Working Papers of the Cornell Phonetics Laboratory
No. 1 (December 1983), pp. 77-90

Production and Perception of the Voicing Contrast in Indian and American English

Katharine Davis and Mary Beckman

Abstract. We measured VOT's of initial stops and tested stop identification in speakers of Indian English (Hindi bilinguals) and American English. Our measurements revealed significant inter- and intra-dialectal differences in distribution of VOT's for the speakers, but only an inter-dialectal difference in identification patterns. Thus production of VOT can vary considerably within a general dialect group without affecting the uniformity of the group's perception norms.

Introduction

The description of phonemic stop contrasts along such dimensions as voiced/voiceless and aspirated/unaspirated was simplified greatly when Lisker and Abramson [1964] discovered that these contrasts can be classified by voice onset time (VOT). The perceptual relevance of this laryngeal timing mechanism has since been confirmed in many experiments showing that VOT is a robust cue in the perception of the phonemic contrasts so classified [cf. Lisker and Abramson 1970, Fischer-Jørgensen 1972, Williams 1977a, Shimizu 1977, Keating et al. 1981].

Lisker and Abramson found that the categories differentiated by VOT seem to fall into three well-defined groups--namely, stops with negative VOT ("voicing lead"), stops with zero or small positive VOT ("short lag"), and stops with large positive VOT ("long lag"). Languages in their study with more than a two-way contrast along the VOT scale used all three of these categories, and those with two-way contrasts always seemed to use adjacent categories. (For example, Dutch and Spanish voiced versus voiceless stop

phonemes contrast voicing lead with short lag, whereas Cantonese unaspirated versus aspirated stop phonemes contrast short lag with long lag.) The importance of this finding has been confirmed by experiments that suggest a general psychoacoustic basis for the divisions between any two adjacent categories [cf. Streeter and Landauer 1976, Miller et al. 1976, Pisoni 1977]. Thus voice onset time as used in the world's languages seems to be a three-step scale rather than a continuum.

We performed an experiment to see whether this finding can be extended to production and perception patterns in two dialects of English that apparently differ in the VOT categories used to contrast intial voiced versus voiceless stops. The two dialects that we studied were American English and the variety of English spoken by Hindi speakers ("Indian English").

Lisker and Abramson's pioneering cross-language study showed that American English initial stops, unlike the similarly labeled Dutch and Spanish phonemes, fall into the short lag versus long lag categories. Three of their four American English speakers consistently produced voiced /b d g/ with zero or short positive VOT, and all four speakers produced voiceless /p t k/ with long VOT's like those in the analogous aspirated stops in Cantonese. On the other hand, informal observations have suggested that Hindi speakers do not have as much aspiration in their English /p t k/ as do American speakers.

If the ostensibly identical phonemic categories in these two dialects of English do indeed differ in VOT, they should fall into different adjacent pairs of the three VOT categories suggested by Lisker and Abramson's study. Moreover, there should be consistent interdialectal perceptual confusion for the intermediate category, short lag, which should be identified with the voiced stops by American English speakers and with the voiceless by Indian English speakers.

We measured VOT in minimally contrasting English words spoken by American and Indian speakers, and then tested these same speakers' perception of the various word tokens produced by the speakers in both dialect groups.

The results of the production test revealed a great deal more idiolectic variation than Lisker and Abramson's study would suggest. Many of the American speakers produced voicing lead for /b d g/ (as did the single "anomalous" speaker in Lisker and Abramson's experiment). Moreover, the

results for most of the Indian speakers suggested a slight influence from American English, although the overall trend confirmed our earlier informal observations.

The results of the perception test, on the other hand, revealed a surprising stability in the perception of these categories. They confirmed all earlier studies showing the perceptually relevant distinction for American English to be short lag versus long lag, even for the speakers who consistently produced voicing lead for /b d g/. Moreover, they revealed that the perceptually relevant distinction for Indian English is voicing lead versus voicing lag, even for those speakers whose production seemed to have been influenced by American English norms.

In this paper, we will describe the experiments that we performed, and discuss the implications of the results for second-language acquisition patterns and sound changes involving voice onset time.

Methods

Production Experiment. Three minimal pairs, differing only in the voicing of the initial stop consonant (namely, puck/buck, tall/doll, cull/gull) were embedded in the frame sentence "Please say_____ again." The sentences containing these target words were placed in a list along with twice as many other sentences containing filler words. The order of the list was random, except that no two sentences containing minimally contrasting target words appeared next to each other. The target sentences occurred five times each on the list.

Ten subjects participated in the experiment. Five were native speakers of American English (four from New York State, one from North Carolina), and five were native speakers of the Hindi dialect spoken in North Central India. The speakers in the latter group had learned Indian English in school between the ages of six and ten. Four of the Indian speakers were long-time residents of the United States, having lived in the country for three to five years. The fifth speaker had arrived only six months previous to the experiment. Nine of the ten subjects were graduate students in engineering at Cornell University; the other was the spouse of a student. All were between the ages of 22 and 32.

The subjects read the list of minimal pair sentences individually in a soundproof booth. The readings were recorded at 7 1/2 ips on a good quality tape recorder. If

any error in reading was made, the subject was asked to repeat the entire sentence.

The target utterances were then digitized at 20 kHz onto the Phonetic Lab's PDP-11/40 computer. VOT was measured from the digitized wave-form, using a wave-form editor developed by David Walter and modified by Mark Pedrotti. In tokens with voicing lead, negative VOT was measured from the drop in intensity signifying stop closure after the preceding vowel to the highest-intensity spike in the release. In tokens with voicing lag, positive VOT was measured from the highest-intensity spike in the release burst to the onset of voicing in the following vowel. When there was a double release (with two equally intense spikes), the measurements were made to the first spike (for negative VOT) or from the second spike (for positive VOT).

Perception Experiment. After all the measurements were completed, the target words were excised from their frames and converted back to an analog recording in a random order on another tape. The tokens were separated on the tape by 1.5-second long silent intervals.

The stimulus tape of 300 tokens was presented in a quiet room to all ten subjects who had participated in the production experiment. The subjects were instructed to identify each stimulus by circling on an answer sheet the word heard. The answer sheet consisted of a list of word pairs, corresponding to the order of presentation on the tape, with the word having the initial voiceless stop on the left and the word having the initial voiced stop on the right.

Results

Production Experiment. The results of the production experiment are shown in Figures 1 and 2. The graphs in these figures are histograms of the distribution of measured VOT values for the two different types of stops as produced by the speakers in the two dialect groups.

As Figure 1a shows, the VOT values for /b d g/ produced by American speakers had a clear bimodal distribution; three-quarters of these stops had voicing lead and the other quarter had short lag. There was no overlap in VOT between these two types of voiced stops. There was also virtually no overlap bewteen the VOT's for /b d g/ with short lag and the VOT's for /p t k/, which had a clear unimodal distribution in

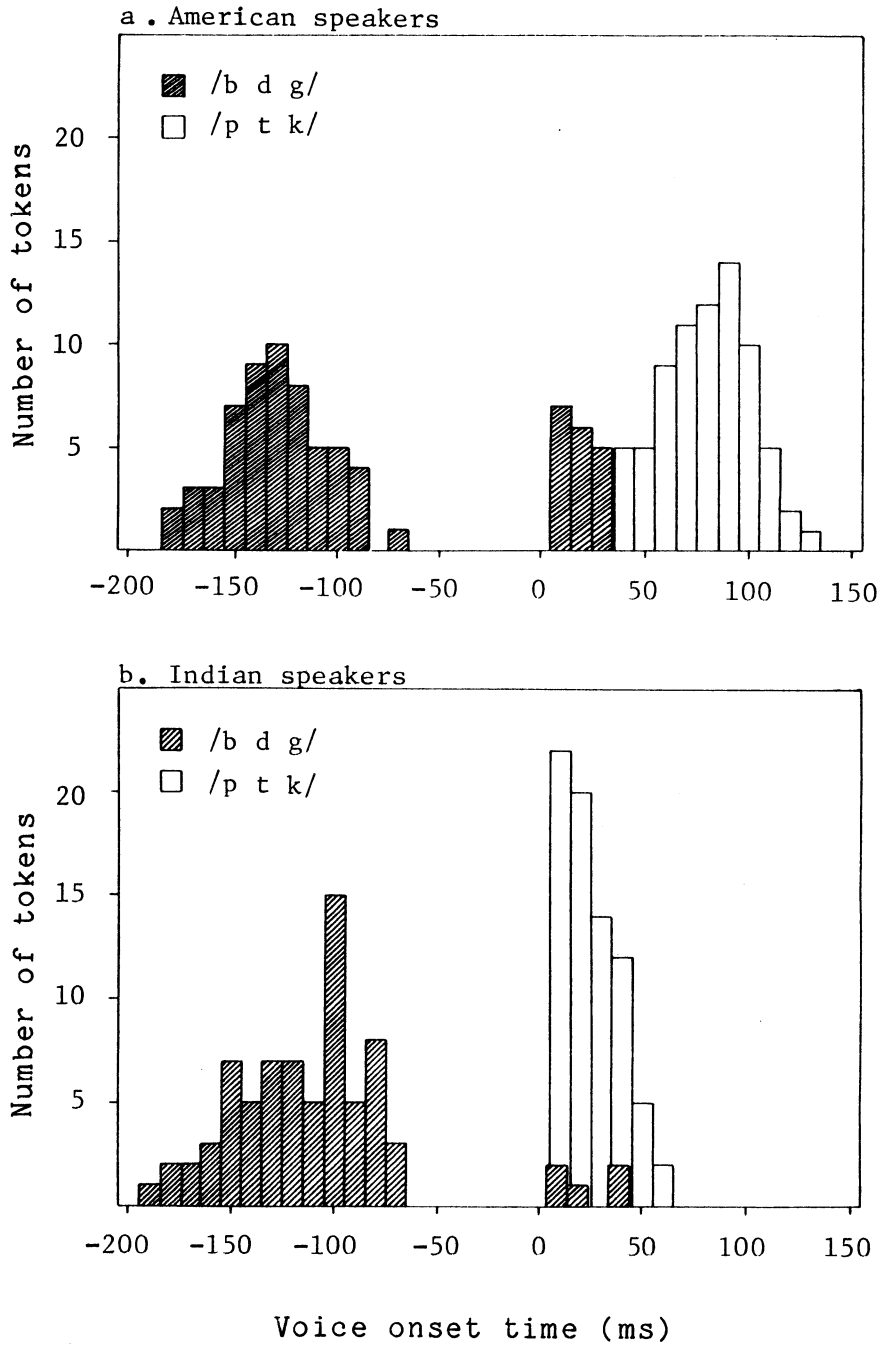

Figure 1. Distribution of measured VOT's in the tokens of words with voiced and voiceless initial stops produced by speakers of American English (upper figure) and by speakers of Indian English (lower figure).

the long lag region, with a mean value of 80 ms (standard deviation 22).

These distributions contrast sharply with the distributions of VOT's for the same stops produced by the Indian speakers. As Figure 1b shows, nearly all of the Indian English tokens of /b d g/ had voicing lead. Only five (7%) of the tokens had short lag. Moreover, the values for these /b d g/ with short lag fell completely within the range covered by the Indian English tokens of /p t k/. The latter group of phonemically voiceless stops had a mean VOT of 25 ms (st. dev. 8). These results indicate a substantial inter-dialectal difference in the distributional patterns for VOT's of initial voiced and voiceless stops.

There is also evidence of intra-dialectal variability, as shown in Figure 2. Figure 2a is a histogram for the VOT values produced by American speaker AB. (For comparison, the two types of dashed lines in the background show again the distribution of values for all five American speakers displayed in Figure 1a.) As can be seen from the figure, speaker AB was responsible for nearly all of the tokens of /b d g/ with short lag. She produced only one token of a phonemically voiced stop that had any voicing during the stop closure. In other words, the bimodal distribution of VOT values for American English /b d g/ reflects an apparent idiolectal split among the speakers.

There is evidence for an idiolectal split among the Indian English speakers as well. Figure 2b is a histogram for the VOT values produced by Hindi speaker SB. (Again, for comparison, the two types of dashed lines repeat the distributions for all speakers in the dialect group shown in Figure 1b.) SB was the Indian speaker with the least exposure to American English. Her tokens of /b d g/ had on the average significantly more voicing lead than did those produced by the other Hindi speakers ($t=5.021$, $p<0.0001$), and she was responsible for none of the five tokens in this phonemic category that had positive VOT. Her tokens of /p t k/, moreover, had significantly less voicing lag than did the voiceless tokens produced by the other Indians ($t=5.126$, $p<0.0001$). These results suggest that the other four speakers, all of whom had been in the United States for three or more years, had been influenced somewhat by American English. Their production of /b d g/ and /p t k/ seems to have shifted somewhat toward the American English norms.

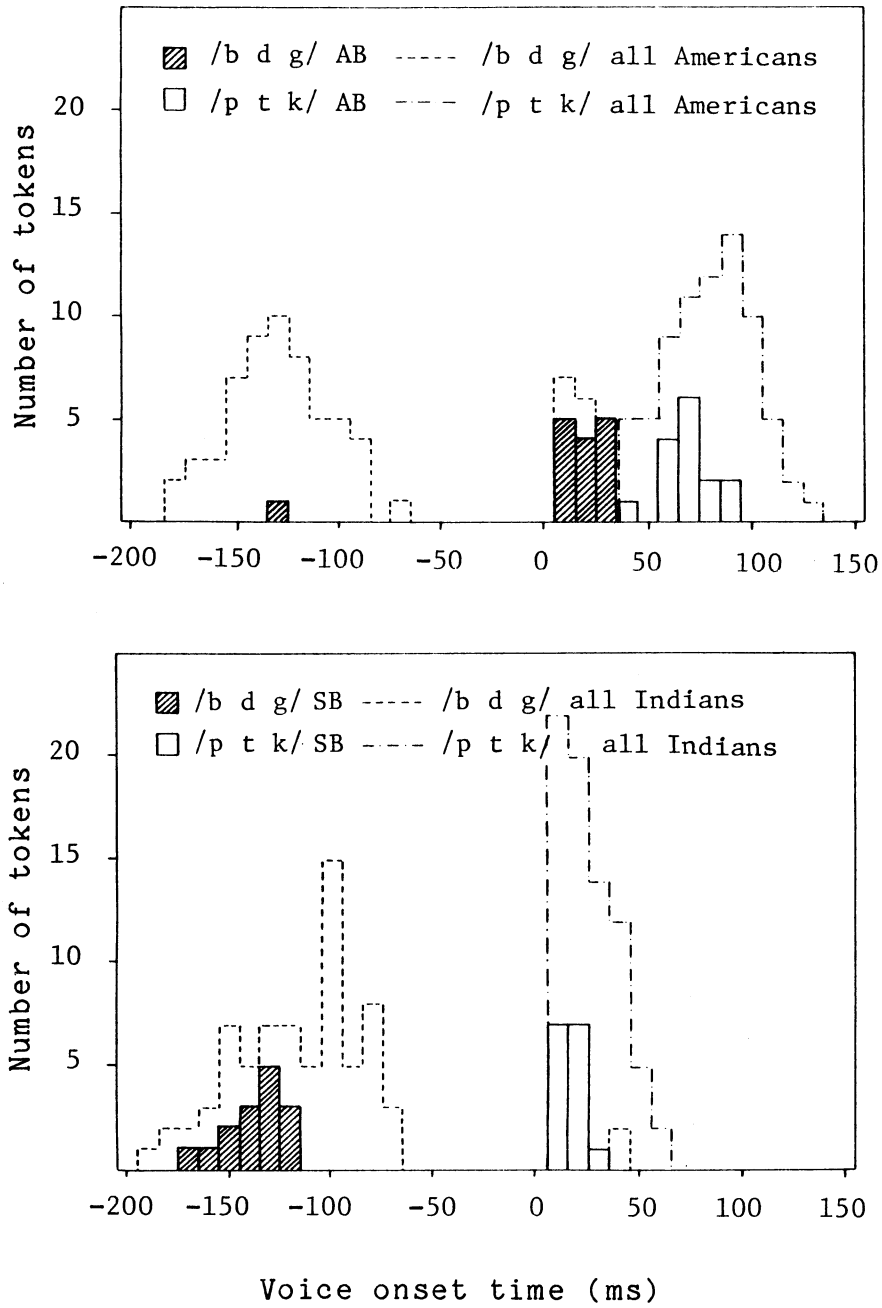

Figure 2. Distribution of measured VOT's in the tokens produced by American speaker AB (upper figure) and by Indian speaker SB (lower figure). The two different types of dashed lines in the figures represent the distribution of VOT's for voiced and voiceless stops by all of the speakers in each dialect group.

Perception Experiment. The results of the perception test are shown in Tables I and II. Table I shows the error rate for the American and Indian listeners divided by the phonemic identity of the stop and the identity of the speaker who produced the token heard.

Table I. Percent error by stop type and listener group for individual speakers.

Listener \ Speaker	American						Indian					
	AB	BS	CC	GC	JB	overall	SL	RB	SB	UM	VB	overall
a. /b d g/												
American	5	0	0	1	1	2	11	5	0	0	0	3
Indian	41	6	1	7	17	14	15	9	0	1	0	5
b. /p t k/												
American	0	4	0	0	0	1	21	28	52	29	15	29
Indian	0	3	3	1	0	1	5	8	0	7	3	5
c. overall												
American	3	2	0	1	1	1	16	17	26	15	7	16
Indian	21	5	2	4	8	8	10	9	2	4	1	5

As the next-to-last column in the table shows, the Americans did better in identifying American English tokens than they did in identifying Indian English tokens. Their overall error rates for these two categories were 1% versus 16%, a highly significant difference ($\chi^2(1)=104.86$, $p<0.005$).

The reason for this difference becomes apparent when the token categories are further broken down by stop type. The Americans misidentified Indian English /b d g/ only 3% of the time (not substantially more often they they did American English tokens of this category), but they misidentified the Indians' /p t k/ 29% of the time. Moreover, when this error rate is even further broken down by the identity of the individual speaker, it is seen that SB's tokens of these phonemically voiceless stops produced significantly higher error rates from the Americans than did those of the other

Indian speakers ($\chi^2(1)=23.91$, $p<0.005$). Since the Hindi speakers generally produced short lag for /p t k/, and since among them SB produced the shortest lag, the breakdown of error suggests a confusion on the part of the American listeners caused by the nearly complete overlap of the Indian English VOT's for voiceless stops with the VOT's identified with /b d g/ in American English.

The error rates in Table I also show a complementary confusion on the part of the Indian listeners. As the last column in the table shows, their overall error rate for American English tokens was not significantly worse than that for Indian English tokens ($\chi^2(1)=4.51$, $p>0.025$). When the tokens heard are broken down by stop type, however, a substantial difference emerges. The Indians misheard 14% of the tokens of /b d g/ produced by the American speakers, as opposed to only 5% of those produced by the Indian English speakers. Moreover, a significant part of the higher error rate is accounted for by speaker AB. Her tokens of the phonemically voiced stops were misheard 41% of the time, significantly more often than those of the four other Americans ($\chi^2(1)=55.17$, $p<0.005$). Since AB was the American speaker who consistently produced short lag for these stops, the high error rate on the part of the Indian listeners suggests confusion caused by the overlap with VOT's identified with /p t k/ in Indian English.

These suggestions are confirmed by the results shown in Table II. This table gives error rates for the tokens broken down by stop type and VOT production category, where "long lag" is defined as 50 ms or more VOT. (50 ms was chosen as the cutoff point between short and long lag because it falls in the middle of the overlap between American English and Indian English /p t k/.)

The highest overall error rates in Table II are for Americans listening to /p t k/ with short lag (27%) and for Indians listening to /b d g/ with short lag. The Americans listeners also misidentified /b d g/ with short lag more often than they did /b d g/ with voicing lead, but their error rate for this category is still significantly better than the Indian listeners' error rate ($\chi^2(1)=29.93$, $p<0.005$).

A second intriguing result evident in Table II is that the identification patterns for SB and AB were not different from those for the other speakers in their respective dialect groups, even though their production patterns were idiosyncractic (as demonstrated above in the results of the production experiment). Thus SB's identification of /b d g/ with short lag was not significantly worse than that of the

Table II. Percent error by individual listener for various VOT production categories.

Listener \ Production category	/b d g/ with voicing lead	/b d g/ with short lag	/p t k/ with short lag	/p t k/ with long lag
American:				
AB	0	13	30	0
BS	0	9	23	1
CC	2	13	24	1
GC	0	0	43	1
JB	2	22	16	1
overall	1	11	27	1
Indian:				
SL	2	30	9	0
RB	6	43	5	3
SB	1	43	9	3
UM	6	48	1	0
VB	4	48	1	1
overall	4	43	5	1

other Indian speakers ($\chi^2(1)=0.01$, $p>0.9$), even though she had had far less exposure to American English, and consequently had less American-like VOT production. Similarly, AB's identification of /b d g/ with short lag was not better than that of the others in her dialect group, even though she was responsible for most of the tokens in this category. Conversely, her error rate for /p t k/ with short lag was not significantly worse than that of the other Americans ($\chi^2(1)=0.46$, $p>0.3$), even though the VOT's in this category overlapped considerably more with her VOT's for /b d g/.

Discussion

Four conclusions can be drawn from these results. First, even within the same general dialect group, there can be marked variation in the VOT categories used for a given phonemic type. The differential treatment of /b d g/ by AB

and the other Americans confirms the results for the distribution of short lag versus voicing lead among the American English speakers in Lisker and Abramson [1964], except that in the present study, the speakers who used voicing lead for /b d g/ were in the majority, rather than in the minority as in Lisker and Abramson's group. Other studies of American English also show more speakers like AB and fewer speakers like the others [cf. Keating et al 1981]. On the other hand, the distribution in the present experiment is similar to that of Caramazza et al. [1973] for Canadian English speakers. Further study is necessary to determine whether this apparently idiolectal variation is due to some finer regional or social dialect variation within the general category "American English."

A second conclusion is that different general dialects of the same language can differ substantially in the VOT categories used for the same phonemic categories. In conformity with our earlier informal observations, the present study shows that Indian English speakers use short lag for /p t k/ where American English speakers use long lag. It is interesting to relate this cross-dialectal difference to the patterns in the native language of this particular group of Indian English speakers. Hindi has a three-way contrast along the VOT scale, so that the Indian English speakers have available a category of aspirated stops that they could identify with English /p t k/, and yet they do not. The Indian speakers in this group, at any rate, claim that the Hindi aspirated stops are very different from the American English initial voiceless stops. Although measurements of Hindi stops elicited from these same speakers show VOT's comparable to those in American English /p t k/, informal observations of the wave-form suggest that the amplitude of aspiration is higher in the Hindi stops, perhaps accounting for the perceptual dissimilarity [cf. Repp 1979].

A third conclusion is that production of VOT seems to be susceptible to influence by other dialects. The Indian English speakers who had been in the United States for several years had mean VOT's for /p t k/ that were significantly more positive than that for the recent arrival, and they even produced a few tokens of /b d g/ with short lag (instead of with the voicing lead that is the norm for this dialect group). On the other hand, these four speakers had not altered their VOT production so far as to produce /p t k/ with the long lag VOT's that are the norm for American English. This slight shift suggests that, in terms of production, VOT is a continuum rather than a three-step scale.

Finally, it can be concluded that perception patterns are not like production patterns for VOT. They seem not to be so variable within the general dialect group, and are far less susceptible to influence by other norms. Thus despite the idiolectal differences evident in the American speakers' production patterns, they all perceived the various VOT categories in the same way.[1] Similarly, despite the differential influence from American English that was evident in the Indian speakers' production data, they all perceived the different tokens according to their group norm. Although VOT may be a continuum in terms of production, it does seem to fall into a small number of well-defined perception categories.

Thus in general, it seems that the perception of VOT is uniform and stable within general dialect groups, whereas production is variable from speaker to speaker within dialects, and susceptible to influence from other dialects. That production can apparently change or differ without an accompanying change or difference in perception is reminiscent of other situtations in which the two types of behavior are independent. For example, small children often acquire the perceptual category for a "difficult" sound such as [s] or [θ], before they can produce the sound; a two-year-old that says [ʃi] for 'see' nevertheless may not accept the same pronunciation from an adult. Conversely, there is some evidence that bilinguals acquire distinct production patterns in speaking their two languages without differentiating their identification patterns in listening to them. This result is seen for VOT production and perception in a study of French-English bilinguals by Caramazza et al. [1973] and in a study of Spanish-English bilinguals by Williams [1977b]. Elman et al. [1977], on the other hand, found that "strong" Spanish-English bilinguals do identify stops in the short lag category differently depending on the context language of the identification task, suggesting that perception may eventually catch up with production.

---

[1] Mikoś et al. [1978] observe an analogous situation in Polish. They say that some of their Polish speakers used long lag for voiceless stops, even though the general distribution for the language is short lag for voiceless versus lead for voiced. Despite this idiolectal difference, however, all of the speakers displayed the same category boundary appropriate to the language's production norm in identifying stimuli along a synthetic /ta/-/da/ continuum.

This result also has important implications for sound change. That the four Hindi speakers with the most exposure to American English had apparently shifted their production somewhat toward the American distribution pattern without shifting their perceptual categorization suggests that some sound changes can be accomplished by a gradual drift along a production continuum. This drift might go unnoticed within the lifetime of any given speaker until it crosses a discontinuity in the perceptual scale, and is reinterpreted by others acquiring the langage for the first time. Such a model of one type of sound change is especially attractive in the case of VOT, which seems to be fairly continuous in terms of articulation but shows strong evidence of having discrete steps perceptually.

Acknowledgements

We thank our informants Sushma Banthia, Vinod Banthia, John Benci, Rajendra Bordia, April Brown, Craig Cameron, Sanjiva Lele, Umesh Mishra, Bill Schaff, and Glenn Swan, for their willing cooperation in this set of experiments. We also thank Mark Pedrotti for technical assistance.

References

Caramazza, A., G.H. Yeni-Komshian, E.B. Zurif, and E. Carbone. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. J. Acoust. Soc. Am. 54: 421-428

Elman, Jeffrey L., Randy L. Diehl, and Susan E. Buchwald. 1977. Perceptual switching in bilinguals. J. Acoust. Soc. Am. 62: 971-974.

Fischer-Jørgensen, Eli. 1972. Perceptual studies of Danish stop consonants: Tape-cutting experiments with Danish stop consonants in initial position. Ann. Rep. Inst. Phonetics, Univ. Copenhagen 6: 104-168.

Keating, Patricia, Michael J. Mikoś, and William F. Ganong. 1981. A cross-language study of range of voice onset time in the perception of initial stop voicing. J. Acoust. Soc. Am. 70: 1261-1271.

Lisker, Leigh, and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustic measurements. Word 20: 384-422.

-----. 1970. The voicing dimension: Some experiments in comparative phonetics. In Proc. 6th Int. Cong. Phonetic Sciences, pp. 563-567 (Prague: Academia).

Mikoś, Michael, Patricia Keating, and Barbara Moslin. 1978. The perception of voice onset time in Polish. J. Acoust. Soc. Am., Suppl. 1 63: S19.

Miller, James D., Craig C. Wier, Richard E. Pastore, William J. Kelly, and Robert J. Dooling. 1976. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. J. Acoust. Soc. Am. 60: 410-417.

Pisoni, David B. 1977. Identification and discrimination of the relative onset time of two-component tones: Implications for voicing perception in stops. J. Acoust. Soc. Am. 61: 1352-1361.

Repp, Bruno H. 1979. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. Language and Speech 22: 173-189.

Shimizu, Katsumasa. 1977. Voicing features in the perception and production of stop consonants by Japanese speakers. Studia Phonologica 11: 25-34.

Streeter, L.A., and T.K. Landauer. 1976. Effects of learning English as a second language on the acquisition of a new phonemic contrast. J. Acoust. Soc. Am. 59: 448-451.

Williams, Lee. 1977a. The voicing contrast in Spanish. J. Phonetics 5: 169-184.

-----. 1977b. The perception of stop consonant voicing by Spanish-English bilinguals. Perception and Psychophysics 4: 289-297.