766

ON TESTING LINGUISTS' PERCEPTION OF INITIAL STOP VOICING: DOES IT REFLECT NATIVE LANGUAGE PHONOLOGY?

LEIGH LISKER
University of Pennsylvania
and
Haskins Laboratories

FOREWORD

In early 1985 I had the good fortune to be a visitor at the Linguistics Department of Osmania University, where it was both pleasant and instructive for me to meet and work with the group of researchers and students then in residence there. I feel it to be particularly appropriate to present, as my small contribution to the present volume honoring Professor Bh. Krishnamurti on the occasion of his sixtieth birthday, a report based in some measure on data that I was enabled to collect because it was by his kind invitation that I was in Hyderabad. To him and to other members of the linguistics faculty, in particular Dr. K. Nagamma Reddy, and to the postgraduate students in linguistics who underwent with patience and good humor the rather tedious perception testing to which they were subjected, I want to express my deep appreciation.

THE QUESTION

Linguists are generally agreed that the stop consonants of English should be divided into two sets of phonological categories: /b, d, g/ and /p, t, k/. These sets are now commonly referred to as voiced and voiceless respectively, although it is recognized that a stop need not be voiced phonetically in order to be classified as a member of the first set. In utterance-initial position, where voiceless /b, d, g/ are commonly found, members of the contrasting set /p, t, k/ are marked by both voiceless closure and voiceless release (aspiration). It is not at all surprising, then, to find that native speakers of the language, including linguistics students in the early stages of phonetic training, are not very sensitive to the presence vs. absence of pre-release voicing in utterance-initial position, or to variations in the duration of this feature.

The situation with regard to the stop consonants of English is unlike that of many other languages of the world, including most of those spoken in India. Hindi and Telugu, for example, have stop consonant categories that are regularly produced with pre-release voicing in initial position as well as

elsewhere, and at the same time they possess consonant categories that are both voiceless and unaspirated. In both these languages, therefore, a [± voiced] difference is said to be linguistically distinctive.

The above-mentioned considerations encourage us to suppose that because speakers of English evaluate variations in initial stop voicing differently from speakers of Telugu and Hindi, they will also "perceive" them differently, that is, they will behave differently when subjected to the so-called perception testing commonly practiced by modern speech researchers. This supposition should, however, be viewed with some caution. First of all, the general belief that sensitivity to an acoustic phonetic property depends crucially on whether that property is said to play a distinctive role in the listener's language can readily be disproven. Thus, while it is true that most English speakers have difficulty in distinguishing the dental from the retroflex stops of Hindi and Telugu, and this difficulty is plausibly explained by the absence of such a distinction in English, English speakers have no trouble in detecting the equally foreign difference between the aspirated and the ejective stops of, for example, Kechua. Therefore we cannot so readily assume that speakers of English do in fact perceive closure voicing differently from speakers of Hindi and Telugu. The question remains to be addressed: Do speakers of languages in which the role of stop voicing differs perceive variations in voicing differently? With the technology now available for phonetic research it is possible to test for a difference in perceptual status quite directly, - by submitting to listerners' judgments sets of acoustic patterns that are identical except for the feature of interest, the experimental variable of closure voicing duration.

EXPERIMENT ONE: TWO HINDI SPEAKERS

A native speaker of Hindi from New Delhi recorded a series of CV syllables, reading from a list of items printed in devanagari. From these two tokens of each of the syllables डा टा ढा ठा [Da Ta D^ha T^ha] were selected for computer storage and processing as follows. After digitization at a sampling rate of 10kHz, a wave editing program was used first to "devoice" each of the voiced stops, yielding stimuli that we might represent as [D_oa] and [D_o^ha]. Then these items, as well as the originally voiceless [Ta] and [T^ha] syllables, were supplied with pre-release voicing derived from the two original tokens of [Da] and [D^ha]. This procedure was followed in order to guard against the possibility of erroneous conclusions based on accidental, i.e. artifactual, differences between natural and unnatural combinations of pre-release and post-release signal intervals. A preliminary test indicated very strongly that 1) the unmodified items were identified by Hindi-speaking listeners as "intended," and 2) the pre-release voicing derived from [Da] and

[D^ha] were perceived identically in all the contexts in which both were presented. In effect, then, the operations just described yielded the following four phonetic types: [Da, D^ha, Ta, T^ha], and four additional types that we will designate [$D_o a$, $D_o^h a$, $T_v a$, $T_v^h a$], but whose phonetic status is yet to be determined. All eight items were read out of computer memory in a randomized order. Two Hindi speakers, both of them trained in linguistics and experienced in teaching Hindi to American students, served as the first test subjects. Their task was to identify each of the stimuli with a written response in devanagari. The test was run twice for each listener. Their responses are given in Table 1.

Labeling Responses: Two Hindi Speakers

Table 1

Stimulus Source (s)		Responses							
		Speaker 1				Speaker 2			
Pre-Release	Post-Closure	Da	Ta	D^ha	T^ha	Da	Ta	D^ha	T^ha
Da/D^ha	Da	32	–	–	–	32	–	–	–
Da/D^ha	D^ha	–	–	32	–	–	–	32	–
–	Ta	–	8	–	–	–	8	–	–
–	T^ha	–	–	–	8	–	–	–	8
–	Da	2	6	–	–	2	6	–	–
–	D^ha	–	–	8	–	–	–	–	8
Da/D^ha	Ta	32	–	–	–	32	–	–	–
Da/D^ha	T^ha	–	–	32	–	–	–	31	1

These data suggest the following conclusions, which must be very tentative because of the small number of listeners tested and the small number of responses elicited. 1) The presence of a voicing signal during an initial closure is an important cue in deciding that an initial stop is /D/ or /D^h/ rather than /T/or /T^h/. 2) deletion of voicing is somewhat less than totally effective in shifting /D/ to /T/ and /D^h/ to /T^h/. 3) The responses of the two linguist-listeners make it appear that the post-closure murmur of [D^h] and [D_o^h] is perceptually distinguishable from the voiceless aspiration of [T^h], sufficiently so for one of the listeners, speaker. 1, to report [D_o^h] as / m / (/D^h/). 4) This difference, however, between the murmured and the voiceless aspirated releases did not prevent either listener from reporting the artificially voiced [T_v^h] as the linguistic equivalent of [D^h]. Thus the prefixing of a "voice bar" to all four syllable tokens with originally voiceless stops effected category shifts, despite the preservation of whatever [-voiced] cues contained in the post-closure intervals.

A REPLICATION OF THE EXPERIMENT: TWO TELUGU SPEAKERS

The same set of stimuli used to elicit category labeling judgments from the two Hindi-speaking linguists in Philadelphia was presented to a small group of postgraduate students and staff members of the Department of Linguistics of Osmania University. In general the data obtained in Hyderabad were rather "noisier" than those gathered in Philadelphia, very possibly because of the use of a small inexpensive tape-recorder rather than high quality equipment. Thus most of the listeners failed to respond uniformly to the unmodified syllables that served as "control" stimuli in the test set. However, two of those who reported Telugu as their first language gave largely "correct" identifications of the control items [Da, Ta, D^ha, T^ha]. Their responses (Table 2) are in general agreement with those of the Hindi-speaking linguists. There is possibly a slightly greater tendency for the Telugu speakers to favour voiceless responses to both the artificially devoiced [D_o D_o^h] and the artificially voiced ([T_v T_o^h]) items, but this might well indicate that the low-frequency and low-amplitude "voice bar" was not as well reproduced by the equipment used in Hyderabad.

TABLE 2

Labeling Responses: Two Telugu Speakers

Stimulus Source (s)		Responses (%)			
Pre-Release	Post-Closure	Da	Ta	D^ha	T^ha
Da/D^ha	Da	100	–	–	–
Da/D^ha	D^ha	3	–	97	–
–	Ta	1	99	–	–
–	T^ha	–	–	–	100
–	Da	25	63	–	12
–	D^ha	–	12	–	88
Da/D^ha	Ta	94	3	3	–
Da/D^ha	T^ha	3	–	91	6

A SECOND EXPERIMENT: VARIABLE VOICING DURATION

In order to obtain crosslanguage data concerning the sensitivity of linguists to variations in the duration of pre-release voicing a single reading of the Hindi syllable दी (*dī*) was selected as the source for a set of stimuli generated, as before, by computer assisted waveform editing. The source syllable was recorded by the same Hindi speaker who had provided the recorded materials used in the first test. The measured duration of the pre-release voiced interval was 65 milliseconds, and six additional test stimuli were generated by reducing this duration by 10 msec decrements to a minimum value of 5 msec. Deletion was accomplished by the removal of signal intervals immediately preceding the release burst, thus preserving the original signal onset characteristics. The minimum duration of 5 msec is well below the psychoascoustic detection threshold, thus equal effectively to a voicing duration of 0 msec (Cf. Pisoni, 1977). Three sets of linguist listeners were asked to report whether or not they detected any pre-release voicing in any to the seven acoustically different test stimuli. Prior to administration of the test there was a "training" session in which the two stimuli with minimum and maximum voicing durations (5 and 65 msec) were presented in pairwise fashion. Each listener was directed to observe that one of each pair of stimuli was clearly produced with initial closure voicing, and that the other was devoid of such signal. In the test itself these items, together with the five others having intermediate durations of voicing, were presented in random order for labeling as either [+ voiced] or [+ voiced]. The linguist listeners were thirteen native speakers of American English, twelve native speakers of Telugu and four whose first language was Hindi. It should perhaps be noted that many of the Americans were essentially monolingual, while all the Indian linguists were competent in more than one language. All twenty-nine linguists ostensibly shared a common notion as to the meaning of [+voiced] and [−voiced] stop consonants.

The data derived from the linguists tested, shown in Figure 1, are radically different from what we should expect. To be sure, the three groups of linguists show differences in their sensitivity to variations in the feature of interest, and the Hindi and Telugu speakers behaved similarly as against the English speakers. Thus it could be said that the data reflect the difference in the reputed linguistic relevance of voicing in Hindi/Telugu versus English. However the relation between labeling behavior and linguistic relevance is the inverse of what is expected. The four Hindi speakers reported *all* stimuli more often [+voiced] than [−voiced], and the Telugu speakers were only slightly more inclined to report some as [−voiced]. The English speakers, on the other hand, shifted systematically from [−voiced] to [+voiced] judgments with increasing duration of the pre-release voicing. It

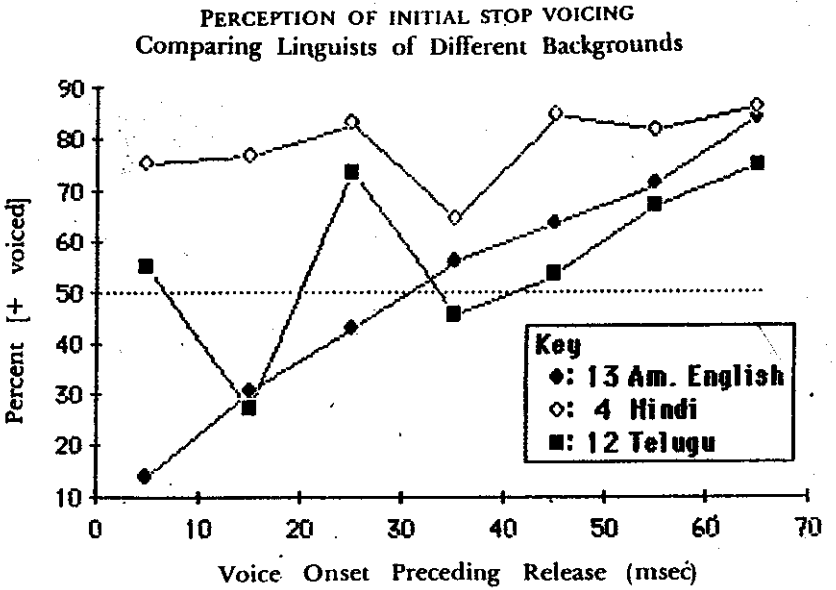

FIGURE 1

is worth noting that all the linguist-listeners said, when questioned after being tested that all the stimuli were "voiced," i.e. members of the categories /d/, /ɖ/ and /ɗ/. For phonetically naive English speakers this is just what we should expect; for Hindi and Telugu speakers it is just what we should *not* expect.

SUMMARY

The conclusions to be drawn from the two tests are not mutually consistent. From the first we should conclude, as expected, that closure voicing is perceptually important from the voiced stop categories of Hindi and Telugu, and that properties of the post-closure signal are, though significant, of lesser importance. From the data of Table 1 we see also that these conclusions may need qualification, since token-specific and listener-specific variations may also be significant determinants of labeling behavior. From the second test, on the other hand, it appears that closure voicing is a dispensible feature of voiced stops, and that the post-release interval of a voiced stop is sufficiently powerful perceptually to make it difficult for Hindi and Telugu speakers, even given linguistic training (which presumably includes some instruction in phonetics), to make a [± voiced] decision based solely on the pre-closure interval. Almost as surprising, the English speakers, despite the phonological irrelevance of initial closure voicing as a defining feature of the English /b,d,g/ categories, *were* able to make a voicing decision based on

the interval. It is true that not all the English speakers tested were equally successful to attending to the feature of interest, so that a few behaved as was initially expected. The stimuli used in the two tests were of course different in several ways: in the first they were sequences of retroflex stops and the vowel [a], while in the second a single token of the syllable *dental stop* +[i] served as the stimulus source. In phonetic descriptive terms both डा ([Da]) and दी ([di]) begin with voiced unaspirated stops, and there is no apriori reason to suppose that the post-closure [+voiced] nature of the syllable डा is more attenuated than that of दी . But this by no means entirely rules out the possibility of a voicing difference between retroflex and dental stops, or of a difference conditioned by the vocalic context, or that individual tokens of the same phonetic and phonological type might differ acoustically to a degree great enough for a phonogically significant difference to emerge as an aftermath of editing. Thus factors of individual listener difference and of individual phonetic token difference must be counted as possible bases for the apparent inconsistencies in our data. On the other hand, the data force us to reckon with the possibility that a phonetic feature's phonological relevance is of less than crucial importance as an explanation of linguists' phonetic labeling behaviour. If it *is* a significant factor, its effect may not be manifested in a way that can be charaterized as perfectly straightforward. Certainly we cannot infer from the phonetic judgments of our experimental subjects just how they manage the feature of closure voicing in speaking their native languages, or what status linguists accord that feature in those languages.

ACKNOWLEDGEMENTS:

In addition to the help of members of the Linguistics Department of Osmania University, I want to acknowledge the assistance of The American Institute of Indian Studies, which provided a travel grant to Hyderabad, and of NICHD grant HD-01994 to Haskins Laboratories, which made it possible to apply the computer facilities used in this study.

REFERENCES

PISONI, DAVID B. 1977. Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America* 61, 1352-1361.