PHONETIC UNIVERSALS AND HINDI SEGMENT DURATION

Manjari Ohala & John J. Ohala
Department of Linguistics
University of Alberta
Edmonton, Alberta
Canada T6G 2E7
and:

San José State University
San José, Calif. 95192

University of California
Berkeley, Calif., 94720

ABSTRACT

We report initial results of an acoustic phonetic study of Hindi designed to further text-to-speech synthesis of the language. We explore the usefulness of prior work on phonetic universals in predicting sub-phonemic phonetic variation. We conclude that phonetic universals cannot always predict such variation but it does serve to direct the search for it.

INTRODUCTION

Text-to-speech systems have been developed or are under development for several languages. But there are hundreds of languages spoken around the world which are potential markets for synthesis systems which still have to be devised. Although each language has its own phonemic norms which have to be discovered inductively, what about the rules which generate the normal phonetic variation which, though not lexically distinctive, makes speech sound natural and more intelligible? Will these also be language-specific or are there phonetic universals which can be put into play? This paper reports efforts to discover relevant phonetic sound patterns in Hindi and to relate them to previously documented phonetic universals.

Table 1. The consonant and vowel phonemes of Hindi.

Stops/Affricts	**Lab**	**Dent**	**Rtflx**	**Pal**	**Vel**	
Voicels unasp.	p(:)	t(:)	ṭ(:)	c(:)	k(:)	
Voicels asp.	p^h(:)	t^h(:)	$ṭ^h$(:)	c^h(:)	k^h(:)	
Voiced	b(:)	d(:)	ḍ(:)	j(:)	g(:)	
Breathy voiced	b^h	d^h(:)	$ḍ^h$(:)	j^h(:)	g^h(:)	
Fricatives						
Voicels	f	s(:)		š		
Voiced		z				
Nasals	m(:)	n(:)				
Approximants	w(:)			j		h
Taps/Flaps						
Voiced		r(:)	ṛ			
Breathy voiced			$ṛ^h$			
Laterals		l(:)				

Vowels	*Front*		*Central*		*Back*	
High	i	ĩ			ũ	u
	I	Ĩ			Ũ	U
Mid	e	ẽ			õ	o
	ɛ	ɛ̃	ə	ə̃	ɔ̃	ɔ
Low	æ		a	ã		

THE SOUNDS OF HINDI

Table 1 gives the principal phonemes of Hindi. Those marked with '(:)' can be realized either as singletons or geminates (distinctively long) in certain environments. (See [18] for further details). (In the figures we use the symbol '@' as a substitute for the schwa, IPA [ə].)

SPEAKERS AND SPEECH CORPUS

The talkers were three male native speakers of Standard Hindi with home towns in Uttar Pradesh, Bihar, and Delhi. Recordings of their speech were made in the Language Laboratory of the Jawahar Lal Nehru University, Delhi, using high-quality analog portable equipment. All test words were read in 2 different random orders in suitable frame sentences as detailed below. We give here means based on a maximum of 3 tokens/word type/speaker or 9 tokens; 3.5% of the target tokens were unusable but never fell below 5/word type.

ANALYSIS METHODS

The recorded speech was band-pass filtered at 68 Hz to 7.8 kHz, digitized at 16.7 kHz and analyzed with the aid of waveform and LPC spectral displays produced by the CSRE speech analysis software and related programs. Segmentation criteria were those conventionally used in phonetics: boundaries between consonants and vowels were based on the location of abrupt changes, especially in amplitude.

RESULTS

Vowel Durations: Effect of Voicing

Citation-form vowel durations before different final consonants were obtained from /pVC/ syllables, most of them nonsense words, spoken in the frame /vo ___ aja/ "he ___ came," where the C was one of a variety of consonants (as reported below) and V was from a set of eight oral vowels (we exclude here the results for /a/ and /ə/).

Fig. 1 gives the mean durations for the eight oral vowels before following voiced and voiceless stops and affricates, categorized separately by the vowel and consonant environment. The short and long vowels averaged 75.3 and 181.9 msec, respectively, or a ratio of 1:2.4 (but recall that this does not include [a]). By comparison, in English this ratio is more like 1:1.3. The long vowels' duration varies inversely with the traditional feature of vowel height, as has been found for many other languages [4,11]. The data in Fig. 1 agree well with those of Dixit [5], reporting on his own speech, except for the vowel [ɛ] which for him was some 13% shorter than [e] but is some 4% longer than [e] here. These data also show that vowels are shorter before voiceless as opposed to voiced stops, a tendency found in numerous other languages [3, 11]. The magnitude of the difference, 16 msec., is congruent with that in most languages except English where vowels before voiced consonants can be as much as 50% longer than before voiceless. It is widely believed that this aspect of vowel duration is phonologized in English, i.e., is not strictly a phonetically-based effect (except that it may be a phonologized exaggeration of an original phonetic effect) [8, 9]. Reinforcing this conclusion about the different nature of these durational differences between English and other languages is the fact that in Hindi this difference seems to be more like a constant additive difference whereas in English the difference has a multiplicative character, i.e., is proportional to the inherent vowel duration [2].

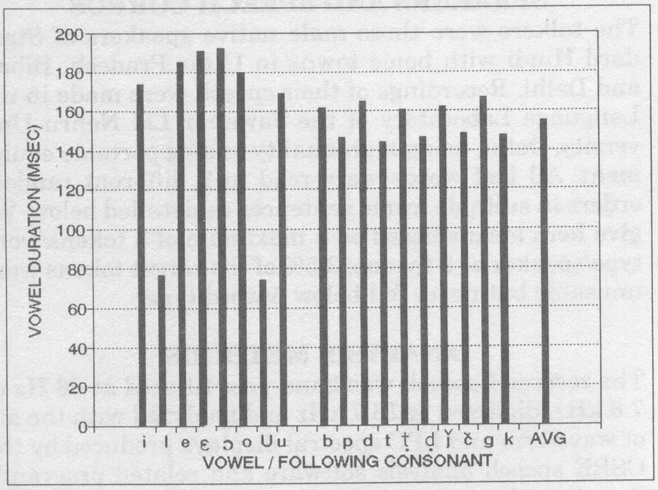

Fig. 1. On the left, mean durations of 8 vowels before all consonants; on the right, before the 10 stops/affricates differing in voicing and the overall mean duration before these same 10 phonemes.

Vowel Durations: Effect of Aspiration

When measuring the duration of vowels following voiceless aspirated or breathy voiced stops, it has often been asked whether the period of aspiration which is colored by the following vowel's resonances, should be counted as part of the vowel's duration or not [19]. To investigate this we used syllables of the form /C$^{(h)}$Vp/ in the above frame, where Cʰ was one of the 8 aspirated or breathy voiced stops and the V = /i u a/. Fig. 2 gives on the left the mean vowel durations following initial voiced and voiceless unaspirated stops, excluding affricates (where /b/ stands for /b d ḍ g/ and /p/ for /p t ṭ k/). Here we assume that the vowel duration measurements are not distorted by the release characteristics of the preceding stop. To the right of these we give the same vowels' durations measured in two ways, first including the aspiration or breathy voice ('R' for 'release') and then excluding it (/pʰ/ stands for /pʰ tʰ ṭʰ kʰ/ and /bʰ/ for /bʰ dʰ ḍʰ gʰ/. If we assume that intrinsic vowel duration should be unaffected by the initial consonant, the "true" vowel starts about half-way into the period of aspiration. This agrees with earlier findings for English [19] and is reminiscent of the placement of the so-called 'P-center', the perceptual onset of syllables [15].

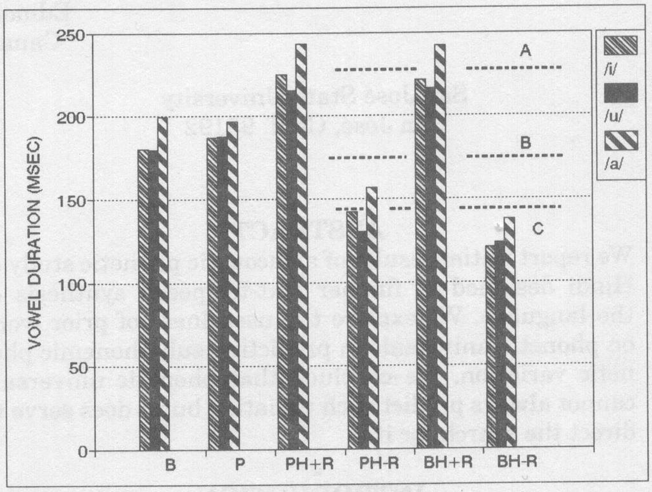

Fig. 2. Mean durations of /i u a/ following (from left to right) voiced, voiceless unaspirated, voiceless aspirated, and breathy voiced stops; in the case of the latter two categories, measurements presented two ways: with the distinctive release (R) and without it. Dashed line A possibly overestimates the vowels' intrinsic durations, line C possibly underestimates them, and B may be the "true" duration (congruent with the durations following the first two categories).

Maddieson [13] reported finding in Hindi and other languages somewhat longer vowels before aspirated as opposed to unaspirated stops. We investigated this using test words of the form /sVCa/ embedded in the frame /vo __ kəho/ "that ___ say," where the V were /a/ or /ə/ and the C were various aspirated or unaspirated stops (including breathy voiced stops which have traditionally been included under the feature 'aspiration'). We used bi-syllabic test words here so that the aspirated stops could appear intervocalically (the [+/- aspirate] contrast can be neutralized in absolute word-final position). The results, Fig. 3, reveal no consistent difference in vowel duration due to aspiration.

Geminate Consonants

Although geminate vs. singleton consonants are usually differentiated by the longer consonantal 'hold' for the geminate, it is acknowledged that other secondary features may keep them apart, too [1, 10, 11, 13]. Moreover, the magnitude of the difference between these classes of consonants can vary from one language to another. To study geminates we used minimal pairs of the sort CVC:V vs. CVCV uttered in the frame /vo ___ do/ "that ___ give." Fig. 4 shows the mean durations for these two types of consonants. On the average the geminates were 76 msec longer than the singletons, a ratio of 1.96:1. It would be important to know whether the difference between these two classes is a constant one or proportional (i.e., whether additive or multiplicative) but the data are perhaps too sparse to tell: after normalizing the two measures, the variance of the absolute difference is somewhat less than the variance of the durational ratios, pointing to an additive factor. Although ratios around 2:1 between geminates and singletons are common in languages [7, 10], the ratio is significantly higher in some cases, e.g., Japanese [6].

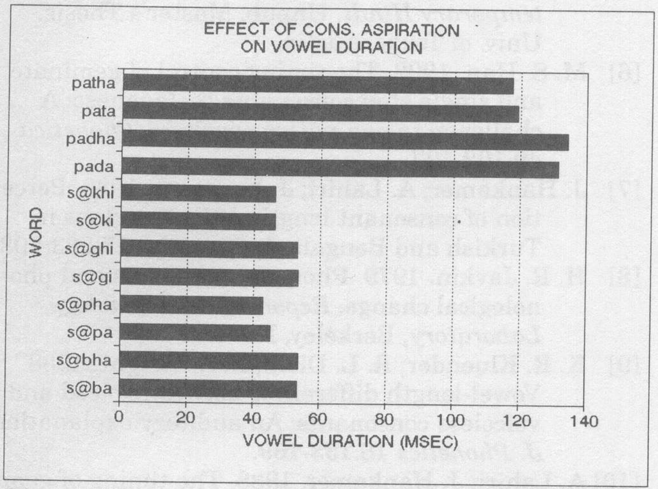

Fig. 3. Mean durations of the first vowel in the indicated minimal pairs which contain medial consonants differing in aspiration or breathy voicing.

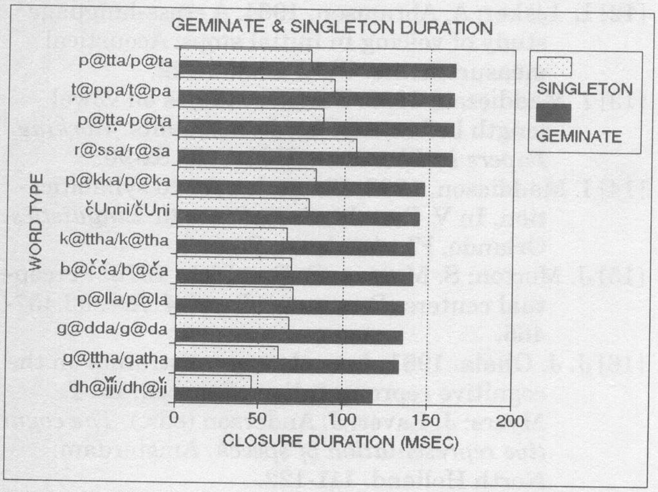

Fig. 4. Mean durations of the medial consonant, geminate or singleton, in the indicated minimal pairs.

Prior phonetic literature testifies to there occasionally being shorter vowel duration before geminates than singletons [10, 13], although the magnitude of this is much less than that of the duration of the consonants, i.e., there is no "compensation" which leads to a relatively constant V + C duration. Fig. 5 shows that Hindi has this pattern, the vowels before geminates being on average 12.4 msec shorter than those before singletons.

Voice Onset Time

Voice Onset Time [12] is a widely used phonetic measure to differentiate classes of obstruents. Hindi makes full use of this continuum for its stops in having the theoretical maximum of three points on it: fully voiced (negative VOT), voiceless unaspirated (zero or slightly positive VOT), and voiceless aspirated (positive VOT typically > 50 msec). A 4th stop type, breathy voiced, uses a distinctive type of voicing in addition to negative VOT. To study VOT we used utterances of the type /CVp/ in the frame /vo ___ aja/, where C was any of the 20 stops (though we omit results for affricates) and the V was /i u a/. In this frame the voiced and breathy voiced stops were fully voiced and so voice onset time, strictly speaking, was not measurable. The overall mean VOT for the remaining two stop types were 19.4 (voiceless unaspirated) and 84.8 msec (voiceless aspirated) in rough agreement with earlier studies [20]. In other languages, the mean values for stops classified as 'voiceless unaspirated' and 'voiceless aspirated' can be quite different [12]. Also, there was considerable systematic contextual variation as shown in Fig. 6. For both stop types, velars show greater VOT than the other places. Within the voiceless unaspirated set, there is a consistent trend for VOT to be greater for the high, close, vowels /i u/ than for /a/. Quite plausibly, these variations as well as those due to consonantal place of articulation can be explained as being due to inherent variations in the impedance these segments offer the exiting airflow. Those articulations with long narrow channels impede the airflow more than others. This delays the achievement of the transglottal pressure difference required for voicing and thus delays VOT [16, 17]. The fact that the retroflex /ṭ/ has the lowest mean VOT is consistent with this: unlike the dental /t/, which involves not only the tongue apex but also much of the tongue blade close the palate, the retroflex at the moment of release has just the tip of the tongue close to the palate, the rest of the tongue body is lowered. Thus the air passage is much shorter for the retroflex. The labial /p/ shows greatest VOT before the vowel /u/ presumably since in addition to the velar constriction there is also a rather long narrow, high impedance channel created by the protrusion of

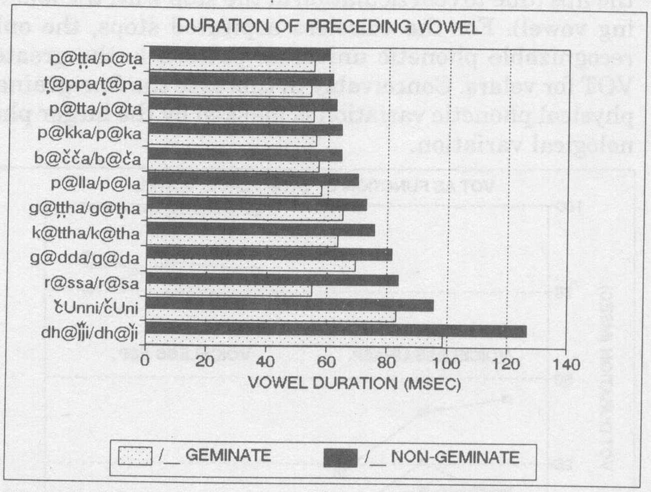

Fig. 5. Mean duration of the vowel preceding medial geminate or singleton consonants in the indicated minimal pairs.

the lips (due to coarticulation of the stop with the following vowel). For the voiceless aspirated stops, the only recognizable phonetic universal pattern is the greater VOT for velars. Conceivably in this case the fine-grained physical phonetic variation is masked by the larger phonological variation.

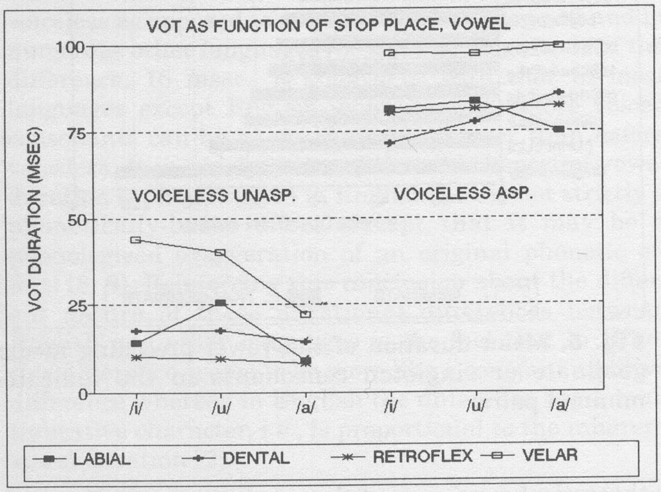

Fig. 6. Mean VOT (Voice Onset Time) of 8 stops types (on the left, 4 voiceless unaspirated, on the right 4 voiceless aspirated) before the vowels /i u a/.

CONCLUSIONS

Some phonetic patterns reviewed above conform closely to posited phonetic universals, e.g., the effect of stop voicing on preceding vowel duration, the effect of vowel height on intrinsic vowel duration. There is especially good agreement in the case of VOT variations in voiceless unaspirated stops as a function of presumed differences in vocal tract aerodynamic impedance. However, in other cases, language-specific differences were evident: the canonical values or ratios in the durations of long vs. short vowels, geminate vs. singleton consonants, VOT values for aspirated vs. unaspirated stops. We must not discount the possibility that an original phonetically 'universal' pattern gets modified via sound change — even ones that do not impact on lexically distinctive variation (phonemic patterns) — and results in language-specific phonetic patterns. The ultimate value of findings on phonetic universals is not to substitute for a detailed inductive discovery of the phonetic details of a given language but rather to *guide* that effort.

ACKNOWLEDGEMENTS

We thank Dr. Anvita Abbi and Mr. V. V. Sahni for their help in the use of facilities at Jawahar Lal Nehru University and Terry Nearey, Terry Baxter, and Tom Wells for help in using the speech analysis systems at the University of Alberta.

REFERENCES

[1] A. S. Abramson. 1987. Word-initial consonant length in Pattani Malay. *Proc. 11th Intl. Congr. Phonet. Sci.* vol. 6.68-70.

[2] M. Caisse. 1988. *Modeling English vowel durations*. Phd. Diss., U. Calif., Berkeley.

[3] M. Chen. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22.129-159.

[4] T. H. Crystal; A. S. House. 1988. The duration of American-English vowels: an overview. *J. of Phonetics* 16.263-284.

[5] R. P. Dixit. 1963. *The segmental phonemes of Contemporary Hindi*. Unpub. Master's Thesis, Univ. of Texas, Austin.

[6] M. S. Han. 1992. The timing control of geminate and single stop consonants in Japanese: A challenge to non-native speakers. *Phonetica* 49.102-127.

[7] J. Hankamer; A. Lahiri; J. Koreman. 1989. Perception of consonant length: Voiceless stops in Turkish and Bengali. *J. Phonetics* 17.283-298.

[8] H. R. Javkin. 1979. Phonetic universals and phonological change. *Report of the Phonology Laboratory*, Berkeley, No. 4.

[9] K. R. Kluender; R. L. Diehl; B. A. Wright. 1988. Vowel-length differences between voiced and voiceless consonants: An auditory explanation. *J. Phonetics* 16.153-169.

[10] A. Lahiri; J. Hankamer. 1988. The timing of geminate consonants. *J. of Phonetics* 16.327-338.

[11] Lehiste, I. 1970. *Suprasegmentals*. Cambridge: MIT Press.

[12] L. Lisker; A. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20.384-422.

[13] I. Maddieson. 1982. Further studies on vowel length before aspirated consonants. *Working Papers in Phonetics* (UCLA) 38.82-90.

[14] I. Maddieson. 1988. Phonetic cues to syllabification. In V. Fromkin (ed), *Phonetic Linguistics*. Orlando, FL: Academic Press.

[15] J. Morton; S. Marcus; C. Frankish. 1976. Perceptual centers (P-centers). *Psychol. Rev.* 93.457-465.

[16] J. J. Ohala. 1981. Articulatory constraints on the cognitive representation of speech. In: T. Myers; J. Laver; J. Anderson (eds.), *The cognitive representation of speech*. Amsterdam: North Holland. 111-122.

[17] J. J. Ohala. 1983. The origin of sound patterns in vocal tract constraints. In: P. F. MacNeilage (ed.), *The production of speech*. New York: Springer-Verlag. 189-216.

[18] M. Ohala. (1983) *Aspects of Hindi phonology*. Delhi: Motilal Banarsidass.

[19] G. E. Peterson; I. Lehiste. 1960. Duration of syllable nuclei in English. *J. Acoust. Soc. Am.* 32.693-703.

[20] K. Shimizu. 1989. A cross-language study of voicing contrasts of stops. *Studia Phonologica* 23.1-12.