Neurophysiologic correlates of cross-language phonetic perception

Anu Sharma[a] and Michael F. Dorman
Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

(Received 6 July 1999; revised 24 January 2000; accepted 28 January 2000)

This study examined neurophysiologic correlates of the perception of native and nonnative phonetic categories. Behavioral and electrophysiologic responses were obtained from Hindi and English listeners in response to a stimulus continuum of naturally produced, bilabial CV stimuli that differed in VOT from −90 to 0 ms. These speech sounds constitute phonemically relevant categories in Hindi but not in English. As expected, the native Hindi listeners identified the stimuli as belonging to two distinct phonetic categories (/ba/ and /pa/) and were easily able to discriminate a stimulus pair across these categories. On the other hand, English listeners discriminated the same stimulus pair at a chance level. In the electrophysiologic experiment N1 and MMN cortical evoked potentials (considered neurophysiologic indices of stimulus processing) were measured. The changes in N1 latency which reflected the duration of pre-voicing across the stimulus continuum were not significantly different for Hindi and English listeners. On the other hand, in response to the /ba/–/pa/ stimulus contrast, a robust MMN was seen only in Hindi listeners and not in English listeners. These results suggest that neurophysiologic levels of stimulus processing reflected by the MMN and N1 are differentially altered by linguistic experience. © *2000 Acoustical Society of America.* [S0001-4966(00)02805-8]

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

INTRODUCTION

A robust finding in the literature on infant speech perception is that infants 6–8 months old can discriminate speech sounds regardless of whether or not they are present in the infant's language environment. However, older infants and adults are less accurate in their perception of speech sounds that are not relevant to their native language (see review by Werker, 1994). These findings demonstrate that listeners' speech perception abilities are altered by experience with a particular language and that a lack of experience with a particular phonetic contrast has the effect of reducing sensitivity to that contrast. Certain nonnative contrasts appear to be easier to discriminate than others, perhaps due to exposure to their phonetic quality through some shared articulatory or phonetic features with sounds from one's own language (Best, 1994; Tees and Werker, 1984; Polka, 1991), and/or acoustic salience of the contrast (Burnham, 1986). Moreover, although for some contrasts re-learning is relatively easy using auditory training methods, for other contrasts it is often very difficult (see reviews by Pisoni *et al.*, 1994; Logan and Pruitt, 1995).

While there is overwhelming evidence for a ''perceptual deficit'' in discrimination of some nonnative speech contrasts for adult second language learners [e.g., Werker, 1994 (review)], the mechanisms of this deficit remain unclear. The prevailing view is that this deficit occurs due to re-alignment of cognitive categories resulting from ''higher-level'' attentional biases rather than a neural-sensory loss (see reviews by Pisoni *et al.*, 1994 and Werker, 1994). In fact, however, little is known about whether underlying neural-sensory representations of speech are actually altered by an individual's linguistic experiences.

A noninvasive method of directly examining neurophysiologic correlates of perceptual processes in humans is recording auditory evoked potentials (AEPs). The N1 and Mismatch Negativity (MMN) auditory evoked potentials, in particular, have proved useful in investigating neural correlates of speech processing. The N1 is an evoked response whose latency and morphology reflects the time of onset of acoustic events *within* speech, such as burst onset relative to voicing onset in a stop-consonant vowel syllable (Sharma and Dorman, 1999) and the onset of friction relative to vowel onset in a fricative-vowel syllable (Kaukoranta *et al.*, 1987; Ostroff *et al.*, 1999). The mismatch negativity (MMN) evoked response, on the other hand, is elicited by changes *between* acoustic signals (Sharma *et al.*, 1993; Sharma and Dorman, 1999). Primary generator sites for the N1 and MMN include the auditory cortical and thalamic areas (Nätäänen and Picton, 1987; Csepe, 1995). Both the N1 and MMN have traditionally been considered neurophysiologic indices of pre-attentive processing since they are recorded without active participation from subjects, whose attention is actively directed away from the eliciting stimuli by reading or watching a movie. Furthermore, the MMN can be elicited when the eliciting stimuli are not behaviorally discriminated, e.g., when stimuli are from within the same place of articulation category (Sams *et al.*, 1990; Sharma *et al.*, 1993; Maiste *et al.*, 1995). For these reasons, the MMN has been considered a measure of ''sensory'' or noncognitive processing (Nätäänen, 1992).

However, recently two studies which have examined

---

[a] Electronic mail: anu.sharma@asu.edu

cross-language speech perception have raised doubts as to whether the MMN reflects a sensory level of processing. In one study, Näätänen *et al.* (1997) examined perception of Finnish vowels in Estonian and Finnish listeners, and in the other study Winkler *et al.* (1999) examined perception of Hungarian and Finnish vowel contrasts in speakers of those two languages. The overall conclusion of these studies was that the MMN responses elicited by contrasts that were not phonetically relevant in the listeners' native language were significantly smaller than the MMN responses evoked by vowels that constituted different phonetic categories in the native languages of those same listeners. One interpretation of these results of diminished MMN responses for nonnative vowel contrasts is that the sensory level of processing traditionally attributed to the MMN is modified by language experience. However, another interpretation of these results is that the MMN does not reflect a entirely sensory level of processing, but rather a higher level susceptible to linguistic influences.

In this report we describe the results of an experiment in which the N1 and MMN evoked potentials were obtained from Hindi and English speakers in response to speech sounds varying in voice-onset-time which constituted different phonetic categories in Hindi and but not in English. If language experience alters the sensory level of processing, then both the N1 and the MMN responses to stimuli along the continuum would be different for Hindi speakers compared to English speakers. On the other hand, if only the MMN (and not the N1) response to stimuli along the continuum were different for Hindi speakers compared to English speakers, then such a finding would strengthen the possibility that the MMN reflects a level of processing which is beyond the sensory level.

I. BEHAVIORAL EXPERIMENT

A. Method

1. Subjects

Ten native Hindi speakers were recruited from the Arizona State University (ASU) campus. These were students who had spent at least the first 20 years of their life in India and had come to ASU for graduate studies. Subjects reported that they were fluent in several Indian languages (including Hindi) as well as English. Additionally, ten monolingual speakers of American–English were recruited from the ASU campus. These subjects reported no previous exposure to the Hindi language. All subjects were paid ($10/hour) for their participation. The subjects reported that they had no history of hearing, speech, or language problems.

2. Stimuli

The first author recorded the syllables /ba/ and /pa/ into a digital signal processing system. The system had 16 bit resolution and ran at 11.2 kHz. The vowels were produced with a slight /r/ coloring. The consonants were produced with pre-voicing as is appropriate in Hindi. The syllables were acceptable to native-English speakers as /ba/ and were acceptable to native-Hindi speakers as ‘‘baar’’ (which means ‘‘again’’ in Hindi) and ‘‘paar’’ (which means ‘‘side’’ in Hindi). To create the experimental stimuli, the pre-voicing segment of the original utterances were edited at zero crossings to create stimuli with pre-voicing durations of 0 to 90 ms (see Fig. 1).

3. Procedures

In the identification portion of the experiment, Hindi subjects were asked to listen to the experimental stimuli through headphones and to classify each stimulus either as a /pa/ (or ‘‘paar’’) or /ba/ (or ‘‘baar’’). Subjects were asked to indicate their responses by clicking with a computer mouse on panels marked PA and BA appearing on a computer screen. Each subject was given an initial practice session where he or she heard each stimulus along the continuum once. After that, ten repetitions of each of the nine stimuli were presented to the subject in a random order.

Following the identification task, subjects participated in a discrimination task. Based on results from a pilot experiment the stimulus pair chosen for the discrimination experiment had -10 and -50 ms VOT. An AX discrimination task was employed. On each trial, subjects heard two stimuli with an interstimulus interval (ISI) of 500 ms. Subjects were asked to determine whether the stimuli in the pair were ‘‘same’’ or ‘‘different.’’ Subjects indicated their responses by clicking on panels labeled ‘‘same’’ or ‘‘different’’ on the computer screen. The presentation of stimulus pairs was randomized within the test and across subjects. To familiarize the subjects with the task, an initial practice session of 20 trials was presented (10 same and 10 different trials). The experimental session consisted of a total of 100 trials (50 same and 50 different) for each stimulus pair.

B. Results

The group mean identification scores for each stimulus token are shown in Fig. 2. As can be seen, Hindi listeners consistently (i.e., $>75\%$ of the time) identified stimuli with VOT's of 0 and -20 as /pa/ and -50 and -90 ms as /ba/. On the other hand, English listeners reliably identified *all* the stimuli as /ba/.

The mean discrimination scores for the -10 and -50 ms VOT stimulus pair are shown in Fig. 3. Hindi subjects were able to discriminate the stimuli as different with a high degree of accuracy, while English listeners discrimination of the stimulus pair was close to chance. A paired *t*-test revealed that Hindi subjects' discrimination of the stimuli was significantly more accurate than their English counterparts ($t=10.8$; $df 18$; $p<0.0001$). Overall, for the Hindi subjects, the identification and discrimination results are consistent with previous reports of categorical perception for similar stimulus continua. On the other hand, for English listeners, the identification and discrimination results are consistent with previous reports that in the absence of training changes in pre-voicing are not phonemically perceived by monolingual speakers of American English [e.g., Werker, 1994 (review)].

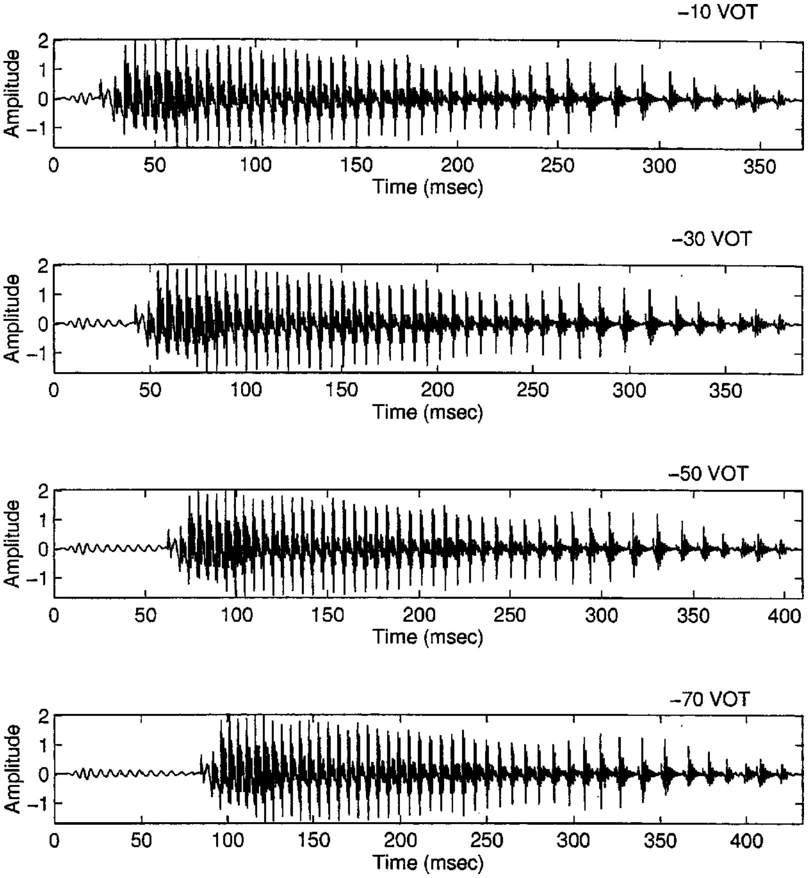

FIG. 1. Stimulus waveforms showing the duration of pre-voicing for four stimuli from the /ba/–/pa/ continuum. The voice-onset time (VOT) is indicated at the upper right-hand corner of the waveforms.

II. ELECTROPHYSIOLOGIC EXPERIMENT

A. Method

1. Subjects

For the electrophysiologic recording sessions the subjects' were seated comfortably in a sound booth. To control for arousal and to minimize subjects attention to the test stimuli, subjects watched a videotaped movie of their choice. Video tape audio levels were kept below 40 dB SPL. Subjects were asked to ignore the test stimuli that were presented through an insert earphones at 75 dB SPL in the right ear.

2. Stimuli

a. MMN. For the MMN recordings, the stimuli were the same as those used in the behavioral discrimination experiment (i.e., the −10 and −50 ms VOT pair). The MMN was elicited using an oddball paradigm in which repetitive presentations of a ''standard'' stimulus were occasionally replaced with a ''deviant'' or ''target'' stimulus. The deviant stimulus had a probability of occurrence of 15%. The −10 ms VOT stimulus was the standard and −50 ms VOT stimulus was the deviant. The stimuli were presented at an offset-to-onset ISI of 510 ms.

b. N1. The stimulus continuum was identical to the one used in the behavioral identification experiment. Repeated presentations of each stimulus separated by an onset-to-offset ISI of 800 ms were used to elicit the N1. The order of presentation of stimuli was counterbalanced across subjects.

3. Recording procedures

MMN and N1 evoked potentials were recorded using a NeuroScan, Inc. data acquisition system. Silver-chloride electrodes were placed on the midline (Fz, Cz, and Pz). The reference electrode was placed on the right mastoid and the ground electrode was placed on the forehead. Eye movements were monitored with a bipolar electrode montage (supraorbital to lateral canthus). Averaging was suspended when the eye channel registered blinks. The recording window included a 100-ms prestimulus period and 500 ms of post-stimulus time. AEPs were bandpass filtered on-line from 0.1 to 100 Hz (slope 12 dB/octave) and the recording gain was set at 1000. Responses that were greater than 100 μV were judged as noisy and rejected offline.

a. MMN. In the oddball paradigm, 2000 sweeps of the response to the standard and 300 sweeps of the response to the deviant stimulus were collected. By definition, the MMN is a response to stimulus change and occurs only when the deviant stimulus is presented in the context of a sequence of standard stimuli. A control condition was employed to ensure that the response reflects a change as opposed to reflecting simply the stimulus difference between the standard stimulus and the deviant stimulus. In this condition, the response which occurred to the deviant stimulus in the oddball paradigm was compared to the response evoked by the same stimulus when it was presented alone. If an MMN is present, then a relative negativity will be apparent only in the response elicited in the context of the oddball paradigm and not when the deviant stimulus was presented alone (Kraus *et al.*, 1995a).

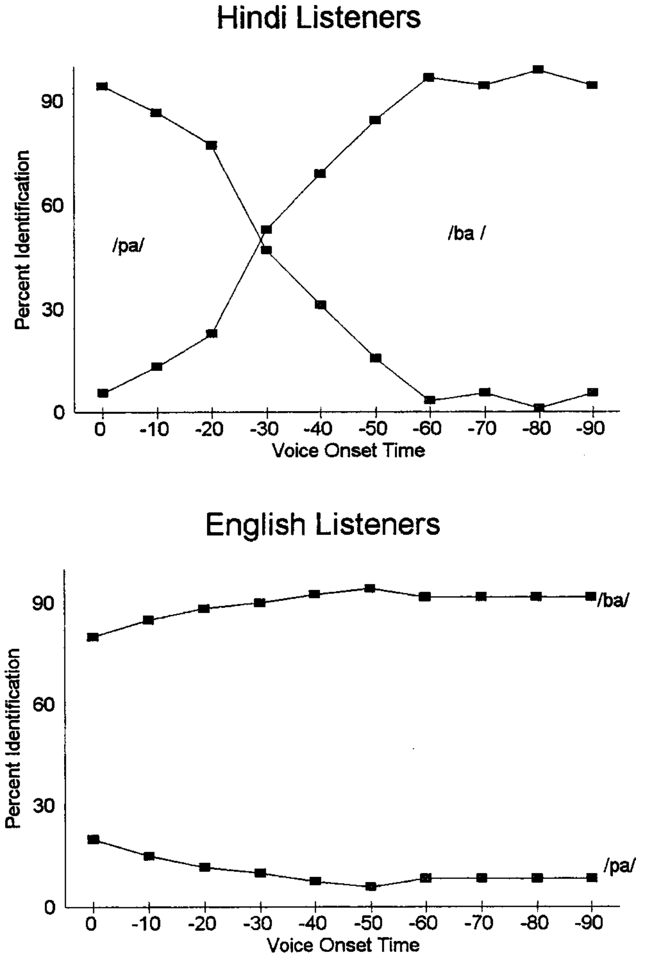

FIG. 2. Mean identification functions for the Hindi listeners (upper panel) and English listeners (lower panel).

b. N1. For the N1 recording, 300 sweeps elicited in response to each stimulus from the continuum were collected.

4. Data analysis

a. MMN. For individual subjects, sweeps were averaged to compute an individual average waveform elicited by the deviant stimulus when it occurred in the context of the standard stimulus (i.e., in the oddball paradigm) and by the deviant stimulus when it was presented alone. A ''difference wave'' was computed by subtracting the response to the deviant stimulus from the deviant-alone stimulus presentation. In individual subjects, the morphologies of the alone, deviant, and difference waves were assessed relative to previously described morphologies of speech-evoked potentials (Kraus *et al.*, 1992). In individual subjects, the MMN was apparent in the difference waveform while the N1 was apparent in the alone and deviant waveforms. The MMN was identified visually as a relative negativity following the N1 peak. The point of maximum negativity of the MMN component was noted and the adjacent relative positive peaks were selected as the MMN onset and offset. To determine the area of the MMN, a line was drawn from the onset to the offset of the MMN in the difference wave. The area was defined as a summation of the point by point multiplication ($\text{ms} \times \mu\text{V}$) of the enclosed difference wave. The area of the MMN was computed for individual subjects from the Fz electrode since the most robust MMNs were observed at that electrode site.

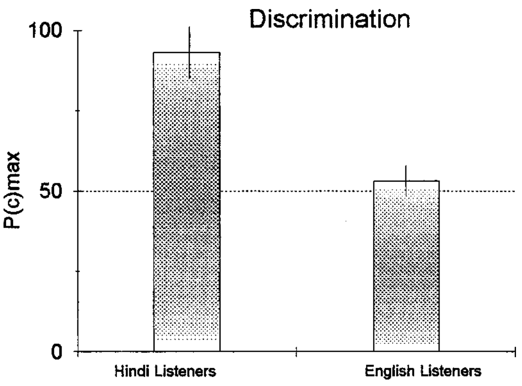

FIG. 3. Mean discrimination scores for the -10 and -50 ms VOT voicing contrast. Hindi subjects were significantly better at discriminating the stimulus pair as compared to English subjects ($p < 0.0001$). Error bars indicate ± 1 standard error.

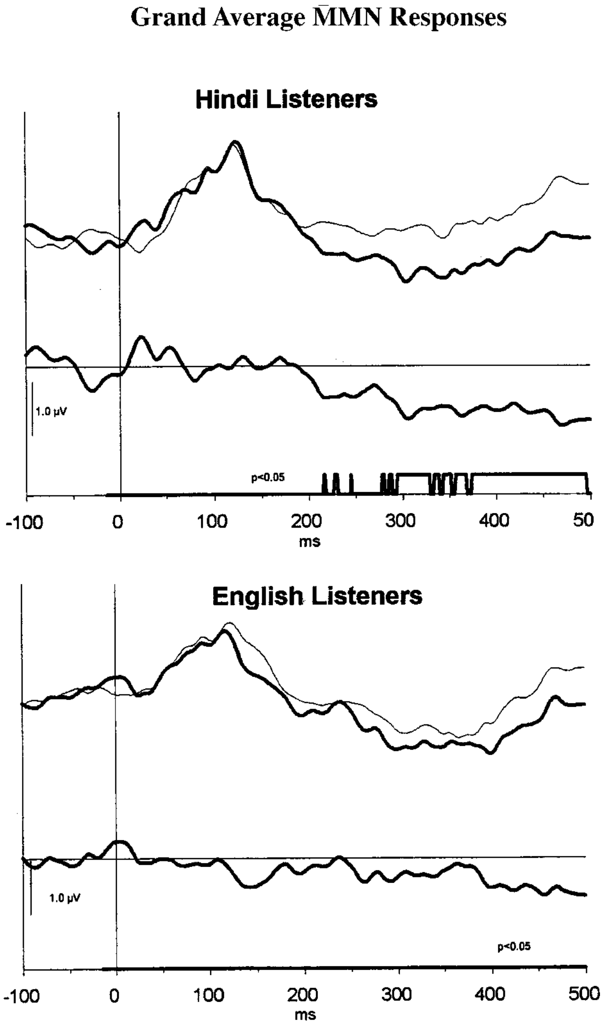

FIG. 4. Grand-averaged responses at the Fz scalp location for the Hindi listeners (left panel) and the English listeners (right panel). The top thin line is the response to the deviant stimulus when it was presented alone. The thick line is the response to the deviant stimulus when it occurred in the mismatch condition. The MMN response is seen in the difference wave (lower thick line) as a negativity. The boxes on the x-axis indicate the region of a significant mismatch response ($p < 0.05$).

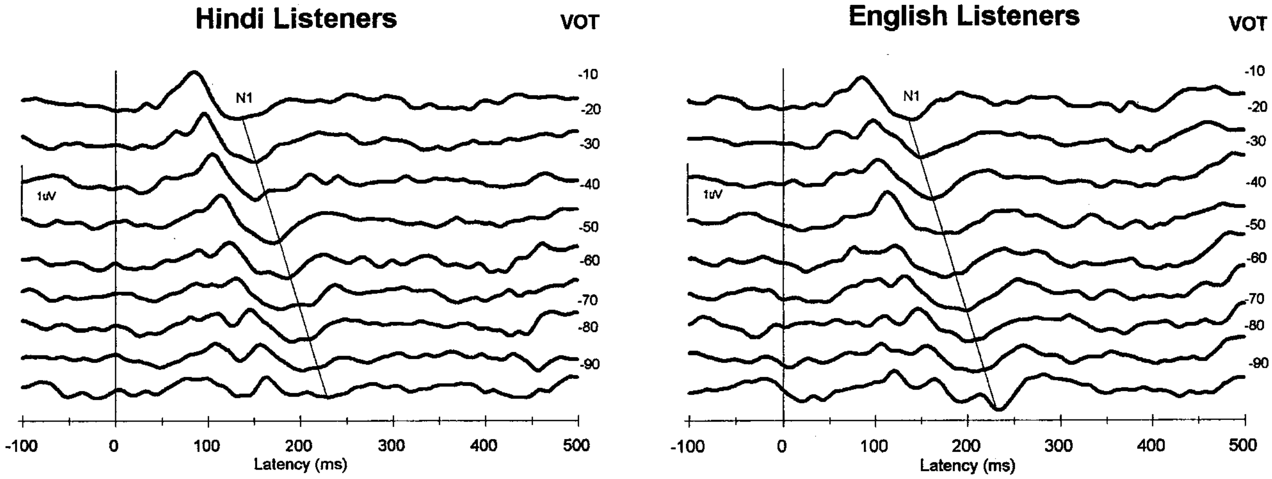

FIG. 5. Grand-averaged N1 responses for Hindi listeners (left panel) and English listeners (right panel). N1 latency was positively correlated with VOT for both groups of listeners [(Hindi group $r=0.79$; $p<0.0001$); (English group $r=0.8$; $p<0.0001$)].

Group averaged difference waves were computed by averaging individual subject difference waves. A point-to-point *t*-test of the values contributing to the group averaged difference waveform determined the period over which the grand averages were significantly different from zero at the $p<0.05$ level. A significant negativity in the mean difference wave was defined as the group MMN. The disadvantage of this analysis is that the number of *t*-tests performed and the high correlation between adjacent points on a waveform increases chances of spurious significant values. To counter this difficulty, Gutherie and Buchwald (1991) based on multiple regression techniques on P300 waveforms, concluded that if a continuous interval of at least 12 sampling points shows significance then the power of the statistical test is sufficient. Because the A/D sampling rates and filter settings for the MMN are essentially similar to the P300, a similar significance interval should give appropriate statistical power (Kraus *et al.*, 1995a; McGee *et al.*, 1997).

b. N1. For individual subjects, sweeps were averaged to compute an individual average waveform. All waveforms were digitally high pass filtered off-line at 4 Hz (filter slope 12 dB/octave). In the group mean waveforms, N1 was identified visually as prominent negative peak within the first half of the time window. In order to aid in peak identification and measurement in data from individual subjects, response windows of +25 ms were created around the peak in the group mean waveforms. Peak latencies were detected based on the recordings from the electrode site Fz, where the response amplitudes were the largest in the group mean waveforms. The latency was typically marked at the center. Group averaged waveforms were computed by averaging across the individual average waveforms.

B. Results

1. MMN

The group averaged waveforms from the Fz electrode site are shown in Fig. 4. A robust and statistically significant MMN was obtained for Hindi listeners. However, a statistically significant MMN was not elicited in English listeners. Furthermore, a *t*-test revealed that the MMN area was significantly larger for the Hindi compared to the English group ($t=2.3$; $df=18$; $p<0.05$). These results are consistent with those of Kraus *et al.* (1995b) and Tremblay *et al.* (1997) who reported that only minimal MMN responses were elicited to Hindi VOT contrasts in untrained English listeners.

To summarize, the native (Hindi) speakers who could behaviorally discriminate the stimulus contrast exhibited a robust MMN to the contrast, while in the nonnative (English) listeners whose perceptual discrimination of the same contrast was at a chance level, a significant MMN in response to the stimulus contrast was not observed.

2. N1

The grand average waveforms showing the variation in N1 latency across the stimulus continuum are shown in Fig. 5. A two way (VOT *x* listener group) ANOVA revealed a significant main effect of VOT ($F=66.45$, $p>0.0001$) but not of listener group ($f=2.15$, $p>0.18$). Correlational analysis revealed that the N1 latency was positively correlated with VOT for both groups of listeners [(Hindi group $r=0.79$; $p<0.0001$); (English group $r=0.80$; $p<0.0001$)], suggesting that the latency of the N1 reflects the VOT of the syllable. Therefore, to summarize, N1 latencies systematically reflected the acoustic change from the pre-voiced portion to the voiced portion within the syllables equally well for *both* the English and Hindi groups.

III. GENERAL DISCUSSION

We have examined behavioral and neurophysiologic correlates of the perception of pre-voicing which provides a phonetically relevant change in Hindi but not in English. The results of the N1 experiment showed that the neurophysiologic representation of the duration of pre-voicing and voicing onset was equally robust in native and nonnative speakers. In contrast, a robust MMN was elicited only for native speakers of Hindi.

As seen in Fig. 5, changes in latency of the N1 component reflected the acoustic change from pre-voicing to voicing onset occurring within a syllable. This result is consistent with previous reports based on animal studies (Steinschnieder *et al.*, 1995; Eggermont, 1995; McGee *et al.*, 1996) and human experiments (Kaukaranta *et al.*, 1987; Ostroff *et al.*, 1998; Sharma and Dorman, 1999) that changes in morphology and latency of the AEP waveform reflect acoustic events ongoing within the speech stimulus. For example, Sharma and Dorman (1999) reported that changes in N1 latency and morphology reflected sequential changes in acoustic events within a /da/–/ta/ stimulus continuum such as the interval between burst and onset of voicing. Similarly, Kaukaranta *et al.* (1987) and Ostroff *et al.* (1999) reported that the N1 latency and morphology reflected the onset of friction relative to vowel onset in a fricative vowel. However, these studies did not specifically examine whether speech-elicited changes in the AEP waveform indicate that the subject can discriminate or accurately identify the speech sounds. The results of the present study sheds some light on this issue. In our stimulus continuum, the duration of prevoicing was the acoustic feature which cued the change in phonetic categories (i.e., from /ba/ to /pa/) for Hindi listeners. For Hindi listeners then, as expected, the onset latency of the N1 response systematically reflected the change from pre-voicing to voicing onset across the stimulus continuum (Fig. 5). However, as can be seen in the same figure, changes in N1 latency also encoded the systematic changes in duration of pre-voicing in English listeners who were unable to behaviorally discriminate short versus long VOTs. This result suggests that neural encoding of acoustic changes ongoing in a speech signal as reflected by the N1 provides necessary but not sufficient information for behavioral discrimination of speech sounds.

In contrast to the N1 component, the MMN elicited in response to a stimulus contrast differed depending on whether that contrast constituted a phonemically relevant change in the listeners native language or not. In response to VOT differences which constituted phonetic categories in Hindi, but not in English, robust MMNs were elicited in Hindi listeners, but not in English listeners. These findings are consistent with those of Nätäänen *et al.* (1997) and Winkler *et al.* (1999) who reported that the MMNs in response to nonnative vowel changes were reduced compared to MMNs elicited by native vowel contrasts in Finnish and Hungarian. The present results are further supported by preliminary results from Phillips *et al.* (1999) who reported that Japanese listeners did not demonstrate an significant MMN response to a /r/–/l/ contrast and by the results of Dehaene-Lambertz (1997) who reported that an MMN-like response (which was recorded in an active rather than a passive listening paradigm) was absent in response to a Hindi retroflex versus dental contrast which was perceptually discriminated at chance by French listeners. Future studies should systematically examine the changes in MMN that accompany relearning of nonnative contrasts in adult second-language learners.

The different pattern of results obtained for the N1 and the MMN (i.e., similar encoding of the duration of prevoicing for both Hindi and English listeners by the N1, and, the presence of an MMN for Hindi listeners but not for English listeners) suggests that the N1 and MMN components of the AEP reflect functionally different levels of processing. The N1 reflects stimulus processing which occurs at a sensory level that is not modified by exposure to the phonetic categories of a language. The MMN reflects a level of processing in which language specific categories play a role. The characterization of the MMN as reflecting an entirely sensory level of processing is not tenable in light of the present results.

ACKNOWLEDGMENTS

We would like to thank the associate editor and two anonymous reviewers for their comments.

Best, C. T. (**1994**). ‘‘The emergence of native-language phonological influences in infants: A perceptual assimilation model,’’ in *The Development of Speech Perception*, edited by J. C. Goodman and H. C. Nusbaum (MIT Press, Cambridge, MA), pp. 167–224.

Burnham, D. (**1986**). ‘‘Developmental loss of speech perception: Exposure to an experience with a first language,’’ Applied Psycholinguistics **7**, 207–240.

Csepe, V. (**1995**). ‘‘On the origin and development of the Mismatch Negativity,’’ Ear Hear. **16**, 91–104.

Dehaene-Lambertz, G. (**1997**). ‘‘Electrophysiological correlates of categorical phoneme perception in adults,’’ NeuroReport **8**, 919–924.

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

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

Kaukaranta, E., Hari, R., and Lounasmaa, O. (**1987**). ‘‘Responses of the human auditory cortex to vowel onset after fricative consonants,’’ Exp. Brain Res. **69**, 19–23.

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

Kraus, N., McGee, T., Carrell, T., King, C., Tremblay, K., and Nicol, T. (**1995b**). ‘‘Central auditory system plasticity associated with speech discrimination training,’’ J. Cogn. Neurosci. **7**, 27–34.

Kraus, N., McGee, T., Sharma, A., Carrell, T., and Nicol, T. (**1992**). ‘‘Mismatch negativity event-related potential elicited by speech stimuli,’’ Ear Hear. **13**, 158–164.

Logan, J. S., and Pruitt, J. S. (**1995**). ‘‘Methodological issues in training listeners to perceive nonnative phonemes,’’ in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Baltimore, MD), pp. 351–378.

Maiste, A., Wiens, A., Hunt, M., Schery, M. and Picton, T. (**1995**). ‘‘Event-related potentials and categorical perception of speech sounds,’’ Ear Hear. **16**, 67–89.

McGee, T., Kraus, N., and Nicol, T. (**1997**). ‘‘Is there really a mismatch negativity? An assessment of methods for determining response validity in individual subjects,’’ Electroencephalogr. Clin. Neurophysiol. **104**, 359–368.

Nätäänen, R. (**1992**). *Attention and Brain Function* (Erlbaum, Hillsdale, NJ).

Nätäänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Livonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. (**1997**). ‘‘Language-specific phoneme representations revealed by electric and magnetic brain responses,’’ Nature (London) **385**, 432–434.

Nätäänen, R., and Picton, T. (**1987**). ‘‘The N1 wave of the human electric and magnetic response to sound: A review and analysis of the component structure,’’ Psychophysiology **24**, 375–425.

Ostroff, J. M., Martin, B. A., and Boothroyd, A. (**1998**). ‘‘Cortical evoked response to acoustic change within a syllable,’’ Ear Hear. **19**, 290–297.

Phillips, C., Marantz, A., Mcginnis, M., Pesetksy, K., Wexler, E. and Yellin. (**1999**). ‘‘Brain mechanisms of speech perception: a preliminary report,’’ MIT working papers in linguistics **26**, 513–191 (MIT Press, Cambridge, MA).

Pisoni, D. B., Lively, S. E., and Logan, J. S. (**1994**). ‘‘Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception,’’ in *The Development of Speech Perception*, edited by J. C. Goodman and H. C. Nusbaum (MIT Press, Cambridge, MA), pp. 121–166.

Polka, L. (**1991**). ‘‘Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions,’’ J. Acoust. Soc. Am. **89**, 2961–2976.

Sams, M., Aulanko, R., Aaltonen, O., and Näätänen, R. (**1990**). ‘‘Event-related potentials to infrequent changes in synthesized phonetic stimuli,’’ J. Cogn. Neurosci. **2**, 344–357.

Sharma, A., and Dorman, M. F. (**1999**). ‘‘Cortical auditory evoked potential correlates of categorical perception of voice-onset-time,’’ J. Acoust. Soc. Am. **106**, 1078–1083.

Sharma, A., Kraus, N., McGee, T., Carrell, T., and Nicol, T. (**1993**). ‘‘Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential,’’ Electroencepholog. Clin. Neurophysiol. **88**, 64–71.

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

Tees, R. C., and Werker, J. F. (**1984**). ‘‘Perceptual flexibility: Maintenance of recovery of the ability to discriminate nonnative speech sounds,’’ Can. J. Psychol. **38**, 579–590.

Tremblay, K., Kraus, N., Carrell, T., and McGee, T. (**1997**) ‘‘Central auditory system plasticity: Generalization to novel stimuli following listening training,’’ J. Acoust. Soc. Am. **102**, 3762–3773.

Werker, J. F. (**1994**). ‘‘Cross-language speech perception: Developmental change does not involve loss,’’ in *The Development of Speech Perception*, edited by J. C. Goodman and H. C. Nusbaum (MIT Press, Cambridge, MA), pp. 93–120.

Winkler, I., Lehtoksoki, A., Alku, P., Vainio, M., Czigler, I., Csepe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, H., Iivonen, A., and Nätäänen, R. (**1999**). Cognitive Brain Research **7**, 357–369.