Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe[a]

Therese McGee, Nina Kraus,[b] Cynthia King, and Trent Nicol
Department of Communication Sciences and Disorders, Northwestern University, 2299 North Campus Drive, Searle Building, Evanston, Illinois 60208

Thomas D. Carrell
Department of Communication Sciences, University of Nebraska, Lincoln, Nebraska 68583

(Received 11 July 1995; accepted for publication 6 February 1996)

Auditory evoked potentials measured from the guinea pig temporal lobe surface reflect acoustic elements of synthesized speech syllables. Eliciting stimuli included a four formant anchor stimulus /ba/, with a 40-ms formant transition duration. The other stimuli differed from /ba/ along simple acoustic dimensions. The /pa/ stimuli differed on a VOT continuum; /da/ stimuli had a higher frequency F_2 onset; /wa/ had a longer (80 ms) formant transition duration; and /bi/ differed in three vowel formant frequencies. The /ba/ and /da/ onset response latencies decreased systematically with increasing F_2 onset frequency. The response to the /pa/ voicing increased in latency with increasing VOT and showed a physiologic discontinuity at VOT of 15–20 ms. Responses to /ba/ and /wa/ showed similar onset morphology but significant amplitude differences at latencies corresponding to vowel onset. Significant amplitude differences in /ba/ and /bi/ responses corresponded in latency to both consonant and vowel portions of the syllables. Similar to previous reports in the awake monkey for VOT, these results demonstrate in the anesthetized guinea pig that acoustic elements essential to speech perception are reflected in aggregate response of ensembles of cortical neurons. © *1996 Acoustical Society of America.*

PACS numbers: 43.64.Sj, 43.71.Pc, 43.71.Es, 43.64.Ri

INTRODUCTION

It has been noted in numerous studies of human speech perception that certain acoustic elements of the speech signal tax the sensory/perceptual system more heavily than others. Although in quiet a normal listener may perceive and discriminate all of the acoustic elements of speech, given adverse listening conditions or impairment to the auditory system, a hierarchy of difficulty is apparent. Phillips and Farmer (1990) cite numerous studies of human primary auditory cortex (AC) lesions in which consonant discrimination was diminished or even destroyed while vowel perception was preserved. Studies of learning-disabled children and adults have described subjects with difficulty discriminating stop consonants, although vowel discrimination remained normal (Reed, 1989; Godfrey *et al.*, 1981; Werker and Teas, 1987; Tallal, 1981; Tallal *et al.*, 1985; Elliott and Hammer, 1988; Elliott *et al.*, 1989). In adverse listening environments, it is the perception of stop consonants that is most severely affected (Helfer, 1994), and hearing impaired listeners can selectively perform poorly on stop consonants, even when audibility is taken into account (Turner and Robb, 1987).

Rosen (1992) provides a framework for relating speech perception and its disruption to the underlying physiology. He suggests that the ease of perception of acoustic elements of speech is related to the neural encoding of temporal information. The implication is that different mechanisms in the auditory system encode a sound depending on its spectro-temporal structure. Furthermore, the locus of those mechanisms may vary. Studies of patients with cortical lesions and PET data from normal subjects have indicated that pitch and phonetic features are processed at different locations in the brain (Zatorre *et al.*, 1992; Auerbach *et al.*, 1982; Sharma *et al.*, 1994). Processing the very rapid spectro-temporal changes that characterize stop consonants seems to require cortical mechanisms. It has been postulated that it is an impairment of the temporal processing performed by cortical mechanisms that underlies deficits in the ability to discriminate rapid transitions (Tallal *et al.*, 1985; Tallal, 1981), and that a high degree of temporal precision is required to encode rapid formant transitions (Phillips *et al.*, 1989; Phillips and Hall, 1990).

A major goal of our group is to arrive at a better understanding of the neurophysiologic bases of normal speech perception and to improve basic understanding and formal assessment of various clinical populations where speech perception is impaired. To this end, we have measured mismatch evoked responses (MMN) elicited by synthesized syllables with phonemic differences to investigate speech discrimination in humans (Kraus *et al.*, 1993, 1994a, 1995). Mismatch responses are elicited by a change in a repetitive sequence of stimuli. Unlike the more familiar P300 response, MMN can be elicited by differences close to the discrimination threshold and is not dependent on attention or a task (Näätänen, 1992). A greater degree of difficulty in discrimi-

---

[a] Portions of this work were presented at the Midwinter Meeting of the Association for Research in Otolaryngology, 1995.

[b] Also with Department of Neurobiology and Physiology, Northwestern University.

nating stop consonants, compared to other speech contrasts, has been observed both behaviorally and electrophysiologically. Responses most vulnerable to disruption were those elicited by consonants with brief (40 ms) formant transitions that differed only in formant onset frequency. The specific syllables used in those studies were /da/ and /ga/, which differed in F_3 onset frequency (Kraus *et al.*, 1994a).

Using a guinea pig model of speech-elicited evoked responses, we have elicited mismatch responses from the thalamus in response to tones and certain speech stimuli (Kraus *et al.*, 1994c). The occurrence of a thalamic mismatch response indicates processing of the discrimination of those stimuli at a subcortical level. In contrast, a minimal mismatch response to /da/ versus /ga/ was apparent at the thalamic level, although a robust mismatch was observed at the epidural surface of the brain, consistent with the premise that discrimination of the /da/–/ga/ contrast requires cortical processing (Kraus *et al.*, 1994b).

Thus we are faced with an intriguing concept, that the acoustic elements that constitute speech show a distributed processing in the auditory pathway, that subcortical processing may be sufficient for relatively steady-state spectral changes such as pitch or vowels, and that a certain set of rapid spectro-temporal changes specifically require cortical processing.

In processing acoustic elements of speech, what is different about cortical processing compared to processing in lower levels of the auditory pathway? The responses of single primary auditory cortex neurons are characterized by a well-defined response to onsets of acoustic events, while steady-state aspects of stimuli (such as periodicity) are poorly represented in cortical neuronal responses (Creutzfeldt *et al.*, 1980; Rouiller *et al.*, 1981; Schreiner and Langner, 1988; Eggermont, 1991). Thus the cortex is said to have poor temporal resolution, although Phillips (1993) has contended that the temporal precision of cortical neuronal firing is excellent, sufficient to encode rapid formant transitions. That is, what is poorly represented in the primary auditory cortex is the temporal structure of steady-state periodicities, while timing of rapid spectro-temporal changes may be precisely represented. This line of reasoning explains how the cortex could be capable of coding stop consonants. It does not explain, however, why the cortex would be better suited than subcortical areas to encode rapid formant transitions.

Possibly a key to this dilemma is in pattern processing mechanisms that may be available only at higher levels. In a study of correlation and synchronization of cortical neurons in cat, Eggermont (1994) concluded that the response of a single cortical neuron is of minimal weight in the encoding of complex stimuli, and the encoding of complex patterns likely is accomplished by synchronous activity across large neural ensembles. This would suggest that the cortical mechanisms for identification of complex sounds which distinguish phonemes likely involve the responses of clusters of neurons rather than single-cell responses. It follows that investigation of the neural coding of phonemes then must of necessity involve measuring the synchronized response of a large number of neurons.

If this is the case, then cortical auditory evoked potentials can be of use in the study of encoding of complex stimuli. The evoked potential consists of an aggregate response from neural ensembles, and the recording technique capitalizes on neural synchrony. If individual neuron responses are extremely diverse or multiple pathways contribute, then the evoked potential, as an aggregate measure, may be a poor picture of the underlying neuronal activity. If, however, neuronal responses have a high degree of homogeneity and synchrony, the summated response should give considerable insight to the underlying activity. Interestingly, Nelken *et al.* (1994) have noted that the summed population response in primary auditory cortex accurately represents unit activity and is statistically more reliable than the unit responses.

In guinea pig, the evoked potential obtained directly from the epidural surface of the temporal lobe appears to be a summation of local cortical activity. Clicks and tone bursts elicit a large triphasic response (up to 2 mV peak-to-peak) with the first peak labeled A, the next trough labeled B, and the second peak, C (Hellweg *et al.*, 1977; McGee *et al.*, 1983). Response changes to rate variations, cortical lesions, neural inactivation, and binaural stimulation indicate that wave A originates in the primary auditory cortex and is maximal contralateral to the stimulus ear. Response topography and depth recordings in thalamus and cortex also indicate that wave A is generated in the primary auditory cortex (Kraus *et al.*, 1988; McGee *et al.*, 1991, 1992; Littman *et al.*, 1992; Kraus and McGee, 1995). Under xylazine and light ketamine anesthesia, a response can be obtained of similar morphology and amplitude to the awake response (McGee *et al.*, 1991).

Steinschneider *et al.* (1994, 1995a) have shown in awake monkeys that the temporal lobe evoked response reflects the cortical encoding of certain acoustic elements of speech stimuli, specifically, variations in voice onset time. They proposed that the cortical evoked response may even reflect demarcations associated with categorical perception. The evoked response morphology corresponded to salient features of simultaneously obtained multiunit activity and current source density recordings. Steinschneider *et al.*'s results indicate that evoked potentials could provide a convenient way to observe or monitor cortical responses. Furthermore, the results link the primary auditory cortex response to a considerable volume of behavioral perceptual data.

In the current study, cortical evoked responses were recorded from guinea pig in response to a synthesized speech syllable /ba/, and four other syllables /pa/, /da/, /wa/, and /bi/, which each differed from /ba/ in a particular acoustic element. Important to this investigation is the pairing of carefully controlled contrasts, so that it can be determined how response differences corresponded to acoustic differences which distinguish phonemes in the eliciting syllables. Investigated here are contrasts in steady-state formants, onset time of voicing, duration of formant transition, and spectral differences in rapid formant transitions. At issue is whether responses from the epidural surface will reflect stimulus differences, and whether those differences support a distribution of processing mechanisms in the auditory pathway.

I. METHODS

A. Subjects

Thirteen albino guinea pigs of either sex, weighing approximately 350 g, were used as subjects. Animals were anesthetized with ketamine hydrochloride (100 mg/kg) and xylazine (7 mg/kg), and maintained at a body temperature of 37.5 °C ($\pm 1°$). Smaller doses (15 mg/kg ketamine; 3 mg/kg xylazine) were administered as needed for the rest of the experiment. Throughout the experiment, the more anesthetic-sensitive surface responses (wave C, a positivity at 60 ms, recorded over the temporal lobe and $M+$, a positivity at 20–25 ms, recorded at the midline) were closely monitored to ensure a stable anesthetic level.

B. Electrophysiologic recording

Epidural sliver bead electrodes (0.5 mm diam) were positioned on the epidural surface at 20 locations: 8 over the midline and 6 over each temporal lobe. The electrodes over the temporal lobe were arranged bilaterally as rectangular grids centered 3 mm caudal and 9 mm lateral to bregma, with electrodes spaced 2 mm apart. Two of the midline electrodes were positioned close to bregma: one 2 mm rostral and to the right, the other 2 mm caudal and to the left of bregma. The remaining six midline electrodes were placed in two rows centered on midline and 9 mm caudal to bregma. An electrode placed 15 mm rostral to bregma and 1 mm lateral to the sagittal suture served as a reference. This distribution allowed assessment of scalp topography with good resolution over the temporal lobes and the posterior midline. Previous mapping of guinea pig cortex has shown these to be optimal sites for recording primary auditory cortex responses over the temporal lobe and for recording subcortical responses over the midline (Kraus *et al.*, 1985, 1988).

Recordings were filtered from 10 to 1500 Hz, 12 dB/oct roll-off. Each averaged response consisted of 200 individual responses with a 512-ms time sweep, including a 100-ms pre-stimulus period, collected at a sampling rate of 500 pts/s. Responses to the /ba/–/pa/ and /ba/–/da/ continua were collected to gain a more detailed view of the initial peak of the temporal lobe response. To obtain greater resolution, these responses were collected with a 256-ms time sweep and a sampling rate of 2000 pts/s.

C. The stimuli

On a Klatt synthesizer (cascade mode), five syllables were constructed: /ba/, /pa/, /da/, /wa/, and /bi/. Stimuli were presented to the right ear at 83 dB $\text{SPL}_{\text{max peak}}$ using Etymotic insert earphones (ER3) through hollow earbars. All stimuli had an effective duration of 140 ms, and were composed of four formants. Stimuli were presented at a repetition rate of 1.7/s.

Figure 1 describes stimulus formant frequencies and voicing time course. For /ba/, the formant transition of the initial consonant occurred from 5 to 40 ms. F_1 frequency changed linearly from 353 to 928 Hz; F_2 from 1135 to 1526 Hz. F_3 and F_4 were constant at 2975 and 4240 Hz, respectively. The fundamental frequency (F_0) ramped linearly from 300 to 280 Hz over 150 ms. A relatively high fundamental frequency was chosen. The glottal pulse will elicit responses that appear as an oscillatory pattern across the evoked potential. If F_0 is relatively high, the oscillation is not only diminished but can be filtered without affecting cortical responses. So that the key formants occurred at frequencies to which the guinea pig has excellent hearing, F_1 and F_2 were shifted upward in frequency. The shifts of F_0, F_1, and F_2 were set to maintain natural-sounding speech; thus a child's voice was synthesized. That the syllables be natural-sounding to humans likely was not necessary, but represents a bias by our speech perception colleague (TC). It was considered essential, however, that phonemic contrasts involve frequencies in the better region of the guinea pig audiogram.

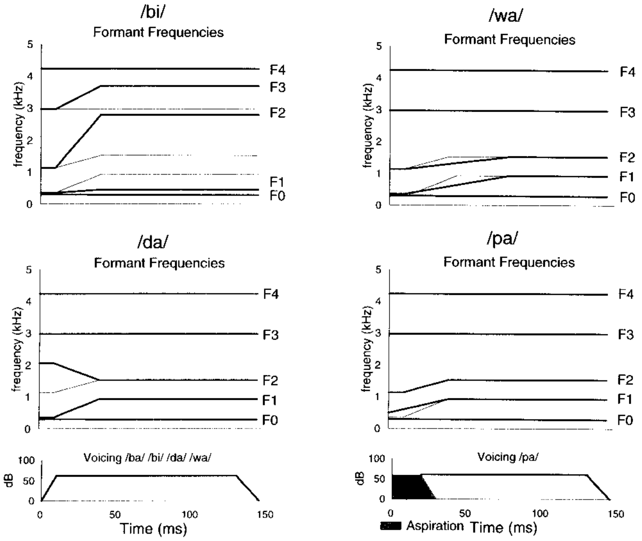

FIG. 1. Top, middle: Schematic diagrams of formant frequency for synthesized syllables /bi/, /wa/, /da/, and /pa/. Thin lines denote the characteristics of /ba/ where there is a phonemic contrast. Bottom: Time course of voicing was identical for /ba/, /bi/, /da/, and /wa/; /pa/ differed in that aspiration occurred and VOT was 20 ms.

Compared to /ba/, each of the other four syllables differed on simple dimensions. The syllable /pa/ differed in voice onset time. For /ba/, (as well as /da/, /wa/, and /bi/) voicing rose to maximum amplitude over the first 10 ms (VOT=10 ms). For /pa/, aspiration occurred from 0–20 ms, and voicing ramped to maximum energy from 10 to 20 ms after aspiration onset (VOT=20 ms). The syllables /da/, /wa/, and /bi/, differed from /ba/ in formant frequency and formant duration. For /da/, initial F_2 was 2050 Hz, 920 Hz higher than /ba/ F_2 onset. The syllable /wa/ differed from /ba/ in the duration of the formant transition. For /wa/, the F_1 and F_2 transitions occurred from 5 to 80 ms. For /bi/, the initial formant frequencies were identical to /ba/. The vowel had different F_1, F_2, and F_3 frequencies (456, 2800, 3700 Hz, respectively), resulting in different transition slopes for those formants, and an overall greater concentration of high-frequency energy.

Five additional stimuli were synthesized for experiments to further investigate response differences noted between /ba/ and /pa/ and between /ba/ and /da/. To construct a /ba/–/pa/ continuum, three new stimuli were synthesized with parameters identical to /pa/, except that the onset of voicing was varied in 5-ms increments. Thus a four stimulus continuum was available with an initial aspiration and VOT=15, 20, 25, 30 ms. To likewise obtain a four stimulus /ba/–/da/ continuum, two additional stimuli were constructed with parameters identical to /ba/, expect the F_2 onset frequency increased in 460-Hz increments. Across the continuum, the four stimuli had F_2 onsets of 1135, 1595, 2055, 2515 Hz.

D. Data analysis

For all 13 guinea pigs, responses were elicited with /ba/. Not all animals were tested with all of the other stimuli. For /da/, /pa/, and /bi/, $n=10$. For /wa/, $n=9$. Waveform and statistical comparisons of responses to each stimulus versus /ba/ were made utilizing the subject set for which both stimuli were recorded on the same animal. Amplitude and latencies of waves A, B, and C were measured for each animal. Using paired *t* tests, data for each of /da/, /pa/, /bi/, and /wa/ were compared to /ba/ responses for the corresponding animals.

In addition, using the individual animal waveforms as a data set, *t* tests were performed at each point of the waveform (point-to-point *t* test), and significance intervals were examined. This type of analysis has the advantage of avoiding subjective identification of peaks. The disadvantage of this analysis is the number of *t* tests performed and the increased chance of spurious significant values. Furthermore, because immediately adjacent points in the waveform are highly correlated, spurious significance may occur across short intervals. In response to this difficulty, Guthrie and Buchwald (1991) have suggested utilizing a criterion for a minimum interval of significance. Using multiple regression techniques on evoked potential waveforms, they concluded that if a continuous interval of a sufficient number of sampling points shows significance then the power of the statistical test is adequate. For the guinea pig data presented here, correlations among adjacent points were insignificant for points separated by 6 ms or more. Therefore, an interval was considered significant only if the duration of the significance interval was greater than 10 ms.

II. RESULTS

A. The response to /ba/

Over temporal lobe, the /ba/ stimulus elicited a triphasic waveform similar to the click-evoked response, so a similar nomenclature was used here, with the first positivity being labeled ''A,'' the following negativity ''B,'' and the next positivity ''C.'' Average wave A latency was 26 ms (s.d. =3.6 ms). This is approximately 14 ms longer than wave A to clicks (McGee *et al.*, 1983), and 11 ms longer than wave A recorded in response to a /ba/ stimulus with an immediate onset of voicing (Kraus *et al.*, 1994b). Maximum A–B amplitude of any animal was 460 μV, with a mean of 309 μV. Wave B occurred at a latency of 40 ms (s.d.=6.5 ms), and average wave C latency was 62 ms (s.d.=7.5 ms).

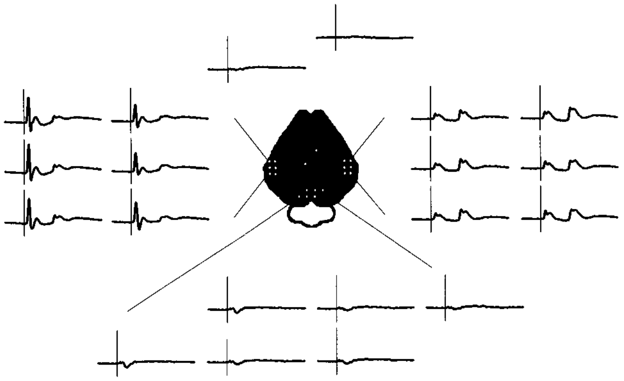

FIG. 2. Grand average waveforms for each electrode position. Shown is response to /ba/. Largest amplitude responses were obtained over the temporal lobe contralateral to the stimulus ear.

B. Topographic distribution of response differences

Similar to clicks, speech stimuli elicited the largest amplitude response over the temporal lobe contralateral to the stimulated ear. An offset positivity was also obtained over the contralateral temporal lobe. Ipsilaterally over the temporal lobe, positivities of 50–65 μV were observed that corresponded to the onset and offset of the stimulus. At the posterior midline, a small (8 μV *re*: prestimulus baseline) positivity was observed at approximately 16 ms poststimulus onset, followed by a negativity at 53 ms with an amplitude of -20 μV. The ipsilateral and midline responses were identical for /ba/, /da/, and /wa/. Response differences between /ba/ and these stimuli were localized to the contralateral temporal lobe. While the largest response difference between /ba/ and /pa/ was localized to the contralateral temporal lobe, small response differences were also observed at ipsilateral and midline sites. These differences appeared to correspond to differences in onset of energy (voicing and aspiration). The /ba/–/bi/ response difference showed a different topography from the other contrasts. For latencies of 10–85 ms, the /ba/–/bi/ response difference was maximal over the temporal lobe, but for latencies greater than 85 ms, the most significant differences were observed over the posterior midline.

The topographic distribution of /ba/ responses, grand averaged across animals, are shown in Fig. 2. Within the electrode array over the contralateral temporal lobe, the electrode placed 3 mm caudal and 10 mm lateral to bregma showed the largest amplitude responses and the greatest response differences for /ba/ versus /da/, /wa/, and /pa/. Detailed comparisons of responses therefore focused on recordings from this electrode. For /bi/, recordings from the electrode 8 mm posterior and 1 mm lateral to bregma (a midline location) were also compared to the /ba/ response in detail.

C. Response differences: /ba/ versus /pa/

As noted above, the /ba/ voice onset comprises a ramp to maximum energy over the first 10 ms. In contrast, the aspiration of /pa/ has an immediate onset, and ramp of voicing rises from 10 to 20 ms after stimulus onset.

For /pa/ (Fig. 3), the latency of peak A appears to correspond to the timing of the onsets of energy. Two peaks with the ''A'' morphology were elicited, and were termed

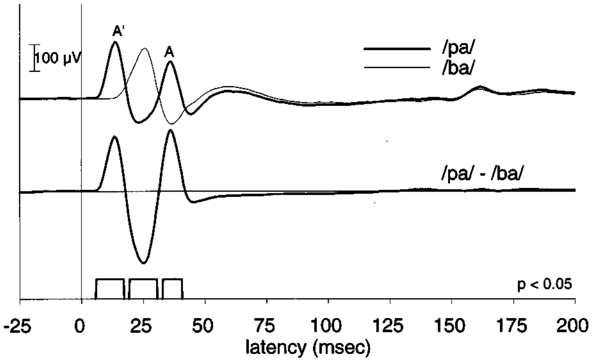

FIG. 3. Top: Grand average waveforms in response to /ba/ (thin line) and /pa/ (thick line); Middle: difference wave; Bottom: intervals where a significant difference occurred between the two waveforms ($n=10$).

$A'_{/pa/}$ and $A_{/pa/}$. The latencies of these peaks appeared to indicate that $A'_{/pa/}$ was elicited by the onset of aspiration, while $A_{/pa/}$ was elicited by the voice onset. The latency of $A'_{/pa/}$ (mean-13.4 ms) is similar to the latency of the click-evoked wave A, about 12 ms (McGee *et al.*, 1983), as would be expected of a response to an immediate onset of energy. $A_{/pa/}$ followed $A_{/ba/}$ in latency by 10 ms which corresponds to the VOT difference of 10 ms between the two syllables. The /pa/ aspiration elicited a very precise onset response. $A'_{/pa/}$ latency intersubject standard deviation was extremely low (0.84 ms), significantly smaller than $A_{/ba/}$ (s.d.=2.67 ms; $F=3.22$, $p<0.05$). No difference in variability was observed between $A_{/pa/}$ and $A_{/ba/}$, consistent with both being elicited by a similar onset ramp of voicing.

To further study whether $A_{/pa/}$ latency is voice onset dependent, in a second experiment VOT was varied, utilizing three additional stimuli, identical to /pa/ except voicing reached maximum amplitude at 15, 25, or 30 ms. Figure 4 shows the responses to the four stimuli. For VOT of 20, 25, and 30 ms, $A_{/pa/}$ latency increases monotonically with increasing VOT. Wave $A'_{/pa/}$ latency (response to aspiration onset) remained constant. For VOT of 15 ms, only $A'_{/pa/}$ was observed, indicating that the temporal spacing between aspiration and voicing was below the minimum needed to obtain a well-defined second wave A. Interestingly, to human listeners, the 15-ms VOT stimulus was subjectively perceived as /ba/, while the remaining stimuli of the continuum were perceived as /pa/.

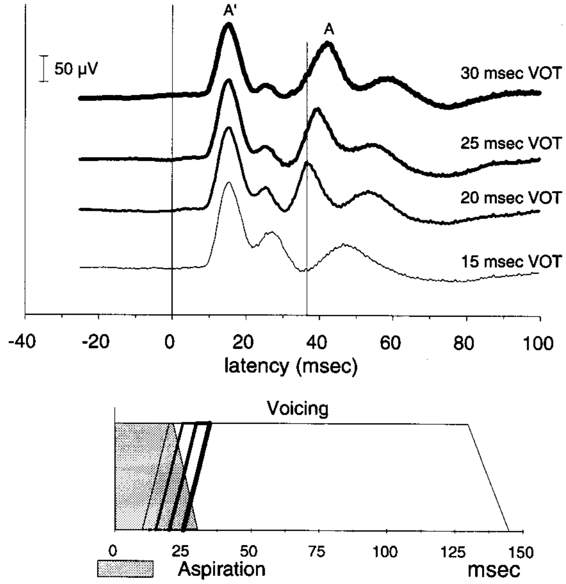

FIG. 4. Average responses ($n=2$) to four stimuli on a /ba/–/pa/ continuum. All stimuli had an initial burst of aspiration at $t=0$ ms. Voice onset varied in 5-ms steps. At VOT=15 ms, no $A_{/pa/}$ was observed. For VOT=20, 25, 30 ms, $A_{/pa/}$ latency showed a monotonic shift, while $A'_{/pa/}$ was unchanged.

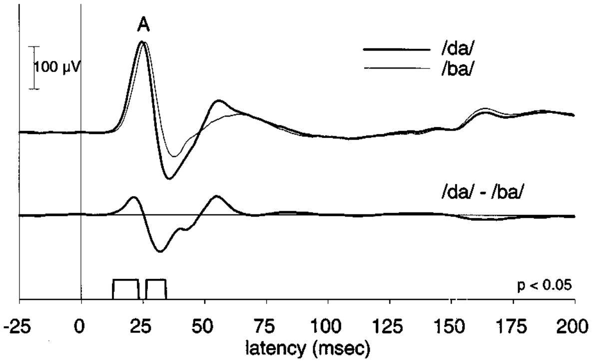

FIG. 5. Top: Grand average waveforms in response to /ba/ (thin line) and /da/ (thick line); Middle: difference wave; Bottom: intervals where a significant difference occurred between the two waveforms ($n=10$).

D. Response differences: /da/ versus /ba/

Only the spectrum of the initial 40-ms formant transition differentiated /da/ and /ba/. At the onset of voicing, F_2 for /da/ was 2055 Hz; F_2 onset for /ba/ was 1135 Hz. Over 40 ms, both ramped linearly to the vowel F_2 of 1526 Hz.

Grand average responses to /ba/ and /da/ are compared in Fig. 5. Wave $A_{/da/}$ latency was earlier than $A_{/ba/}$. Although the difference seemed small, it was consistent and highly significant across subjects ($p<0.01$), raising speculation that the latency of wave A reflected the F_2 onset frequency difference.

To further investigate whether a systematic effect could be demonstrated between F_2 onset frequency and wave A latency, an additional experiment was performed with two guinea pigs. Two additional stimuli were synthesized with onset F_2 of 1595 and 2515 Hz, but identical to /ba/ in all other parameters. With the original /ba/ and /da/, a four stimulus continuum was then available. Wave A latency showed a monotonic decrease with increasing F_2 onset frequency. Figure 6 summarizes these results. The monotonicity of the effect is particularly apparent in cross-correlation functions of the /ba/ response versus the responses to each of the other stimuli.

E. Response differences: /wa/ versus /ba/

The F_1/F_2 transition of /wa/ was 40 ms longer than the /ba/ F_1/F_2 transition. The initial onset spectrum of /wa/ was identical to /ba/. The responses showed no onset differences. No significant differences were observed in wave A ampli-

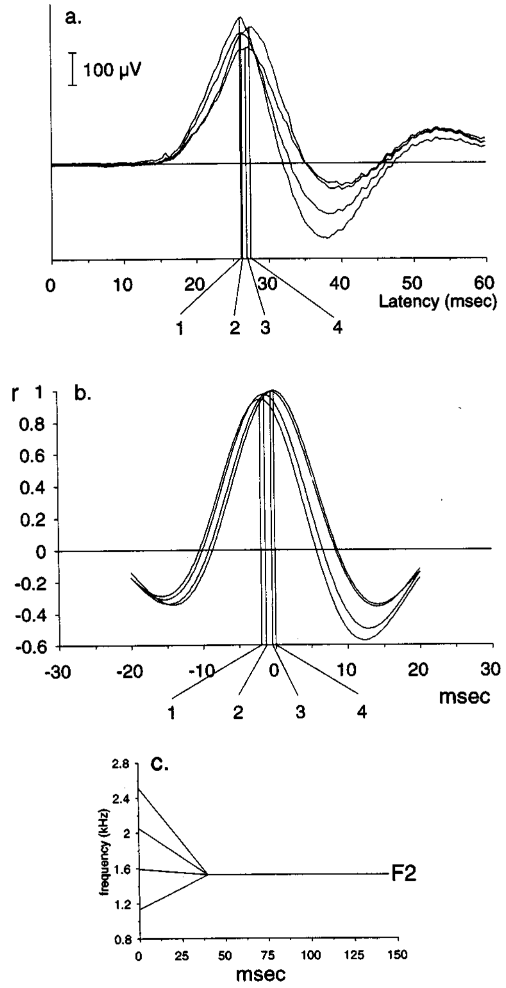

FIG. 6. (a) Initial 60-ms segment of grand average responses ($n=2$) to four stimuli on the /ba/–/da/ continuum; (b) Cross correlation of each response with the response to /ba/, where $F_2=1135$ Hz. (c) Schematic of second formant frequencies for the duration of the stimulus. Vertical lines in (a) and (b) indicate peak latency: F_2 onset=2515 Hz (1), 2055 Hz (2), 1595 Hz (3), 1135 Hz (4).

tude or latency. A point-to-point comparison of the grand averaged waveforms showed two intervals of significant differences occurring later in the waveform, centered at 59 and 101 ms (Fig. 7). That is, differences were not observed in the identifiable peaks and troughs. However, the spacing of the intervals of significance (≈40 ms) correspond to the timing difference, between /ba/ and /wa/, of the onsets of the steady-state portions of the syllables.

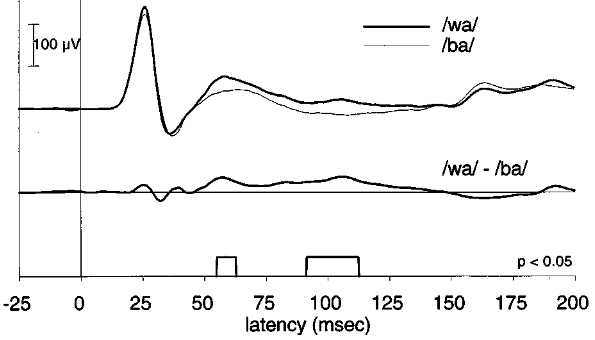

FIG. 7. Top: Grand average waveforms in response to /ba/ (thin line) and /wa/ (thick line); Middle: difference wave; Bottom: intervals where a significant difference occurred between the two waveforms ($n=9$).

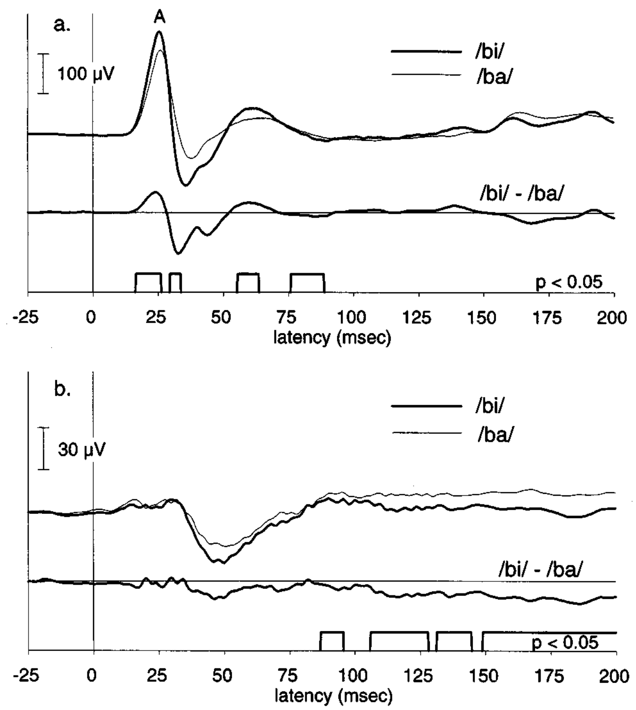

FIG. 8. Top: Grand average waveforms in response to /ba/ (thin line) and /bi/ (thick line); Middle: difference wave; Bottom: intervals where a significant difference occurred between the two waveforms ($n=10$). (a) from the temporal lobe surface; (b) from the midline.

F. Response differences: /bi/ versus /ba/

The syllables /ba/ and /bi/ were similar in their onset spectrum, but the slope of the F_1, F_2, and F_3 formant transitions differed radically and F_1-F_3 formant frequencies differed in the vowel portion. Thus except for a very brief interval at the onset, spectral differences occurred throughout the syllables.

From over the temporal lobe, although wave A latencies were not significantly different for /ba/ and /bi/, wave A amplitude was significantly larger for /bi/, and point-to-point comparison of the grand averaged waveforms showed intervals of significant difference from the onset of wave A to a latency of ≈90 ms. Over the midline, a long significance interval was observed from a latency of 85 to 280 ms, even extending beyond stimulus offset (Fig. 8). The maximal difference for latencies >85 ms was observed at the electrode 8 mm caudal and 1 mm lateral to the left of bregma (contralateral to the stimulus ear).

III. DISCUSSION

Because the responses of primary auditory cortex neurons are characterized by a well-defined response to onsets of acoustic events, it can be expected that the aggregate response, the evoked potential, will reflect acoustic onsets, and that the response to a well-defined onset will be the summation of precise, highly synchronized neuronal responses, thus producing a well-defined peak. With simple stimuli (clicks or tone bursts) the definition of an ''onset'' is clear. Dynamic continuing stimuli, such as speech, present a more complex situation. Within a speech sound, voicing, bursts, and transitions may each be an ''onset,'' an acoustically unique event eliciting a well-defined pattern of cortical response.

Results presented here indicate that acoustic elements of synthetic speech stimuli are reflected in surface-recorded potentials. Latency and amplitude differences in the peaks appear to reflect spectro-temporal differences in the stimulus onsets. For each stimulus comparison, the latency of response differences corresponded to the timing of the acoustic differences between stimulus pairs. Onset spectral or energy differences resulted in response differences in the initial portions of the waveform. Stimuli with similar initial spectra, but with later-occurring acoustic differences elicited responses that differed at later latencies.

Specifically, for /ba/ and /pa/, the initiation of aspiration and voicing each produced a well-defined peak from the electrode over primary auditory cortex. Across subjects, the response to aspiration, with its immediate onset, was extremely stable in latency. Systematic latency differences in the second peak were associated with variations in VOT. Similar double-peaked responses have been observed in response to /ta/ versus /da/ in awake monkeys (Steinschneider *et al.*, 1994, 1995a). In humans, Kaukoranta *et al.* (1987) demonstrated a double-peaked cortical magnetic response elicited by a synthesized ''hay,'' consisting of aspiration and a voice onset. Latency of the second peak of the magnetic response increased with increasing VOT.

Here, the second peak was elicited only for aspiration-VOT onset differences of 20 ms or more. An onset difference of 15 ms elicited only the initial response to aspiration. Similarly, EPs from awake monkey (Steinschneider *et al.*, 1994, 1995a) and single cell responses from cat (Eggermont, 1995) show an electrophysiologic discontinuity for VOT. These investigators suggest that the discontinuity may be a substrate for categorical perception of phonemic contrasts that differ in voice onset time.

Responses to /ba/ and /wa/ had similar onset morphologies but showed significant differences at latencies that corresponded to the onset of the steady-state vowel portion. Waveform differences in the responses to /ba/ and /bi/ corresponded in latency to both the consonant and vowel portion of the stimuli. Thus, the timing of differences in the cortical response to the stimuli corresponded to the timing of stimulus differences. The syllable /bi/ also elicited larger amplitude responses. Possibly the concentration of energy in higher frequencies for /bi/ affected response amplitude.

The most salient /ba/–/da/ stimulus difference is in the onset spectra and a small but significant latency difference was observed in the first peak (A) of the evoked cortical response. Wave A latency decreased systematically with increasing F_2 onset frequency in response to a four stimulus continuum. Ostensibly, there is no timing difference in /ba/ versus /da/, but a dynamic spectral difference. Yet a consistent timing difference is observed in the aggregate activity of the neural ensemble. By what specific mechanism this is accomplished is not clear. Given the size of the electrodes, small variations in electrode position across animals, and possible variations in brain organization, it seems unlikely that responses would be differentially affected by a bias to a particular characteristic frequency in the tonotopic map. The frequency difference in /ba/ versus /da/ is less than 1 kHz. Hellweg *et al.* (1977) and Redies *et al.* (1989) indicate CFs differing by 3–6 kHz within 0.5 mm across auditory cortex. It also seems unlikely that the cortical timing differences are a reflection of cochlear travel time. The spectral differences are in $F2$, but $F3$ and $F4$ carry energy at higher frequencies. Thus the shortest cochlear travel times should be to formant energy that is unchanged from /da/ to /ba/. Differences in cortical population should be the basis of any /ba/–/da/ response difference. Steinschneider *et al.* (1995b) have shown in monkeys that /da/ and /ba/ elicit responses from different neural populations. Interestingly, multiunit activity at high frequencies showed an earlier latency to /da/ in that study, but no statistical tests of significance were performed.

The results for /ba/–/pa/ and /ba/–/da/ reflect a high degree of neural precision and synchrony from an ensemble of cortical neurons. Eggermont (1994) documented the synchronization properties of 1290 pairs of cortical neurons in response to clicks and amplitude modulated noise bursts. His data demonstrate that cortical responses are precise and highly synchronized across the neuronal ensemble. That a neural ensemble can improve acuity compared to a single neuron has been demonstrated in electric fish and barn owl (Carr *et al.*, 1986; Kawasaki *et al.*, 1988; Carr, 1993). In cat ventral cochlear nucleus, neurons show a reduction of jitter compared to eighth nerve neurons, likely because of a convergence of many inputs (Rhode and Smith, 1986). At higher levels of the system, temporal precision can be sharpened and jitter reduced by collective action across a neural ensemble.

Considering the poor response of cortical neurons to repetitive stimuli presented at frequencies faster than 20 Hz (Creutzfeldt *et al.*, 1980; Rouiller *et al.*, 1981; Schreiner and Langner, 1988; Eggermont, 1991), one might expect that onsets spaced closely in time would not elicit well-synchronized summated responses. The timing difference between aspiration and voice onset, 20 ms (50 Hz), should preclude a strong response to the /pa/ voice onset. The surprisingly rapid neural recovery time indicated by the evoked response to /pa/ is consistent with responses to VOT recorded from a population of cortical single neurons (Eggermont, 1995). Possibly, this is also an effect of encoding by the neural ensemble, in that spectral differences between aspiration and voicing allow a more rapid second firing across the population than would be expected based on repetitive click or tone burst experiments. Alternatively, the higher stimulus level may allow better locking on temporal fluctuations (Creutzfeldt *et al.*, 1980) or some subset of neurons such as the ''thin-spike lockers'' described by de Ribaupierre *et al.* (1972) are able to respond at higher repetition rates.

Also evident in the /pa/ response were differences in the standard deviations of the latencies, with the response to aspiration showing a low standard deviation. This appears to support Phillips' (1993) focus on the precision of firing. That is, a well-defined onset elicits a response that is well defined in time. Similarly, Mainen and Sejnowski (1995) have noted increased reliability of cortical spike timing with well-defined onsets.

Topographic comparisons showed response differences consistent with a distribution of processing at different locations on the auditory pathway depending on the stimulus contrast. Responses to /ba/ and /da/ were identical at all electrodes except at the position directly over primary auditory cortex. Responses to /ba/ and /wa/ were different over a more widespread area of contralateral auditory cortex, but were identical at midline and cortical sites ipsilateral to the stimulus ear. /Ba/ and /pa/ differences were apparent at all locations, but were maximal over the contralateral temporal lobe. /Ba/ and /bi/ differences were apparent at all electrodes, but were most significant for the early portion of the waveform over the temporal lobe, and for the latter portion over the posterior midline area, possibly demonstrating a consonant–vowel segmentation in topography. In previous guinea pig studies, the midline surface region has been an advantageous location for recording subcortical activity (Kraus *et al.*, 1988; McGee *et al.*, 1991).

The topographic differences among the responses cannot definitively locate key areas for discrimination, but they do suggest that encoding of different acoustic elements involves different neural locations, and that in general, encoding of speech stimuli involves mechanisms at distributed locations along the auditory pathway.

IV. CONCLUSIONS

(1) Differences can be seen among cortical evoked responses that correspond to various acoustic differences between syllables that provide phonemic contrasts. Voice onset time is particularly salient, but response changes to spectral and temporal changes in formant transitions are also evident.

(2) The temporal structure of the evoked potential is extremely sensitive to the spectral and temporal aspects of the stimulus. The evoked response over the auditory cortex likely reflects neural encoding of elemental acoustic properties of the stimulus, even for complex signals such as speech.

(3) Results support the premise that the various acoustic elements that distinguish speech stimuli are processed at different locations along the auditory pathway.

ACKNOWLEDGMENT

Supported by NIH Grant No. R01 DC000264.

Auerbach, S. H., Allard, T., Naeser, M., Alexander, M., and Albert, M. L. (**1982**). ‘‘Pure word deafness: analysis of a case with bilateral lesions and defect at the prephonemic level,’’ Brain **105**, 271–300.

Carr, C. E. (**1993**). ‘‘Processing of temporal information in the brain,’’ Annu. Rev. Neurosci. **16**, 223–243.

Carr, C. E., Heiligenberg, W., and Rose, G. J. (**1986**). ‘‘A time-comparison circuit in the electric fish midbrain. I. Behavior and physiology,’’ J. Neurosci. **6**, 107–119.

Creutzfeldt, O., Hellweg, F.-C., and Schreiner, C. (**1980**). ‘‘Thalamocortical transformation of responses to complex auditory stimuli,’’ Exp. Brain Res. **39**, 87–104.

Eggermont, J. (**1991**). ‘‘Rate and synchronization measures of periodicity coding in cat primary auditory cortex,’’ Hear. Res. **56**, 153–167.

Eggermont, J. (**1994**). ‘‘Neural interaction in cat primary auditory cortex II. Effects of sound stimulation,’’ J. Neurophys. **71**, 246–270.

Eggermont, J. (**1995**). ‘‘Representation of a voice onset time continuum in primary auditory cortex of the cat,’’ J. Acoust. Soc. Am. **98**, 911–920.

Elliott, L., and Hammer, M. (**1988**). ‘‘Longitudinal changes in auditory discrimination in normal children and children with language-learning problems,’’ J. Speech Hear. Disord. **53**, 467–474.

Elliott, L., Hammer, M., and Scholl, M. (**1989**). ‘‘Fine-grained auditory discrimination in normal children and children with language-learning problems,’’ J. Speech Hear. Res. **32**, 112–119.

Godfrey, J. J., Syrdal-Lasky, A. K., Millay, K. K., and Knox, C. M. (**1981**). ‘‘Performance of dyslexic children on speech perception tests,’’ J. Exp. Child Psychol. **32**, 1981.

Guthrie, D., and Buchwald, J. S. (**1991**). ‘‘Significance testing of difference potentials,’’ Psychophysiology **28**, 240–244.

Helfer, K. (**1994**). ‘‘Binaural cues and consonant perception in reverberation and noise,’’ J. Speech Hear. Res. **37**, 429–438.

Hellweg, F. C., Koch, R., and Vollrath, M. (**1977**). ‘‘Representation of the cochlea in the neocortex of guinea pigs,’’ Exp. Brain Res. **29**, 467–474.

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.

Kawasaki, M., Rose, G., and Heligenberg, W. (**1988**). ‘‘Temporal hyperacuity in single neurons of electric fish,’’ Nature **336**, 173–174.

Kraus, N., and McGee, T. (**1995**). ‘‘The middle latency response generating system,’’ Electroencephalogr. Clin. Neurophysiol. Suppl. **44**, 93–101.

Kraus, N., Smith, D. I., and Grossman, J. (**1985**). ‘‘Cortical mapping of the auditory middle latency response in the unanesthetized guinea pig,’’ Electroencephalogr. Clin. Neurophysiol. **62**, 219–226.

Kraus, N., Smith, D. I., and McGee, T. (**1988**). ‘‘Midline and temporal lobe MLRs in the guinea pig originate from different generator systems: A conceptual framework for new and existing data,’’ Electroencephalogr. Clin. Neurophysiol. **70**, 541–558.

Kraus, N., McGee, T., Micco, A., Sharma, A., Carrell, T., and Nicol, T. (**1993**). ‘‘Mismatch negativity in school-age children to speech stimuli that are just perceptibly different,’’ Electroencephalogr. Clin. Neurophysiol. **88**, 123–130.

Kraus, N., McGee, T., Ferre, J., Hoeppner, J., Carrell, T., Sharma, A., and Nicol, T. (**1994a**). ‘‘Mismatch negativity in the neurophysiologic/behavioral evaluation of auditory processing deficits: A case study,’’ Ear Hear. **14**, 223–234.

Kraus, N., McGee, T., King, C., Littman, T., and Nicol. T. (**1994b**). ‘‘Discrimination of speechlike signals in auditory thalamus and cortex,’’ J. Acoust. Soc. Am. **96**, 2758–2768.

Kraus, N., McGee, T., Littman, T., Nicol, T., and King, C. (**1994c**). ‘‘Encoding of acoustic change involves non-primary auditory thalamus,’’ J. Neurophysiol. **72**, 1270–1277.

Kraus, N., McGee, T., CArrell, T., and Sharma, A. (**1995**). ‘‘Neurophysiologic bases of speech discrimination,’’ Ear Hear. **16**, 19–37.

Littman, T., Kraus, N., McGee, T., and Nicol, T. (**1992**). ‘‘Binaural stimulation reveals functional differences between midline and temporal components of the MLR in the guinea pig,’’ Electroencephalogr. Clin. Neurophysiol. **84**, 362–372.

Mainen, Z. F., and Sejnowski, T. J. (**1995**). ‘‘Reliability of spike timing in neocortical neurons,’’ Science **268**, 1503–1506.

McGee, T., Ozdamar, O., and Kraus, N. (**1983**). ‘‘Auditory middle latency responses (MLRs) in the guinea pig,’’ Am. J. Otolaryngol. **4**, 116–122.

McGee, T., Kraus, N., Comperatore, C., and Nicol, T. (**1991**). ‘‘Subcortical and cortical components of the MLR generating system,’’ Brain Res. **54**, 211–220.

McGee, T., Kraus, N., Littman, T., and Nicol. T. (**1992**). ‘‘Contributions of the subcomponents of the medial geniculate body on the MLR,’’ Hear. Res. **61**, 221–220.

Näätänen, R. (**1992**). *Attention and Brain Function* (Erlbaum, Hillsdale, NJ), pp. 136–210.

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.

Phillips, D. P. (**1993**). ‘‘Representation of acoustic events in the primary auditory cortex,’’ J. Exp. Psychol.: Human Percep. Perform. **19**, 203–216.

Phillips, D. P., and Farmer, M. E. (**1990**). ‘‘Acquired word deafness, and the temporal grain of sound representation in the primary auditory cortex,'' Behav. Brain Res. **40**, 85–94.

Phillips, D. P., and Hall, S. E. (**1990**). ''Response timing constraints in the cortical representation of sound time structure,'' J. Acoust. Soc. Am. **88**, 1403–1411.

Phillips, D. P., Hall, S. E., and Hollett, J. L. (**1989**). ''Repetition rate and signal level effects on neuronal responses for brief tone pulses in cat auditory cortex,'' J. Acoust. Soc. Am. **85**, 2537–2549.

Redies, H., Sieben, U., and Creutzfeldt, O. D. (**1989**). ''Functional subdivisions in the auditory cortex of the guinea pig,'' J. Comp. Neurol. **282**, 473–488.

Reed, M. A. (**1989**). ''Speech perception and the discrimination of brief auditory cues in reading disabled children,'' J. Exp. Child Psychol. **48**, 270–292.

Rhode, W. S., and Smith, P. H. (**1986**). ''Encoding timing and intensity in the ventral cochlear nucleus of the cat,'' J. Neuorphys. **56**, 261–286.

Ribaupierre, F. de, Goldstein, M., and Yeni-Komshian, G. (**1972**). ''Cortical coding of repetitive acoustic pulses,'' Brain Res. **46**, 205–225.

Rosen, S. (**1992**). ''Temporal information in speech: Acoustic, auditory and linguistic aspects,'' Philos. Trans. R. Soc. London **336**, 367–373.

Rouiller, E., Ribaupierre, Y. de, Toros-Morel, A., and Ribaupierre, F. de (**1981**). ''Neural coding of repetitive clicks in the medial geniculate body of cat,'' Hear. Res. **5**, 81–100.

Schreiner, C. E., and Langner, G. (**1988**). ''Coding of temporal patterns in the central auditory nervous system,'' in *Auditory Function*, edited by G. M. Edelman, W. E. Gail, W. M. Cowan (Wiley, New York), pp. 337–361.

Sharma, A., Kraus, N., Carrell, T., and Thompson, C. (**1994**). ''Physiologic bases of pitch and place of articulation perception: A case study,'' J. Acoust. Soc. Am. **95**, 3011.

Steinschneider, M., Arezzo, J. C., Schroeder, C., and Vaughan, H. G., Jr. (**1994**). ''Speech evoked activity in primary auditory cortex: effects of onset time,'' Electroencephalogr. Clin. Neurophysiol. **92**, 30–43.

Steinschneider, M., Arezzo, J. C., Schroeder, C., and Vaughan, H. G., Jr. (**1995a**). ''Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: Temporal response patterns,'' Brain Lang. **48**, 326–340.

Steinschneider, M., Reser, D., Schroeder, C., and Arezzo, J. C. (**1995b**). ''Tonotopic organization of responses reflecting a stop consonant place of articulation in primary auditory cortex (A1) of the monkey,'' Brain Res. **674**, 147–152.

Tallal, P. (**1981**). ''Language disabilities in children: perceptual correlates,'' Int. J. Pediatr. Otorhinolaryngol. **3**, 1–13.

Tallal, P., Stark, R., and Mellitis, F. (**1985**). ''The relationship between auditory temporal analysis and receptive language development: Evidence from studies of developmental language disorder,'' Neuropsychologia **23**, 314–322.

Turner, C. W., and Robb, M. P. (**1989**). ''Audibility and recognition of stop consonants in normal and hearing-impaired subjects,'' J. Acoust. Soc. Am. **81**, 1566–1573.

Werker, J. E., and Teas, R. C. (**1987**). ''Speech perception in severely disabled and average reading children,'' Can. J. Psychol. **41**, 48–61.

Zatorre, R. J., Evans, A. C., Meyer, E., and Gjedde, A. (**1992**). ''Lateralization of phonetic and pitch discrimination in speech processing,'' Science **256**, 846–849.