

Relationship between N1 evoked potential morphology and the perception of voicing

Anu Sharma,^{a)} Catherine M. Marsh, and Michael F. Dorman

Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

(Received 12 July 1999; accepted for publication 22 August 2000)

Auditory evoked potential (AEP) correlates of the neural representation of stimuli along a /ga/–/ka/ and a /ba/–/pa/ continuum were examined to determine whether the voice-onset time (VOT)-related change in the N1 onset response from a single to double-peaked component is a reliable indicator of the perception of voiced and voiceless sounds. Behavioral identification results from ten subjects revealed a mean category boundary at a VOT of 46 ms for the /ga/–/ka/ continuum and at a VOT of 27.5 ms for the /ba/–/pa/ continuum. In the same subjects, electrophysiologic recordings revealed that a single N1 component was seen for stimuli with VOTs of 30 ms and less, and two components (N1' and N1) were seen for stimuli with VOTs of 40 ms and more for both continua. That is, the change in N1 morphology (from single to double-peaked) coincided with the change in perception from voiced to voiceless for stimuli from the /ba/–/pa/ continuum, but not for stimuli from the /ga/–/ka/ continuum. The results of this study show that N1 morphology does not reliably predict phonetic identification of stimuli varying in VOT. These findings also suggest that the previously reported appearance of a “double-peak” onset response in aggregate recordings from the auditory cortex does not indicate a cortical correlate of the perception of voicelessness. © 2000 Acoustical Society of America. [S0001-4966(00)00512-9]

PACS numbers: 43.71.Pc, 43.71.An, 43.64.Qh [CWT]

I. INTRODUCTION

The aim of the research reported here was to examine the relationship between the morphology of the N1 component of the cortical auditory evoked potential (AEP) and the perception of the voicing contrast in syllable initial position. In a previous study (Sharma and Dorman, 1999), we recorded behavioral responses and AEPs in response to stimuli along a synthesized /da/–/ta/ continuum in which the voice-onset time (VOT) of the syllables varied from 0 to 80 ms. The results of the behavioral experiment revealed, as expected, a sharp category boundary between /da/ and /ta/ around the same VOT for all listeners. The results of the electrophysiologic experiment showed two distinct onset responses (N1' and N1) of the AEP that behaved differently in response to VOT. Based on the latencies of the two components, Sharma and Dorman (1999) concluded that the first component (N1') occurred in response to the burst at the beginning of each syllable and that the second component (N1) occurred in response to voicing onset. Of particular interest was their finding that for sounds with a short lag in voicing (perceived as /da/) only a single component (N1) was seen, while for sounds with longer VOTs (perceived as /ta/) two distinct components (N1' and N1) were observed in the AEP waveform. That is, in Sharma and Dorman's study the physiologic discontinuity in N1 morphology coincided with the change in perception from /da/ to /ta/ across the category boundary.

Sharma and Dorman's (1999) findings were consistent with some earlier human and animal work which had also examined VOT encoding in the auditory cortex. For ex-

ample, Steinscheider *et al.* (1995) and McGee *et al.* (1996), who examined evoked responses from the auditory cortex of awake moneys and guinea pigs, respectively, and Steinscheider *et al.* (1999), who examined intracortical evoked responses in humans, reported that voiced syllables and syllables with short VOTs elicited a “single-onset” response in the AEP waveform, and that a robust “double-onset” response was observed only for sounds with longer VOTs (i.e., sounds that are typically perceived as voiceless by human listeners).

In all these studies, the appearance of a clearly apparent double-peaked onset response in the AEP waveform coincided with a stimulus identified as voiceless. Therefore, it is possible that the change in N1 from a single to double-peaked component actually *predicts* a change in perception of phonetic categories. If this were the case, then the N1 might provide a clinically useful technique to measure VOT perception. Towards this end, we examined N1 correlates of neural representation of VOT in a /ba/–/pa/ and /ka/–/ga/ stimulus continuum to determine whether the appearance of VOT-related changes in the N1 waveform is a reliable indicator of perceptual identification of voiced and voiceless sounds. The /ba/–/pa/ and /ga/–/ka/ continua were chosen because the voicing boundaries differ for the two continua.

II. BEHAVIORAL EXPERIMENT

A. Method

1. Subjects

Five females and five males ranging in age from 20–30 years with a mean age of 25 years were paid \$5/h to partici-

^{a)}Electronic mail: anu.sharma@asu.edu

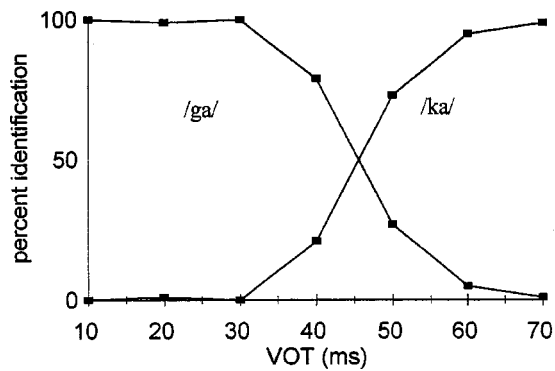


FIG. 1. Mean identification functions for the /ga/-/ka/ continuum. Stimuli with VOTs of 0–30 ms were identified as /ga/, while stimuli with VOTs of 60–70 ms were identified as /ka/. The category boundary between /ga/ and /ka/ was at a VOT of 46 ms.

pate as subjects. All subjects were monolingual native speakers of American English who reported no history of speech or hearing disorders.

2. Stimuli

Two sets of continua of CV speech sounds varying in VOT were generated using the Klatt (1980) speech synthesizer. One continuum consisted of seven tokens varying perceptually from /ga/ to /ka/ and the other continuum consisted of six tokens varying perceptually from /ba/ to /pa/. In each syllable, a brief burst of friction noise was present at syllable onset and from syllable onset to $F1$ onset the higher formants were excited using aspiration noise. For the /ga/-/ka/ continuum, the aspirated interval corresponded to VOT which was varied from 0–70 ms. The center frequency of $F1$ was 310 Hz for the first 10 ms and increased to its steady-state value of 521 Hz at 60 ms. $F2$ decreased from 1920 to 1568 Hz over a 50-ms transition duration and $F3$ remained at 2500 Hz throughout the syllable. For the /ba/-/pa/ continuum, the aspirated interval corresponded to VOT which was varied from 0–60 ms. The center frequency of $F1$ was 435 Hz for the first 10 ms of the syllable and increased to its steady-state value of 650 Hz at 40 ms. $F2$ increased from 1250 to 1490 Hz over a 40-ms transition duration and $F3$ increased from 2300 to 2470 Hz over a 50-ms transition

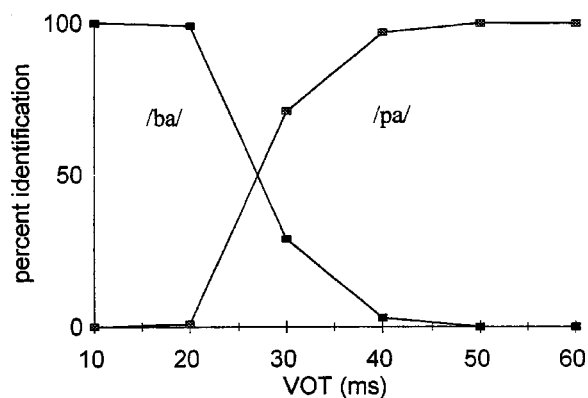


FIG. 2. Mean identification functions for the /ba/-/pa/ continuum. Stimuli with VOTs of 0–20 ms were identified as /ba/, while stimuli with VOTs of 40–60 ms were identified as /pa/. The category boundary between /ba/ and /pa/ lay at a VOT of 27.5.

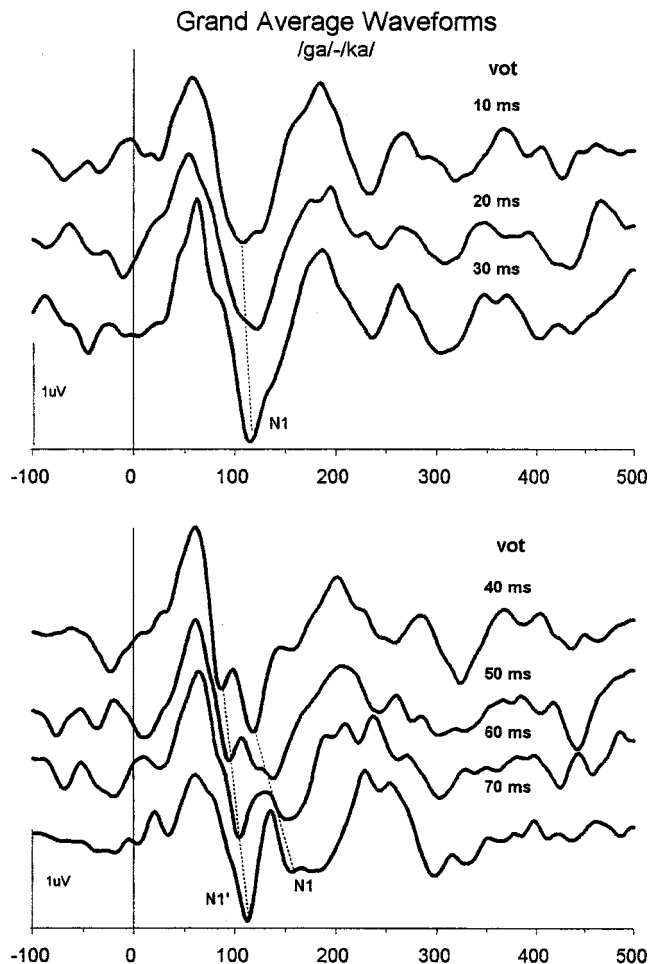


FIG. 3. Grand-averaged $N1$ responses elicited by stimuli in the /ga/-/ka/ continuum. Responses to stimuli with VOTs of 30 ms or less showed a single negativity ($N1$) indicated by the dashed line. Responses to stimuli with VOTs of 40 ms or more showed two distinct negative components ($N1'$ and $N1$) indicated by dashed lines.

duration. In both continua, the overall duration of each syllable was 200 ms. Stimuli were equated in intensity within 1 dB of each other and were presented to subjects at 75 dB SPL binaurally over headphones.

3. Procedure

Subjects performed a two-choice identification experiment with the stimuli from each stimulus continuum. That is, subjects were asked to listen to the syllables and classify them as either /ba/, /pa/ or /ga/, /ka/. Subjects were asked to indicate their responses by clicking with a computer mouse on panels marked BA, PA, GA, and KA appearing on the computer screen. Subjects were given an initial practice session where they heard every stimulus from each continuum once in order. After the practice session, ten repetitions of each stimulus in a continuum were presented to the subject in random order. The order of presentation of the /ga/-/ka/ and /ba/-/pa/ stimulus continua was randomized across subjects.

B. Results

Responses to each of the 13 tokens in the two continua were collapsed across subjects. The group means for each

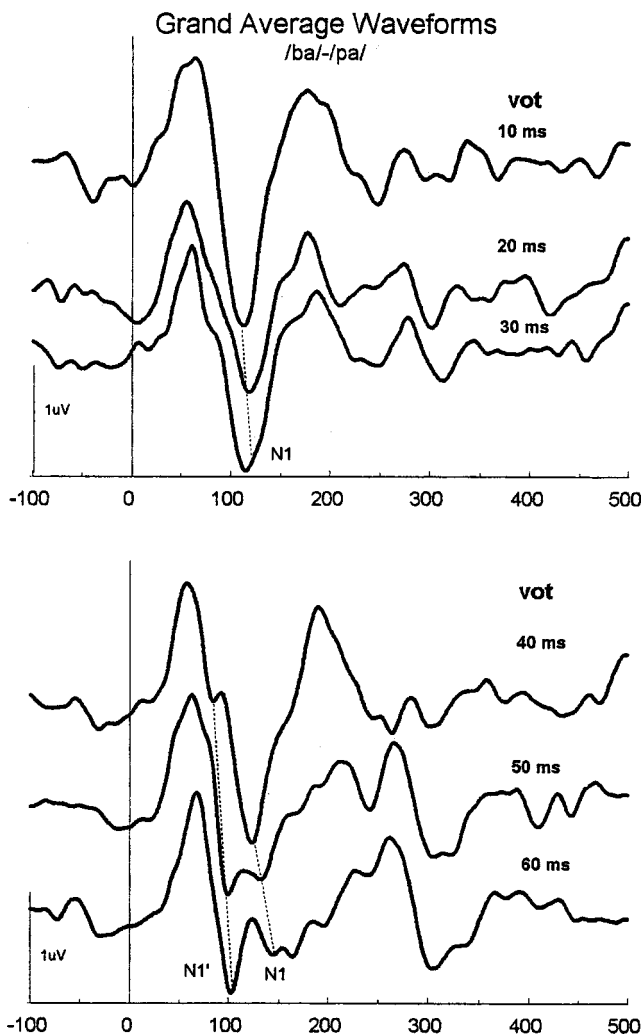


FIG. 4. Grand-averaged *N1* responses elicited by stimuli in the /ba-/pa/ continuum. Responses to stimuli with VOTs of 30 ms or less showed a single negativity (*N1*) indicated by the dashed line. Responses to stimuli with VOTs of 40 ms or more showed two distinct negative components (*N1'* and *N1*) indicated by dashed lines.

token are shown in Fig. 1 for /ga-/ka/ and Fig. 2 for /ba-/pa/. As can be seen in Fig. 1, for the /ga-/ka/ continuum, stimuli with VOT of 0–30 ms were consistently identified as /ga/ while stimuli with VOT of 60–70 ms were consistently identified as /ka/. The boundary (i.e., the 50% identification point) between the /ga-/ka/ categories occurred at a VOT of 46 ms. As can be seen in Fig. 2, for the /ba-/pa/ continuum, stimuli with VOT of 0–20 ms were consistently identified as /ba/ while stimuli with VOT of 40–60 ms were consistently identified as /pa/. The boundary (i.e., the 50% identification point) between the /ba-/pa/ categories occurred at a VOT of 27.5 ms.

III. ELECTROPHYSIOLOGIC EXPERIMENT

A. Methods

1. Subjects

Electrophysiologic recordings were obtained in the same ten subjects who performed in the behavioral experiment.

N1 /ga-/ka/

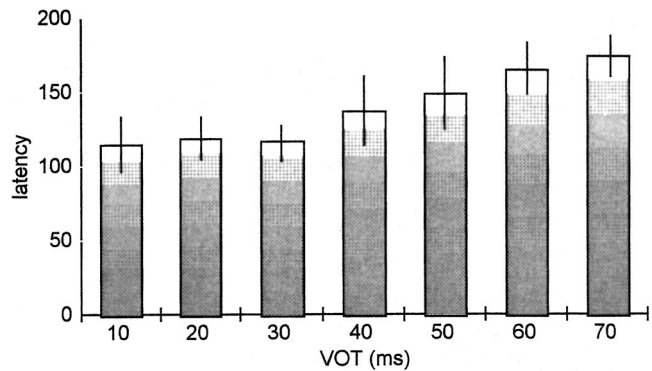


FIG. 5. Mean *N1* response latency values for the /ga-/ka/ continuum. Error bars indicate ± 1 standard deviation.

2. Stimuli

The /ga-/ka/ and /ba-/pa/ stimulus continua were identical to those used in the behavioral identification experiment. Repeated presentations of each stimulus separated by an onset-to-offset interstimulus interval (ISI) of 800 ms were used to elicit the *N1*. The order of presentation of stimuli was counterbalanced across subjects.

3. Recording procedures

During *N1* recordings, subjects were seated comfortably in a sound-treated booth. To control for arousal state and to minimize their attention to the test stimuli, subjects watched a videotaped movie of their choice. Subjects were asked to ignore the stimuli that were presented through an insert earphone at 75 dB SPL in the right ear. Videotape audio levels were kept below 40 dB SPL.

A Neuroscan Inc. data acquisition system was utilized to record the auditory evoked potentials. Silver-chloride electrodes were placed on the scalp at midline (Fz, Cz, and Pz) locations. A reference electrode was placed on the right mastoid and a ground electrode was positioned on the forehead. Eye movements were monitored with a bipolar electrode montage (supraorbital to lateral canthus). Averaging was suspended when the eye channel recorded blinks. The recording window included a 100-ms prestimulus period and 500-ms poststimulus time. Evoked responses were recorded in response to each stimulus; 250–300 sweeps analog filtered on-line from 0.1 to 100 Hz were collected. The digitization rate was 2 kHz. Total testing time for each subject was approximately 2 h and breaks were provided to subjects as necessary.

4. Data analysis

For individual subjects, sweeps were corrected to the average baseline, and sweeps that were greater than ± 100 microvolts were automatically rejected. Sweeps were then averaged to compute an individual averaged waveform. Waveforms were digitally high-pass filtered off-line at 4 Hz (filter slope 12 dB/octave).

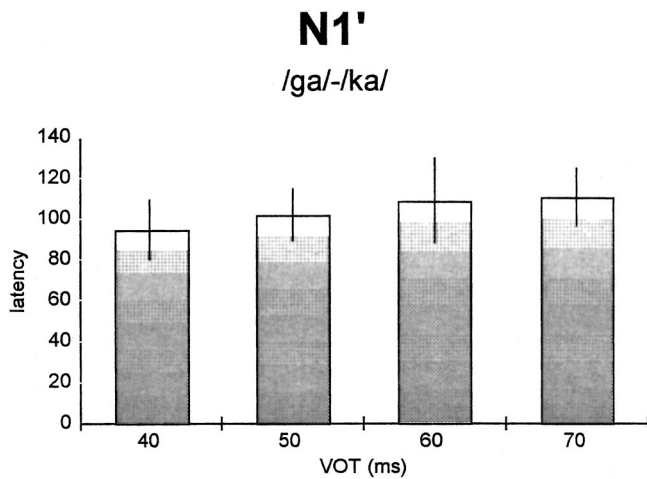


FIG. 6. Mean $N1'$ response latency values for the /ga/-/ka/ continuum. Error bars indicate ± 1 standard deviation.

Response windows were defined around the peak or peaks in the group mean waveforms to aid in peak identification and measurement in data from individual subjects. The time frame was defined as 0–200 ms when one $N1$ component was present. When it was determined that two $N1$ components were present in the grand average waveform, the time window was defined as 0–120 ms for the first ($N1'$) component and 120–200 ms for the second ($N1$) component. Peak latencies were detected based on the recordings from the Cz electrode site because the response amplitudes were the largest at this site in the group mean waveforms. $N1$ and $N1'$ amplitudes were measured relative to preceding positive peak. The author (CM) who marked the $N1$ and $N1'$ latencies and amplitudes was unaware of the eliciting stimulus continuum or stimulus number. Finally, group-averaged waveforms were computed by averaging across the individual average waveforms for the appropriate stimulus condition. Group-averaged waveforms were low-pass filtered at 40 Hz (12 dB/octave) to smooth the waves for the final figures.

B. Results

Grand average waveforms elicited in response to stimuli from the /ga/-/ka/ continuum and in response to stimuli from the /ba/-/pa/ continuum are shown in Fig. 3 and Fig. 4, respectively. Distinct morphological changes related to encoding of VOT are seen in the AEP waveforms. For both the /ga/-/ka/ and the /ba/-/pa/ continua, stimuli with short VOTs (0–30 ms) elicited a single negativity. However, in response to stimuli with long VOTs (40–70 ms) two negative components ($N1'$ and $N1$) were apparent in the AEP waveform.

1. Latency

Mean latency values for the $N1$ and $N1'$ components are shown in Fig. 5 and Fig. 6, respectively, for the /ga/-/ka/ continuum. For the /ga/-/ka/ continuum, a one-way repeated-measures analysis of variance (ANOVA), which revealed a significant main effect of VOT ($F=28$, $p < 0.000001$) on $N1$ latencies. Correlation analysis revealed a significant positive correlation between $N1$ latency and

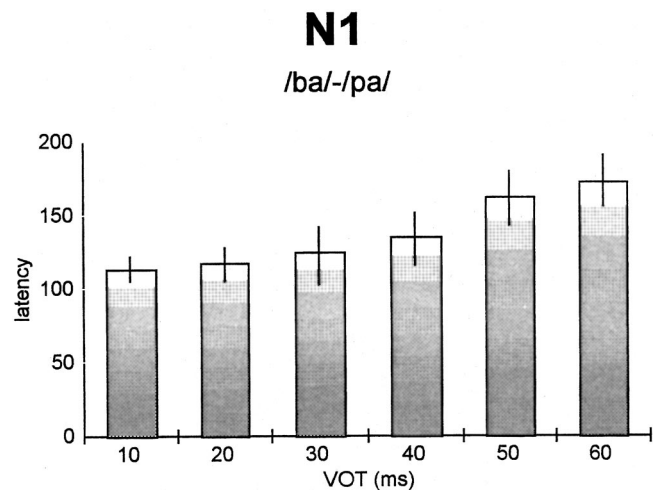


FIG. 7. Mean $N1$ response latency values for the /ba/-/pa/ continuum. Error bars indicate ± 1 standard deviation.

VOT ($r=0.77$, $p < 0.0001$). For the $N1'$ component (Fig. 6), a one-way repeated measures ANOVA did not demonstrate a main effect for VOT ($F=3.73$, $p > 0.01$).

Mean latency values for the $N1$ and $N1'$ components for the /ba/-/pa/ continuum are shown in Fig. 7 and Fig. 8, respectively. For the /ba/-/pa/ continuum, a one-way repeated measures ANOVA was performed which revealed a significant main effect of VOT and $N1$ ($F=43.8$, $p < 0.000001$). Furthermore, a correlation analysis revealed that the $N1$ latency was significantly positively correlated with VOT ($r=0.82$, $p < 0.0001$). For the $N1'$ component (Fig. 8) a one-way repeated measures ANOVA did not demonstrate a main effect for VOT ($F=3.32$, $p > 0.01$).

2. Amplitude

For the /ba/-/pa/ continuum, a repeated measures ANOVA for $N1$ amplitude revealed a significant main effect of VOT ($F=13$, $p < 0.00001$). A correlation analysis revealed that the $N1$ amplitude was significantly negatively

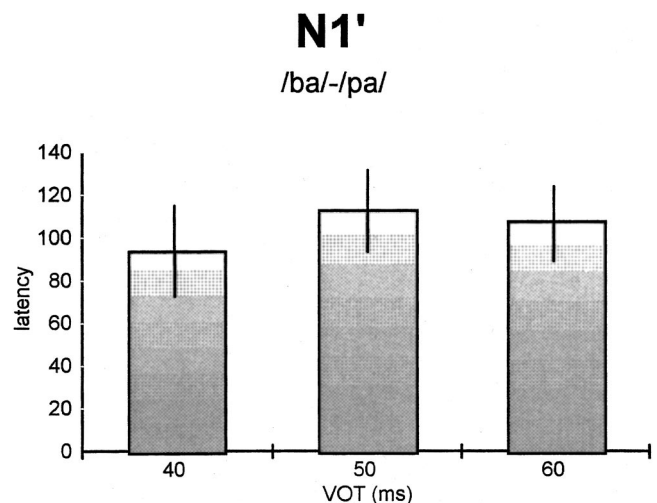


FIG. 8. Mean $N1'$ response latency values for the /ba/-/pa/ continuum. Error bars indicate ± 1 standard deviation.

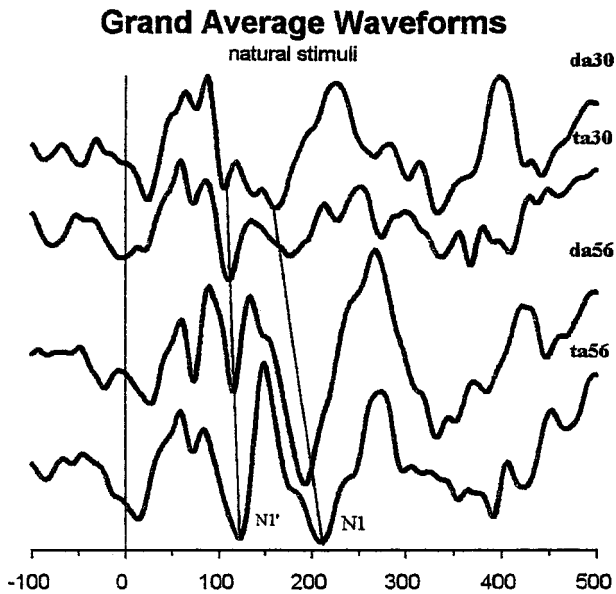


FIG. 9. Mean responses elicited by natural speech stimuli with VOTs of 30 and 56 ms which were perceived as /da/ and /ta/, respectively. Responses to all stimuli show two negative components ($N1'$ and $N1$).

correlated with VOT ($r=0.59$, $p<0.0001$). A one-way repeated measures ANOVA for $N1'$ amplitude did not show a main effect of VOT ($F=1$, $p>0.05$).

For the /ga/-ka/ continuum, a repeated measures ANOVA for $N1$ amplitude revealed a significant main effect of VOT ($F=15.4$, $p<0.0000001$). A correlation analysis revealed that the $N1$ amplitude was significantly negatively correlated with VOT ($r=0.59$, $p<0.0001$). A repeated measures ANOVA for $N1'$ amplitude did not show a main effect of VOT ($F=0.9$, $p>0.05$).

To summarize the results of the electrophysiologic experiment, a double-peaked $N1$ response was observed for stimuli with 40-ms VOT in both the /ga/-ka/ and the /ba/-pa/ continua. As pointed out in the stimulus description section, each stimulus in the two continua began with a brief burst and frication noise. For both continua, given the time frame of the $N1'$ peak latency and given that the latency did not show a significant correlation with VOT, it would appear that the $N1'$ component occurred in response to the burst at syllable onset. Taking into consideration the time frame of the peak latency of the $N1$ component and its significant positive correlation with VOT, it appears that this component occurs in response to the onset of voicing in the syllables.

With respect to the amplitude data, since the burst at syllable onset was identical for stimuli from within a continuum, the amplitude of the component which occurred in response to the burst (i.e., $N1'$) was not significantly different for stimuli from within both continua. However, as the duration between the burst and the onset of voicing increased, the response to the burst became more distinct to visual inspection, and the amplitude of the second component which occurred in response to voicing onset (i.e., $N1$) decreased. These results for the amplitude data are consistent with those of Steinschneider *et al.* (1999) and Simos *et al.* (1998), who have reported that the amplitude of AEP responses elicited by stimuli with long VOTs (i.e., 40, 60, and

80 ms) were significantly smaller than the amplitude of AEP responses elicited by stimuli with short VOTs (i.e., 0 and 20 ms). Overall, results from the present study are consistent with those of Sharma and Dorman (1999), who described similar $N1'$ and $N1$ components in the AEP waveform in response to stimuli varying in VOT across a /da/-/ta/ continuum.

IV. GENERAL DISCUSSION

In the case of the /ba/-/pa/ stimulus continuum, the results of the behavioral experiment revealed that listeners' perception of a change in phonetic categories from voiced to voiceless occurred reliably at a VOT of 40 ms (Fig. 2). The results of the electrophysiologic experiment for the same continuum showed that the change in the AEP waveform morphology from a single to double-peaked $N1$ component also occurred at a VOT of 40 ms (Fig. 4). These results are consistent with those from an earlier study (Sharma and Dorman, 1999), in which we found that the change in $N1$ morphology from single to double-peaked coincided with a change in perception from voiced to voiceless for a /da/-/ta/ continuum. On the other hand, in the present study, for the /ga/-/ka/ stimulus continuum, listeners' behavioral perception of a change in phonetic categories occurred reliably at a VOT of 60 ms (Fig. 1). As seen in Fig. 4, the change in $N1$ morphology from single- to double-peaked occurred at a VOT of 40 ms. That is, in the case of the /ga/-/ka/ continuum, the change in $N1$ morphology to a double-peaked component did not signal behavioral perception of a voiceless sound. Therefore, the results of the present study demonstrate that $N1$ morphology is not a reliable indicator of perception of voicing contrasts in syllable-initial position.

The data were also analyzed with respect to the correspondence between individual boundaries and individual changes in $N1$ morphology. For each subject, for each continuum, the individual VOT boundary was determined (the 50% identification point). We then determined for individual subjects the postboundary stimulus that was perceived as voiceless greater than 75% of the time. Finally, we noted for individual subjects if the $N1$ morphology changed from a single-peaked to a double-peaked response at this stimulus (i.e., the stimulus identified greater than 75% as voiceless). For the /ba/-/pa/ continuum, there was a correspondence between perceptual boundaries and changes in $N1$ morphology for 6 of 11 subjects. For the /ga/-/ka/ continuum, however, there was no correspondence for any subject. Therefore, the individual data conform to the overall findings for the group that the change in $N1$ from single- to double-peaked responses is not a reliable indicator of the perception of voiced and voiceless sounds.

Examination of the grand average waveforms for the two continua in Figs. 3 and 4, indicates that the two components (i.e., $N1'$ and $N1$) are fused in the AEP waveform at short VOTs and that a minimum temporal separation of 40 ms between the burst and voicing onset is required for the two components to be seen discretely in the temporal waveform. As stated above, our results with the synthesized speech continua show that the minimum VOT value (i.e., 40 ms) required for this temporal separation of $N1$ components

is dependent on acoustic properties of the stimulus rather than the perceptual categorization of the stimulus. In a subsequent experiment we confirmed this result using natural speech stimuli (Marsh, 1999). In that study, a /da/ and a /ta/ syllable (with VOTs of 30 and 56 ms, respectively) were recorded by a male speaker. Then, 26 ms of silence was added to the /da/ syllable (between the burst and onset of voicing) to create a new /da/ syllable with a VOT of 56 ms, and 26 ms of aspiration was removed from the /ta/ syllable to create a new /ta/ syllable with a VOT of 30 ms. Behavioral measurements in ten subjects revealed that subjects identified both the /da/ with a 30-ms VOT and the /da/ with a 56-ms VOT as /da/ and identified both the /ta/ with a 56-ms VOT and the /ta/ with a 30-ms VOT as /ta/. *N1* recordings in the same ten subjects (see Fig. 9) revealed that all four stimuli elicited double-peaked *N1* responses (albeit, not with identical morphology). Thus, the presence of a double-peaked *N1* component was independent of the phonetic categorization of the stimulus.

Yet another recent study confirms this outcome. Sharma and Dorman (2000) explored the changes in *N1* latency and morphology which are correlated with changes in VOT ranging from 0 to -90 ms across a prevoiced /ba/-/pa/ continuum. The results of a labeling experiment revealed, as expected, that the sounds along the continuum could be grouped into two categories (/ba/ and /pa/) by Hindi listeners but only into a single category (/ba/) by English listeners. However, *N1* responses were observed to change from single to double peaked at approximately 70 ms of prevoicing for both English listeners and Hindi listeners. Furthermore, the latencies of the *N1* components were not significantly different for the two groups of listeners. Thus, changes in *N1* waveform morphology and latency were independent of phonetic categorization of voicing.

Taken together, the results from the present study, Marsh (1999) and Sharma and Dorman (1999, 2000) suggest that the appearance of the double-peaked *N1* is influenced by multiple stimulus properties, including VOT, burst duration, amplitude of aspiration, and height of *F1*. More studies are

needed to systematically examine the effects of these and other acoustic properties of voicing on the morphology and the topography of the AEP waveform. The present study which examined evoked responses only from the scalp midline shows neither the presence of a double-peaked *N1* response, nor that the latency and amplitude of this response are related to the phonetic categorization of voicing. It is possible that recordings made at other scalp recordings or at intracortical locations may show a different relationship between *N1* morphology and phonetic categorization. In addition, it is possible that a different relationship will be found in children or adults learning phonetic categories.

In conclusion, our results show that a “double-peak” onset response as reported in recordings from the auditory cortex (e.g., Steinschneider *et al.*, 1995, 1999; Eggermont, 1995; McGee *et al.*, 1996, and Sharma and Dorman, 1999) is not a cortical correlate of the perception of voicelessness.

- Eggermont, J. (1995). “Representation of a voice onset time continuum in the primary auditory cortex of the cat,” *J. Acoust. Soc. Am.* **98**, 911–920.
- Klatt, D. (1980). “Software for cascade/parallel formant synthesizer,” *J. Acoust. Soc. Am.* **67**, 971–995.
- Marsh, C. M. (1999). “Cortical Auditory Evoked Potential Correlates of VOT,” Masters thesis, Arizona State University.
- McGee, T., Kraus, N., King, C., and Nicol, T. (1996). “Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe,” *J. Acoust. Soc. Am.* **99**, 3606–3614.
- Sharma, A., and Dorman, M. (1999). “Cortical auditory evoked potential correlates of categorical perception of voice-onset time,” *J. Acoust. Soc. Am.* **106**, 1078–1083.
- Sharma, A., and Dorman, M. (2000). “Neurophysiologic correlates of cross-language phonetic perception,” *J. Acoust. Soc. Am.* **107**, 2697–2703.
- Simos, P. G., Diehl, R. L., Breier, J. L., Molis, M. R., Zouridakis, G., and Papanicolaou, G. (1998). “MEG correlates of categorical perception of a voice onset time continuum in humans,” *Brain Res. Cognit. Brain Res.* **7**, 215–219.
- Steinschneider, M., Schroeder, C., Arezzo, J., and Vaughan, H. (1995). “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., Volkov, I., Noh, D., Garell, P., and Howard, M. (1999). “Temporal encoding of voice onset time phonetic parameter by field potentials recorded directly from the human auditory cortex,” *J. Neurophysiol.* **82**(5), 2346–2357.