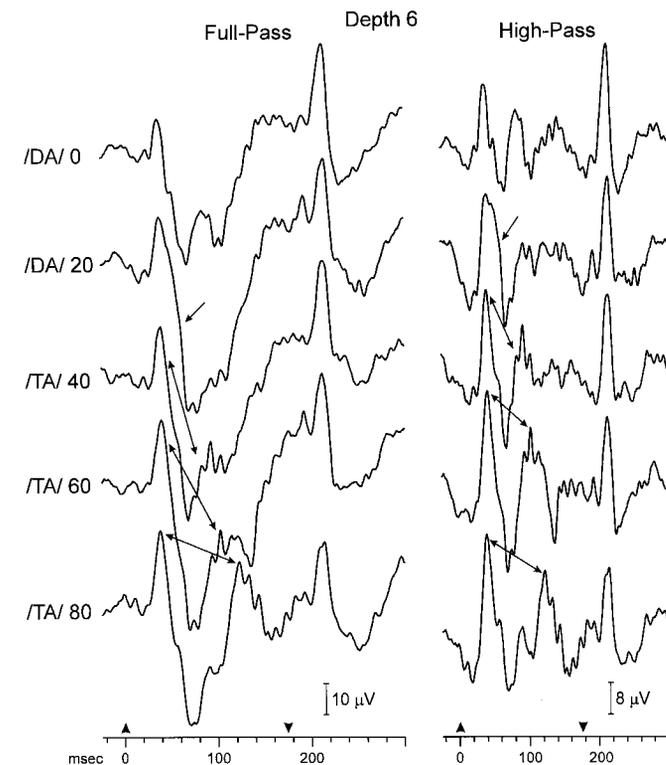
FIG. 4. *Patient 2*: categorical-like AEPs evoked by the 5 CV syllables that vary in their VOT and recorded from *depth 6*. AEPs evoked by the 3 /ta/ stimuli contain positive-going response components time-locked to consonant release and voicing onset (↔). /Da/ with the 20-ms VOT does not elicit a discrete response to voicing onset. →, location of the expected response to voicing onset. These effects are observed in the full-frequency waveforms (*left*), and the AEPs high-pass filtered >10 Hz to minimize the slower frequency components (*right*).

increments. Field potentials evoked by the /ta/ stimuli with 40-, 60-, and 80-ms VOTs contain two positive polarity components that are time-locked to stimulus onset and the onset of voicing (↔). There is a progressive 20 ms shift of the second component time-locked to voicing onset as the VOT is reduced from 80 to 60 and 40 ms. In contrast, the AEP evoked by /da/ with the 20-ms VOT fails to exhibit a discrete response time-locked to voicing onset. The expected location of this response is marked by the single-headed arrow. To ensure that the time-locked response to voicing onset is not masked by the large-amplitude negative slow wave that follows the initial positivity, the AEPs were high-pass filtered >10 Hz. This filter setting was chosen after Fourier analysis indicated that the slow wave energy was <10 Hz. Filtered responses are shown in Fig. 4, *right.* The /ta/ stimuli still evoke response components time-locked to both consonant release and voicing onset (↔), whereas /da/ with the 20-ms VOT fails to evoke a discrete response to voicing onset (→). Instead, the initial positive wave contains a shoulder of activity marking the location of the response to voicing onset.

Similar changes in AEP morphology are replicated in the field potentials recorded by the low-impedance electrode contacts in *patient 1* (Fig. 5). AEPs evoked by the stimuli with a prolonged VOT (40–80 ms) and recorded at *depth 2* display positive (solid arrow) and negative (asterisk) components time-locked to voicing onset that are delayed from the initial response complex by an amount equal to the VOT. The syllable with the 20-ms VOT does not elicit a well-defined response to voicing onset (open arrow) and instead evokes a response more similar to that evoked by /da/ with the 0-ms VOT. The syllables evoke a similar pattern of activity at *depth 3,* though the response to voicing onset for the 40-ms VOT stimulus is seen only as a positive-going deflection on the following positivity (→). Although differences are noted in the AEPs at *depth 1,* no pattern reflecting a discrete representation of voicing onset is evident. This suggests that the synchronized response pattern reflecting VOT is generated in specific regions of auditory cortex and is supported by a similar localization of activity exhibited in the AEPs recorded from *patient 2.*

Findings are highlighted by the 10-Hz high-pass-filtered AEPs evoked by the VOT series and recorded from *depths 2* and *1* (Fig. 6, *A* and *B*). Activity at *depth 2* (*A*) contains prominent positive waves time-locked to voicing onset for the three stimuli with more prolonged VOTs (←), whereas a markedly diminished positivity, barely above baseline, is evoked by the syllable with the 20-ms VOT. In contrast, simultaneously recorded activity at *depth 1* (*B*) consists of prominent ON components and oscillatory responses phase-locked to the syllable f_0 that occur after a nearly constant delay after the response to consonant onset. Additional evidence revealing major differences between the responses evoked by consonants with short and long VOTs is shown in *C*. These waveforms are derived by subtracting the response evoked by the completely voiced syllable /da/ with the 0-ms VOT from each of the other AEPs. The waveforms derived using the /ta/-evoked responses contain a large-amplitude negative wave that shifts in peak latency by increments of 20 ms as the VOT is increased by identical amounts (←). These waves are diminished markedly when the response is derived from the AEP evoked by the syllable with the 20-ms VOT. Thus, activity evoked by the two short VOT syllables are nearly identical,

whereas those evoked by CV syllables with prolonged VOTs elicit profoundly different responses.

Previous studies examining magnetic responses evoked by syllables or two-tone analogues of the VOT continuum have reported a categorical-like decrease in amplitude for the major surface negative wave when VOT or tone separation increased from 20 to 40 ms (Simos et al. 1998a–c). Data from these studies were acquired after low-pass filtering to improve signal quality. We examined whether equivalent findings would be demonstrated in the intracortical data by digitally filtering the syllable-evoked responses <20 Hz, thus mimicking AEPs that might be observed with scalp recordings using typical filter settings. Filtered AEPs recorded at *depth 2* are shown in Fig. 6D. Similar to the noninvasive recordings of the previous studies, there is a decrease in the peak (↓) and trough-to-peak amplitudes for only the responses evoked by the prolonged VOT stimuli. Additionally, two sequential positive waves separated by the VOT interval are only observed for the responses evoked by /ta/ (*). These findings support the noninvasively acquired data and suggest that one reason for the amplitude decrement is a truncation of the principal slow waves by the introduction of components time-locked to voicing onset for the /ta/ stimuli.

Results for the low-pass filtered AEPs at all three depths in *patient 1* were quantified by measuring the maximum baseline-to-positive peak and trough-to-peak deflections in the waveforms. For both measures, there is a dramatic reduction in amplitude of the slow wave when VOT increases from 20 to 60 ms at all three sites (Fig. 7). The summed trough-to-peak
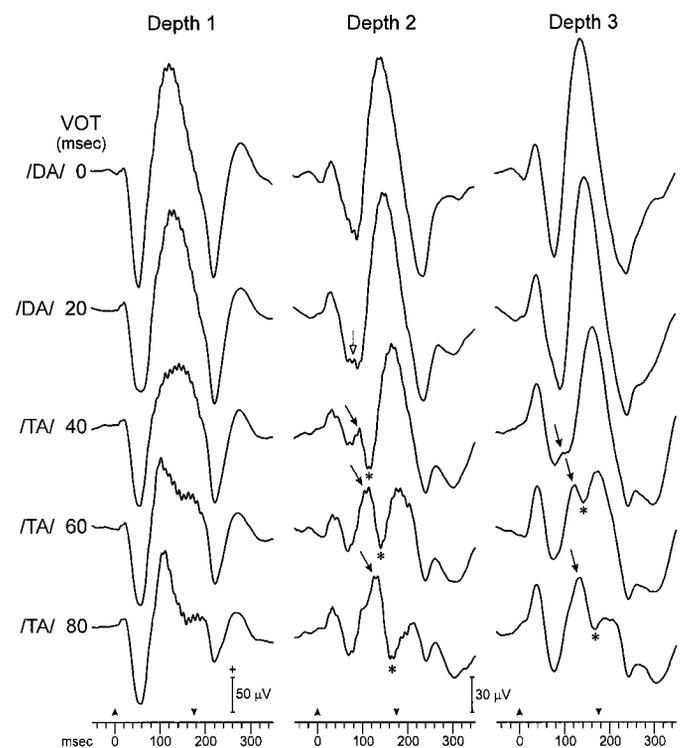
FIG. 5. AEPs evoked by the syllables varying in their VOT and recorded in Heschl's gyrus of *patient 1* also exhibit categorical-like features. Activity at *depths 2* and *3* contain positive-going (solid arrows) and negative-going (asterisks) components time-locked to voicing onset. These components are absent in the responses to the stimulus with the 20-ms VOT. Instead, the stimuli with a 0- and 20-ms VOT evoke AEPs with similar morphologies. The unfilled arrow marks the location of the expected positive-going wave at *depth 2.*

excursion was maximal for the syllables with the 0-ms (100%) and 20-ms (94%) VOTs. Syllables with the 40-, 60-, and 80-ms VOT intervals elicited responses 65, 52, and 54% of maximum, respectively. When baseline-to-peak measurements were made, syllables with the 0- and 20-ms VOT intervals evoked the largest responses (99 and 100%, respectively). Evoked responses for the 40-, 60-, and 80-ms VOT stimuli were 70, 47, and 52% of maximum, respectively.

Heschl's gyrus responses evoked by acoustic transients embedded in click trains

The previous data indicate that acoustic transients evoked by the onsets of consonant release and voicing are represented partially by synchronized responses within auditory cortex. Furthermore, the markedly diminished response evoked by

FIG. 7. Amplitudes of the trough-to-peak excursions of the major AEP positivity at all 3 depths recorded in *patient 1*. For all sites and for their sum, there is a nonlinear drop in amplitude as VOT changes from 0 to 40 ms. This pattern mirrors identification functions observed for the categorical perception of voiced from unvoiced stop consonants.

voicing onset embedded in the syllable with the 20-ms VOT suggests that there is a relative refractory time after stimulus onset when it is difficult to generate a second transient response. This refractory period may be partially responsible for defining the psychoacoustical boundary for voiced and unvoiced stop consonants. To further investigate this phenomenon, the ability of auditory cortex to respond to acoustic transients in the form of click trains was examined in *patient 2* (Fig. 8). Phase-locked responses to the individual clicks were present at rates up to 50 Hz at *depths 5* and *6* and 100 Hz at *depth 4*. Additionally, lower frequency click trains evoked well-defined second transient responses (solid arrows). However, there was a marked decrease in amplitude of the second phase-locked response at 40 and 50 Hz (unfilled arrows), corresponding to an interpulse interval of 20–25 ms. Further, the AEPs evoked by these click trains are similar in morphology to that evoked by the 100-Hz train. High-pass filtering the AEPs >10 Hz also failed to reveal a transient response peak to the second click in the higher frequency trains. The presence of time-locked activity to the individual clicks later in the waveforms, including trains of higher rates, suggest that stimulus onset engages a sequence of cortical events that inhibit the generation of a transient response within a time window similar to that of the VOT boundary.

VOT representation in nonprimary auditory cortex

Speech-evoked AEPs recorded from subdural strip electrodes placed on the lateral surface of the superior temporal gyrus in *patients 1* and *3* also contain features that accentuate differences between syllables with short and long VOTs (Fig. 9, *patient 3*). The illustrated electrode site was located lateral to the posterior margin of Heschl's gyrus. Only the syllables with a VOT of 40 to 80 ms evoke two positive waves separated by an interval equal to the VOT (*A*, 2nd peak denoted by solid arrows). These positive waves are preceded by two negative waves the separation of which also equals the extended VOT. In contrast, the syllables with a VOT of 0 and 20 ms elicit a single positive and negative wave complex. AEPs were high-

FIG. 6. *A*: *patient 1*, AEPs recorded at *depth 2* and high-pass filtered >10 Hz also display marked differences in the amplitude of the positivity time-locked to voicing onset (←). Responses are prominent with a VOT associated with the unvoiced consonant /t/ and nearly absent with the 20-ms VOT syllable. *B*: simultaneously recorded high-pass filtered AEPs at *depth 1* do not exhibit responses reflecting the VOT and instead contain an initial response time-locked to consonant release followed by oscillations phase-locked to the fundamental frequency. *C*: *depth 2* waveforms generated by subtracting the AEP evoked by the completely voiced syllable, /da/ with the 0-ms VOT, from waveforms elicited by the other syllables. A large component time-locked to voicing onset is only present for comparisons between activity of the /ta/ stimuli and /da/ with the 0-ms VOT (←). *D*: AEPs at *depth 2* low-pass filtered below 20 Hz to mirror recording parameters of noninvasive recordings also demonstrate categorical-like features. Positive-going waves time-locked to syllable and voicing onset are only evoked by the /ta/ stimuli (*). Note the decrease in amplitude of the major positivity with the /ta/ stimuli (↓). See text for details.

FIG. 8. AEPs evoked by click trains of various rates (*left*) and recorded at the 3 depths in *patient 2*. Prominent components time-locked to the 2nd click in the trains are only present for rates <40–50 Hz (solid arrows). Unfilled arrows mark the locations of the expected responses to the 2nd click at the higher rates. Note that phase-locked activity to the faster rates is present later in the waveforms, indicating that auditory cortex is capable of generating these responses. See text for details.

pass filtered >10 Hz to investigate whether similar waves are masked by slower frequency components for the 20-ms VOT stimulus (*B*). Again, the stimuli with prolonged VOTs evoke two prominent positive waves time-locked to consonant release and voicing onset (↔). The filtered AEP evoked by the syllable with the 20-ms VOT contains only a single positive wave preceded by a small positive-going deflection embedded in the initial negativity (asterisk). Similar findings also were observed in the high-pass filtered AEPs recorded from the subdural electrodes overlying the posterior portion of the superior temporal gyrus in *patient 1* (not shown).

Response components linearly related to the VOT also are present in the waveforms, indicating that auditory cortex is capable of generating responses that differentiate among all five syllables regardless of their consonant perception. This is evident in the progressive 20-ms shift of the later negative AEP component as VOT is increased from 0 to 80 ms (Fig. 9*A*, unfilled arrows). Additionally, a significant difference wave is generated when /da/ with the 0-ms VOT is subtracted from the syllable with the 20-ms VOT (*C*). The difference wave includes a positive component that shifts in peak latency with VOT (asterisks). Although this component is present in the difference wave between the two short VOT syllables, it is only 52% of the peak amplitude generated from /ta/ with the 60-ms VOT. The other /ta/ stimuli evoke a positive difference wave that is 85% of the maximum.

Several considerations suggest that the VOT-related activity recorded from the lateral subdural region reflect local responses and not just far-field AEPs generated in Heschl's gyrus. First, the reference electrode for the subdural recordings shown in Fig. 9 failed to record AEPs that reflected syllable VOT when it was referenced to a subdural electrode located on the inferior surface of the temporal lobe. The absence of relevant activity emphasizes the focal nature of the responses encoding VOT and indicates that the previous AEPs primarily were generated by activity recorded from the lateral surface subdural electrode. Furthermore, AEPs recorded from the subdural site and evoked by click trains contained fundamentally different patterns than that seen in Heschl's gyrus (Fig. 9*D*). Whereas well-defined phase-locked responses that diminish with increasing frequency (low-pass) were recorded from He-
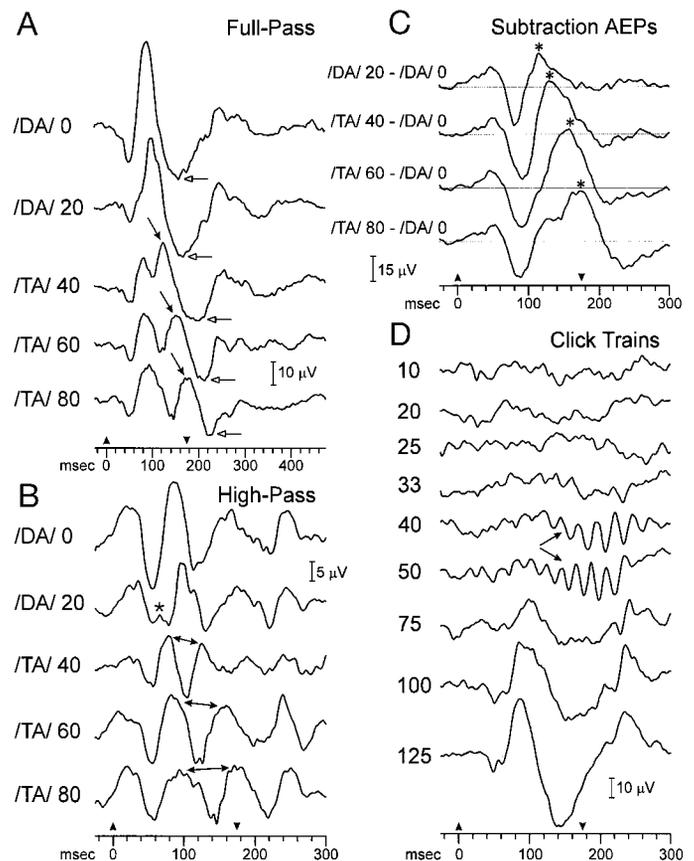
FIG. 9. *A*: *patient 3*, full-pass AEPs evoked by /da/ and /ta/ recorded at a subdural electrode located on the lateral convexity of the superior temporal gyrus, adjacent to the posterior margin of Heschl's gyrus. Marked changes in temporal representation of VOT are observed with positive-going components only present for the /ta/ stimuli (solid arrows). A progressive, and noncategorical, increase in the latency of the following negativity is also seen (unfilled arrows). *B*: high-pass filtered AEPs also demonstrate a pronounced change in amplitude of the response evoked by voicing onset with responses time-locked to consonant release and voicing onset only present for the /ta/ stimuli (double-headed arrows). Twenty-millisecond VOT stimulus only evokes a small positive deflection (asterisk) that precedes the following positivity. *C*: subtraction waveforms illustrate that although /ta/ stimuli evoke a prominent response to voicing onset (asterisks), a similar but smaller wave is elicited when comparing the AEPs evoked by the 2 short VOT syllables. *D*: AEPs evoked by click trains. In contrast to the low-pass characteristics of the click train-evoked responses in Heschl's gyrus (Fig. 8), those at this site are predominantly band-pass limited ~40–50 Hz. This suggests that the AEPs recorded at this site are generated by a nonprimary auditory cortical field and are not reflections of volume-conducted activity from Heschl's gyrus.

schl's gyrus (Fig. 8), very weak phase-locking was present to low-frequency click trains. In contrast, phase-locked responses at the lateral subdural site were maximal to click train rates of 40 and 50 Hz, demonstrating band-pass characteristics (*D*, arrows).

AEPs also were recorded from three high-impedance sites on an electrode placed in the planum temporale of *patient 2* (not shown). Waveforms evoked by the syllables consisted of a triphasic complex with peaks at 33, 61, and 107 ms. Latency differences between medial and lateral recording sites were not observed. Surprisingly, speech sounds failed to elicit response components time-locked to voicing onset. However, phase-locked responses were evoked at all three sites by click trains presented at rates of 10, 20, and 40 Hz, and were absent at a stimulus rate of 100 Hz. Similar to the findings recorded from Heschl's gyrus, the second phase-locked response evoked by the 40-Hz click train either was absent or markedly reduced in amplitude.

DISCUSSION

Relationship to perceptual investigations

Numerous psychoacoustical studies have established the importance of VOT as a cue for the perceptual discrimination of voiced from unvoiced English stop consonants (e.g., Borsky et al. 1998; Lisker 1975; Summerfield and Haggard 1977). This study demonstrates that the VOT temporal cue is represented at least partly within auditory cortex by synchronized responses of neuronal ensembles time-locked to both consonant release and voicing onset. These neural responses exhibit features that may facilitate categorical encoding of stop consonants. AEPs evoked by syllables with a short VOT contain a large response time-locked to consonant release followed by a variable, low-amplitude component time-locked to voicing onset. In contrast, AEPs evoked by CV syllables with a longer VOT contain prominent components time-locked to both stimulus and voicing onset. The VOT at which an AEP component to voicing onset becomes obvious is the same VOT at which the perceptual boundary occurs in most listeners. Present findings do not exclude additional phonetic encoding mechanisms or minimize the importance of speech parameters other than VOT (e.g., 1st formant onset frequency), for making the discrimination of voiced from unvoiced stop consonants (Liberman et al. 1958; McClaskey et al. 1983; Pegg and Werker 1997; Soli 1983; Summerfield and Haggard 1977; Treisman et al. 1995). Furthermore, present results need to be tempered by the fact that epileptic foci were located in nearby tissue and may have disrupted normal auditory cortical function and that AEPs were recorded from variable sites in the nondominant right hemisphere. Additionally, recording paradigms did not allow AEP patterns to be correlated tightly with perceptual responses. Despite these procedural limitations, results imply that auditory cortex is activated in synchronized fashion by temporal speech features and that these components need to be incorporated into neural models that characterize initial stages of language perception.

Temporal activity patterns representing VOT may contribute to categorical perception, wherein stimuli drawn from a physical continuum naturally segregate into discrete categories, and there is an increase in perceptual discrimination between stim-

uli located on either side of a perceptual boundary (Studdert-Kennedy et al. 1970). In this scheme, rapid segregation into categories of voiced and unvoiced stop consonants could be accomplished by determining whether the syllable evokes one or two discrete response bursts to consonant release and voicing onset. Heightened discrimination around the perceptual boundary would be based on comparisons between syllables that evoke one versus two response bursts. In contrast, discrimination between two consonants located on the same side of the perceptual boundary would require the more difficult task of comparing responses with a similar temporal response pattern. Human listeners are capable, under certain experimental conditions, to discriminate stop consonants located on the same side of the VOT perceptual boundary (Carney et al. 1977; Kewley-Port et al. 1988; Tremblay et al. 1997). Responses reported here mirror these findings, as subtle differences between the AEPs evoked by the two stimuli with short VOTs were observed, and the /ta/ stimuli could conceivably be discriminated by identifying differences in timing of the response burst to voicing onset relative to that evoked by consonant release.

AEP profiles demarcating English language voiced from unvoiced stop consonants are likely indexing a natural psychoacoustic boundary reflecting constraints imposed on temporal processing in the auditory cortex. For instance, many languages use a different voicing contrast from that used in English, with a distinction made between prevoiced stop consonants (vocal cord vibrations precede consonant release) and those with a short VOT. Studies in both adults and young infants whose home language does not use the short versus long VOT distinction demonstrate a natural tendency for enhanced discriminability of consonants that straddle the English boundary of 20–40 ms (Keating et al. 1981; Lasky et al. 1975; Streeter 1976). These findings suggest that the VOT phonetic boundary is not limited to learned language-specific categories but is based on the more general ability to identify the temporal sequence of two acoustical events, such as consonant release and voicing onset (Pisoni 1977). This conclusion is supported by categorical-like VOT discrimination in animals, which clearly do not possess language-specific capacities (Kuhl and Miller 1978; Kuhl and Padden 1982; Sinnott and Adams 1987). In his classic study, Hirsh (1959) found that a 15 to 20 ms separation in the onset of two sounds was required for their temporal sequence to be perceived. Subsequent studies found categorical perception with boundaries similar to those of speech for the temporal ordering of various other nonspeech stimuli (e.g., Formby et al. 1993; Miller et al. 1976; Phillips et al. 1997; Pisoni 1977; Stevens and Klatt 1974). Our findings suggest that these perceptual limitations are manifested in auditory cortex by a threshold for generating discrete responses to the onsets of consonant release and voicing onset. The markedly decreased amplitude of a response to the second click in trains with rates of 40–50 Hz reinforces this suggestion.

Relationship to physiological investigations

Timing and morphology of the intracranial AEPs observed in this study are consistent with previously reported values, suggesting that although recording sites are limited, they represent a valid sample of activity in human auditory cortex. Onset of activity evoked by clicks in Heschl's gyrus is 8–10

ms (Liégeois-Chauvel et al. 1991), compared with 11–12 ms for the syllables. The slightly prolonged latency may reflect the slower rise time and lower frequency content of the speech sounds relative to clicks. The increase in latency as recordings shift from more medial to lateral sites in Heschl's gyrus supports previous observations (Liégeois-Chauvel et al. 1994). Variability among the few intracranial studies reported makes exact comparisons with previous data difficult, but several reports describe the first large amplitude wave as peaking at ~30 ms (Celesia 1976; Liégeois-Chauvel et al. 1994). Wave A of the present study likely represents this component. The large amplitude wave B overlaps in time with the prominent positivities seen in more lateral sites of Heschl's gyrus, whereas wave C likely corresponds to the N120 wave also observed in lateral Heschl's gyrus (Liégeois-Chauvel et al. 1994). Polarity-inversion of waves A–C from previously reported AEP components is expected, as our electrodes were located on the underside of the current dipoles within auditory cortex. AEPs recorded from the lateral surface of the superior temporal gyrus have component latencies consistent with the N1, P2, and N3 click-evoked AEP components reported by Celesia (1976). Additionally, AEPs generated in the planum temporale have a triphasic morphology with polarity and peaks designated N30, P50, and N100 (Liégeois-Chauvel and al. 1994). These components are nearly identical in latency to the waves seen from the electrode in the planum temporale. Finally, it is evident from the body of intracranial data that single dipole models of scalp-recorded AEPs and magnetic responses are extreme simplifications of a complex, overlapping sequence of activation extending from koniocortex to secondary auditory areas. Similar conclusions have been stressed by other investigators (Lütkenhöner and Steinsträter 1998; Schreiner 1998), highlighting the concern that interpretations of results obtained via noninvasive physiological techniques should be viewed with caution and be confirmed by direct intracranial recordings.

Given these concerns, it is worthwhile to examine other reports of VOT encoding by responses time-locked to both consonant release and voicing onset (Kaukoranta et al. 1987; Kuriki et al. 1995). Both studies concluded that the magnetic responses equivalent to the N100 AEP component (N100m) and evoked by the two speech components were generated by separate sources in auditory cortex. The same conclusion was reached from analysis of magnetic responses evoked by a noise burst-square wave complex and two-tone stimuli, both analogous to syllables varying in their VOT (Mäkelä et al. 1988; Simos et al. 1998a). The latter two studies are especially relevant because they also report categorical-like changes in the amplitude of the N100m with a temporal boundary of 20–30 ms (see also Simos et al. 1998b,c). These conclusions are not wholly supported by the present data. Locally generated AEPs contain comparable response features reflecting VOT that are due to activity from single sources within auditory cortex. Present findings indicate that the nonlinear decrease in component amplitude is based on attenuation of the slower waves by new responses time-locked to voicing onset when VOT is more prolonged than 20 ms. In similar manner, categorical-like magnetic responses have been obtained with double-click stimulation, where the sources of activation for the two components should be identical (Celesia 1976; Joliot et al. 1994). Nonlinear changes in this pattern of physiological activity can be viewed as an example of forward masking. As

such, it requires interaction between the two response components, wherein activation evoked by the initial portion of the stimulus (e.g., consonant release) modifies the ability to generate a response to the second stimulus segment (e.g., voicing onset). The ability of stimuli with disparate frequency components to generate activity at single cortical locations is consonant with the widespread activation of auditory cortex by suprathreshold tones. (Bakin et al. 1996; Howard et al. 1996a; Phillips et al. 1994; Schreiner 1998). Thus, speech sounds presented at conversational levels should produce activation patterns in auditory cortex that contain regions where responses evoked by frequency-specific formants can interact to elicit forward masking phenomena.

In this study, activity recorded from three sites in the planum temporale of one patient did not reflect the VOT of the syllables. While this observation is consistent with previous findings that AEPs from the superior temporal plane were not highly dependent on the physical characteristics of acoustic stimuli (Halgren et al. 1995), multiple other reasons may be responsible for this negative result. Sites representing VOT with responses time-locked to consonant release and voicing onset could have been missed by the limited sampling of the tissue. AEPs were recorded in the right hemisphere, which, in this patient, was nondominant for speech. Temporal response patterns representing VOT may be restricted to locations in the homologous cytoarchitectonic area of the dominant, language hemisphere. Finally, other activity patterns that do not represent VOT with synchronized responses of neuronal ensembles time-locked to consonant release and voicing onset may be the relevant physiological encoding mechanism at these nonprimary sites. Considerations such as these indicate that it is premature to place too much emphasis on this negative result for generating schemes related to speech sound processing.

In contrast, the positive finding that AEPs recorded from the convexity of the posterior superior temporal gyrus maintain temporal representation of VOT supports the evolving concept that lateral belt areas of auditory cortex participate in the pattern recognition of sound, including speech (Rauschecker 1998; Rauschecker et al. 1995). In this model of cortical sound encoding, auditory cortex is organized into hierarchical streams of processing that occur within distinct pathways. Spatial features of sounds activate a dorsal pathway leading to the parietal cortex, whereas spectral and temporal patterns of complex sounds are represented in parallel within a ventral pathway that includes regions of the lateral superior temporal gyrus. Functional neuroimaging studies demonstrating greater activation of the lateral superior temporal gyrus bilaterally with complex acoustic stimuli from that evoked by simple tones or noise bursts add additional support for the existence of a ventral pathway involved in the pattern recognition of sound (Strainer et al. 1997; Zatorre et al. 1992).

A critical facet of this study is its complementary relationship with work performed in experimental animals (for review, see Phillips 1998). Temporal representation of VOT in the human AEP lends relevance to similar findings in A1 of monkeys, cats, and guinea pigs (Eggermont 1995a; McGee et al. 1996; Schreiner 1998; Steinschneider et al. 1982, 1994, 1995b). Detailed investigation in A1 of the anesthetized cat indicate that temporal representation of VOT is dependent on stimulus intensity and is likely a graded, nonlinear, and saturating function of the interval between consonant release and

voicing onset (Eggermont 1995b, 1999). In a more general framework, present results reinforce the hypothesis that synchronized activity within neuronal ensembles is a viable mechanism for encoding specific features of complex acoustic stimuli (deCharms and Merzenich 1996; Eggermont 1994; Wang et al. 1995). Although human AEPs offer relevance to the animal work, experimental models allow details of physiological processing to be identified at a level that are otherwise unobtainable. Studies in monkeys demonstrate a response transformation between thalamocortical fibers and A1 that accentuates the acoustic transients of consonant release and voicing onset and have suggested the synaptic events that generate the speech-evoked human AEP in A1 (Steinschneider et al. 1994). Detailed spatiotemporal patterns of A1 activation induced by speech sounds have been sequenced (Schreiner 1998). Some forms of developmental language disorders may be based on temporal processing deficits (e.g., Mannis et al. 1997; Merzenich et al. 1996; Tallal et al. 1996; Wright et al. 1997), which in turn may be associated with dysfunction or dysgenesis of the medial geniculate and auditory cortex (Galaburda et al. 1985, 1994; Humphreys et al. 1990; Nagarajan et al. 1999). Experimentally induced cortical dysgenesis produces temporal processing deficits specific for rapidly presented sounds that mirror abnormalities present in language-impaired children (Fitch et al. 1994; Herman et al. 1997). These exciting findings emphasize that animal models may not only assist in defining normal processes associated with carefully selected features of speech encoding, but may help clarify neural mechanisms underlying aberrant language development.

REFERENCES

BAKIN, J. S., KWON, M. C., MASINO, S. A., WEINBERGER, N. M., AND FROSTIG, R. D. Suprathreshold auditory cortex activation visualized by intrinsic signal optical imaging. *Cereb. Cortex* 6: 120–130, 1996.

BINDER, J. R., FROST, J. A., HAMMEKE, T. A., COX, R. W., RAO, S. M., AND PRIETO, T. Human brain language areas identified by functional magnetic resonance imaging. *J. Neuroscience* 17: 353–362, 1997.

BORSKY, S., TULLER, B., AND SHAPIRO, L. P. "How to milk a coat:" the effects of semantic and acoustic information on phoneme categorization. *J. Acoust. Soc. Am.* 103: 2670–2676, 1998.

CARNEY, A. E., WIDIN, G. P., AND VIEMEISTER, N. F. Noncategorical perception of stop consonants differing in VOT. *J. Acoust. Soc. Am.* 62: 961–970, 1977.

CELESIA, G. G. Organization of auditory cortical areas in man. *Brain* 99: 403–414, 1976.

DECHARMS, R. C. AND MERZENICH, M. M. Primary cortical representation of sounds by the coordination of action-potential timing. *Nature* 381: 610–613, 1996.

EGGERMONT, J. J. Neural interaction in cat primary auditory cortex. II. Effects of sound stimulation. *J. Neurophysiol.* 71: 246–270, 1994.

EGGERMONT, J. J. Neural correlates of gap detection and auditory fusion in cat auditory cortex. *Neuroreport* 6: 1645–1648, 1995a.

EGGERMONT, J. J. Representation of a voice onset time continuum in primary auditory cortex of the cat. *J. Acoust. Soc. Am.* 98: 911–920, 1995b.

EGGERMONT, J. J. Neural correlates of gap detection in three auditory cortical fields in the cat. *J. Neurophysiol.* 81: 2570–2581, 1999.

FITCH, R. H., TALLAL, P., BROWN, C. P., GALABURDA, A. M., AND ROSEN, G. D. Induced microgyria and auditory temporal processing in rats: a model for language impairment? *Cereb. Cortex* 4: 260–270, 1994.

FORMBY, C., BARKER, C., ABBEY, H., AND RANEY, J. J. Detection of silent temporal gaps between narrow-band noise markers having second-formantlike properties of voiceless stop/vowel combinations. *J. Acoust. Soc. Am.* 93: 1023–1027.

GALABURDA, A. M., MENARD, M. T., AND ROSEN, G. D. Evidence for aberrant auditory anatomy in developmental dyslexia. *Proc. Natl. Acad. Sci. USA* 91: 8010–8013, 1994.

GALABURDA, A. M., SHERMAN, G. F., ROSEN, G. D., ABOITIZ, F., AND GESCHWIND, N. Developmental dyslexia: four consecutive patients with cortical anomalies. *Ann. Neurol.* 18: 222–233, 1985.

HALGREN, E., BAUDENA, P., CLARKE, J. M., HEIT, G., LIÉGEOIS, C., CHAUVEL, P., AND MUSOLINO, A. Intracerebral potentials to rare target and distractor auditory and visual stimuli. I. Superior temporal plane and parietal lobe. *Electroencephalogr. Clin. Neurophysiol.* 94: 191–220, 1995.

HERMAN, A. E., GALABURDA, A. M., FITCH, R. H., CARTER, A. R., AND ROSEN, G. D. Cerebral microgyria, thalamic cell size and auditory temporal processing in male and female rats. *Cereb. Cortex.* 7: 453–464, 1997.

HIENZ, R. D., ALESZCZYK, C. M., AND MAY, B. J. Vowel discrimination in cats: thresholds for the detection of second formant changes in the vowel /ɛ/. *J. Acoust. Soc. Am.* 100: 1052–1058, 1996.

HIRSH, I. J. Auditory perception of temporal order. *J. Acoust. Soc. Am.* 31: 759–767, 1959.

HOWARD, M. A., III, VOLKOV, I. O., ABBAS, P. J., DAMASIO, H. M., OLLENDIECK, M. C., AND GRANNER, M. A. A chronic microelectrode investigation of the tonotopic oganization of human auditory cortex. *Brain Res.* 724: 260–264, 1996a.

HOWARD, M. A., III, VOLKOV, I. O., GRANNER, M. A., DAMASIO, H. M., OLLENDIECK, M. C., AND BAKKEN, H. E. A hybrid clinical-research depth electrode for acute and chronic in vivo microelectrode recording of human brain neurons. *J. Neurosurg.* 84: 129–132, 1996b.

HUMPHREYS, P., KAUFMANN, W. E., AND GALABURDA, A. M. Developmental dyslexia in women: neuropathological findings in three patients. *Ann. Neurol.* 28: 727–738, 1990.

JOLIOT, M., RIBARY, U., AND LLINÁS, R. Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. *Proc. Natl. Acad. Sci., USA* 91: 11748–11751, 1994.

KAUKORANTA, E., HARI, R., AND LOUNASMAA, O. V. Responses of the human auditory cortex to vowel onset after fricative consonants. *Exp. Brain Res.* 69: 19–23, 1987.

KEATING, P. A., MIKÓS, M. J., AND GANONG, W. F., III. A cross-language study of range of voice onset time in the perception of initial stop voicing. *J. Acoust. Soc. Am.* 70: 1261–1271, 1981.

KEWLEY-PORT, D., WATSON, C. S., AND FOYLE, D. C. Auditory temporal acuity in relation to category boundaries; speech and nonspeech stimuli. *J. Acoust. Soc. Am.* 83: 1133–1145, 1988.

KLUENDER, K. R. AND LOTTO, A. J. Effects of first formant onset frequency on [-voice] judgments result from auditory processes not specific to humans. *J. Acoust. Soc. Am.* 95: 1044–1052, 1994.

KUHL, P. K. AND MILLER, J. D. Speech perception by the chinchilla: identification functions synthetic VOT stimuli. *J. Acoust. Soc. Am.* 63: 905–917, 1978.

KUHL, P. K. AND PADDEN, D. M. Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Percept. Psychophys.* 32: 542–550, 1982.

KUHL, P. K. AND PADDEN, D. M. Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *J. Acoust. Soc. Am.* 73: 1003–1010, 1983.

KURIKI, S., OKITA, Y., AND HIRATA, Y. Source analysis of magnetic field responses from the human auditory cortex elicited by short speech sounds. *Exp. Brain Res.* 104: 144–152, 1995.

LASKY, R. E., SYRDAL-LASKY, A., AND KLEIN, R. E. VOT discrimination by four to six and a half month old infants from Spanish environments. *J. Exp. Child Psychol.* 20: 215–225, 1975.

LIBERMAN, A. M., DELATTRE, P. C., AND COOPER, F. S. Some cues for the distinction between voiced and voiceless stops in initial position. *Lang. Speech* 1: 153–167, 1958.

LIÉGEOIS-CHAUVEL, C., MUSOLINO, A., BADIER, J. M., MARQUIS, P., AND CHAUVEL, P. Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr. Clin. Neurophysiol.* 92: 204–214, 1994.

LIÉGEOIS-CHAUVEL, C., MUSOLINO, A., AND CHAUVEL, P. Localization of the primary auditory area in man. *Brain* 114: 139–153, 1991.

LISKER, L. Is it VOT or a first-formant transition detector? *J. Acoust. Soc. Am.* 57: 1547–1551, 1975.

LOTTO, A. J., KLUENDER, K. R., AND HOLT, L. L. Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *J. Acoust. Soc. Am.* 102: 1134–1140, 1997.

LÜTKENHÖNER, B. AND STEINSTRÄTER, O. High-precision neuromagnetic study of the functional organization of the human auditory cortex. *Audiol. Neurootol.* 3: 191–213, 1998.

MÄKELÄ, J. P., HARI, R., AND LEINONEN, L. Magnetic responses of the human auditory cortex to noise/square wave transitions. *Electroencephalogr. Clin. Neurophysiol.* 69: 423–430, 1988.

MANNIS, F. R., MCBRIDE-CHANG, C., SEIDENBERG, M. S., KEATING, P., DOI, L. M., MUNSON, B., AND PETERSEN, A. Are speech perception deficits associated with developmental dyslexia? *J. Exp. Child Psychol.* 66: 211–235, 1997.

MCCLASKEY, C. L., PISONI, D. B., AND CARRELL, T. D. Transfer of training of a new linguistic contrast in voicing. *Percept. Psychophys.* 34: 323–330, 1983.

MCGEE, T., KRAUS, N., KING, C., AND NICOL, T. Acoustic elements of speech-like stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. *J. Acoust. Soc. Am.* 99: 3606–3614, 1996.

MERZENICH, M. M., JENKINS, W. M., JOHNSON, P., SCHREINER, C., MILLER, S. L., AND TALLAL, P. Temporal processing deficits of language-learning impaired children ameliorated by training. *Science* 271: 77–81, 1996.

MILLER, J. D., WIER, C. C., PASTORE, R. E., KELLY, W. J., AND DOOLING, R. J. Discimination and labeling of noise-buzz sequences with varying noise-lead times: an example of categorical perception. *J. Acoust. Soc. Am.* 60: 410–417, 1976.

NAGARAJAN, S., MAHNCKE, H., SALZ, T., TALLAL, P., ROBERTS, T., AND MERZENICH, M. M. Cortical auditory signal processing in poor readers. *Proc. Natl. Acad. Sci. USA* 96: 6483–6488, 1999.

OHL, F. W. AND SCHEICH, H. Learning-induced dynamic receptive field changes in primary auditory cortex of the unanaesthetized Mongolian gerbil. *J. Comp. Neurol. [A]* 181: 685–696, 1997.

PEGG, J. E. AND WERKER, J. F. Adult and infant perception of two English phones. *J. Acoust. Soc. Am.* 102: 3742–3753, 1997.

PHILLIPS, D. P. Sensory representations, the auditory cortex, and speech perception. *Semin. Hear.* 19: 319–332, 1998.

PHILLIPS, D. P., SEMPLE, M. N., CALFORD, M. B., AND KITZES, L. M. Level-dependent representation of stimulus frequency in cat primary auditory cortex. *Exp. Brain Res.* 102: 210–226, 1994.

PHILLIPS, D. P., TAYLOR, T. L., HALL, S. E., CARR, M. M., AND MOSSOP, J. E. Detection of silent intervals between noises activating different perceptual channels: some properties of "central" auditory gap detection. *J. Acoust. Soc. Am.* 101: 3694–3705, 1997.

PISONI, D. P. 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, 1977.

POEPPEL, D., PHILLIPS, C., YELLIN, E., ROWLEY, H. A., ROBERTS, T.P.L., AND MARANTZ, A. Processing of vowels in supratemporal auditory cortex. *Neurosci. Lett.* 221: 145–148, 1997.

RAUSCHECKER, J. P. Parallel processing in the auditory cortex of primates. *Audiol. Neurootol.* 3: 86–103, 1998.

RAUSCHECKER, J. P., TIAN, B., AND HAUSER, M. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111–114, 1995.

RECANZONE, G. H., SCHREINER, C. E., AND MERZENICH, M. M. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* 13: 87–103, 1993.

SCHREINER, C. E. Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. *Audiol. Neurootol.* 3: 104–122, 1998.

SHARMA, A. AND DORMAN, M. F. Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. *J. Acoust. Soc. Am.* 104: 511–517, 1998.

SHARMA, A., KRAUS, N., MCGEE, T., CARRELL, T., AND NICOL, T. Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalogr. Clin. Neurophysiol.* 88: 64–71, 1993.

SIMOS, P. G., BREIER, J. I., ZOURIDAKIS, G., AND PAPANICOLAOU, A. C. MEG correlates of categorical-like temporal cue perception in humans. *Neuroreport* 9: 2475–2479, 1998a.

SIMOS, P. G., BREIER, J. I., ZOURIDAKIS, G., AND PAPANICOLAOU, A. C. Magnetic fields elicited by a tone onset time continuum in humans. *Cognit. Brain Res.* 6: 285–294, 1998b.

SIMOS, P. G., DIEHL, R. L., BREIER, J. I., MOLIS, M. R., ZOURIDAKIS, G., AND PAPANICOLAOU, A. C. MEG correlates of categorical perception of a voice onset time continuum in humans. *Cognit. Brain Res.* 7: 215–219, 1998c.

SINNOTT, J. M. AND ADAMS, F. S. Differences in human and monkey sensitivity to acoustic cues underlying voicing contrasts. *J. Acoust. Soc. Am.* 82: 1539–1547, 1987.

SINNOTT, J. M., BROWN, C. H., MALIK, W. T., AND KRESSLEY, R. A. A multidimensional scaling analysis of vowel discrimination in humans and monkeys. *Percept. Psychophys.* 59: 1214–1224, 1997.

SOLI, S. D. The role of spectral cues in discrimination of voice onset time differences. *J. Acoust. Soc. Am.* 73: 2150–2165, 1983.

SOMMERS, M. S., MOODY, D. B., PROSEN, C. A., AND STEBBINS, W. C. Formant frequency discrimination by Japanese macaques (*Macaca fuscata*). *J. Acoust. Soc. Am.* 91: 3499–3510, 1992.

STEINSCHNEIDER, M., AREZZO, J. C., AND VAUGHAN, H. G., JR. Speech evoked activity in the auditory radiations and cortex of the awake monkey. *Brain Res.* 252: 353–365, 1982.

STEINSCHNEIDER, M., SCHROEDER, C. E., AREZZO, J. C. AND VAUGHAN, H. G., JR. Speech-evoked activity in primary auditory cortex: effects of voice onset time. *Electroencephalogr. Clin. Neurophysiol.* 92: 30–43. 1994.

STEINSCHNEIDER, M., RESER, D. SCHROEDER, C. E., AND AREZZO, J. C. Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. *Brain Res.* 674: 147–152, 1995a.

STEINSCHNEIDER, M., SCHROEDER, C. E., AREZZO, J. C. AND VAUGHAN, H. G., JR. Physiological correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: temporal response patterns. *Brain Lang.* 48: 326–340, 1995b.

STEVENS, K. N. AND BLUMSTEIN, S. E. Invariant cues for place of articulation in stop consonants. *J. Acoust. Soc. Am.* 64: 1358–1368, 1978.

STEVENS, K. N. AND KLATT, D. H. Role of formant transitions in the voiced-voiceless distinction for stops. *J. Acoust. Soc. Am.* 55: 653–659, 1974.

STRAINER, J. C., ULMER, J. L., YETKIN, F. Z., HAUGHTON, V. M., DANIELS, D. L., AND MILLEN, S. J. Functional MR of the primary auditory cortex: An analysis of pure tone activation and tone discrimination. *Am. J. Neuroradiol.* 18: 601–610, 1997.

STREETER, L. A. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature* 259: 39–41, 1976.

STUDDERT-KENNEDY, M., LIBERMAN, A. M., HARRIS, K. S., AND COOPER, F. S. Motor theory of speech perception: a reply to Lane's critical review. *Psychol. Rev.* 77: 234–239, 1970.

SUMMERFIELD, Q. AND HAGGARD, M. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *J. Acoust. Soc. Am.* 62: 435–448, 1977.

TALLAL, P., MILLER, S. L., BEDI, G., BYMA, G., WANG, X., NAGARAJAN, S. S., SCHREINER, C., JENKINS, W. M., AND MERZENICH, M. M. Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science* 271: 81–84, 1996.

TREISMAN, M., FAULKNER, A., NAISH, P.L.N., AND ROSNER, B. S. Voice-onset time and tone-onset time: the role of criterion-setting mechanisms in categorical perception. *Q. J. Exp. Psychol.* 48A: 334–366, 1995.

TREMBLAY, K., KRAUS, N., CARRELL, T. D., AND MCGEE, T. Central auditory system plasticity: generalization to novel stimuli following listening training. *J. Acoust. Soc. Am.* 102: 3762–3773, 1997.

WANG, X., MERZENICH, M. M., BEITEL, R., AND SCHREINER, C. E. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J. Neurophysiol.* 74: 2685–2706, 1995.

WEINBERGER, N. M. Learning-induced receptive field plasticity in the primary auditory cortex. *Semin. Neurosci.* 9: 59–67, 1997.

WISE, R., CHOLLET, F., HADAR, U., FRISTON, K., HOFFNER, E., AND FRACKOWIAK, R. Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain* 114: 1803–1817, 1991.

WRIGHT, B. A., LOMBARDINO, L. J., KING, W. M., PURANIK, C. S., LEONARD, C. M., AND MERZENICH, M. M. Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature* 387: 176–178, 1997.

ZATORRE, R. J., EVANS, A. C., MEYER, E., AND GJEDDE, A. Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256: 846–849, 1992.

ZATORRE, R. J., MEYER, E., GJEDDE, A., AND EVANS, A. C. PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cereb. Cortex.* 6: 21–30, 1996.