THE VOICING DIMENSION: SOME EXPERIMENTS IN COMPARATIVE PHONETICS*

LEIGH LISKER**—ARTHUR S. ABRAMSON***

Speech synthesis has made it possible to produce controlled variations along acoustic dimensions to test their perceptual relevance for phonemic distinctions. Zones of perceptual ambiguity are compared with boundaries between ranges of acoustic values measured in speech. We are interested in doing this in cross-language comparisons. For each of the languages we want to know the number of phonological categories along a dimension used in common. For languages with the same number of categories the boundaries need not be the same.

Our research has led us to believe that in many languages some phoneme categories are distinguished by the timing of glottal adjustments relative to supraglottal articulation, and that this timing relation determines not only the voicing state as narrowly defined, but the degree of aspiration and certain features associated with the so-called force of articulation as well. For word-initial stops in non-whispered speech, this relation is realized acoustically by what we have called voice onset time (VOT), i.e., the interval between the release burst and the onset of laryngeal pulsing.[1] A recent pilot study with synthetic speech demonstrated the distinctive power of VOT.[2] The present study takes a closer look at its perceptual relevance in three languages, English, Spanish and Thai. The resulting data also furnish a basis for discrimination experiments.

We used the Haskins Laboratories parallel resonance synthesizer, which has three

* This research was supported by the National Institute of Child Health and Human Development of the National Institutes of Health and the Information Systems Branch of the Office of Naval Research.

** University of Pennsylvania and Haskins Laboratories, New York, N. Y.

*** University of Connecticut and Haskins Laboratories, New York, N. Y.

[1] L. Lisker and A. S. Abramson, "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements," *Word* 20 (1964), 384—422; L. Lisker and A. S. Abramson, "Stop Categorization and Voice Onset Time", *Proc. 5th Intl. Cong. Phon. Sci.*, E. Zwirner and W. Bethge, eds. (Basel, 1965), pp. 389—391.

[2] A. S. Abramson and L. Lisker, "Voice Onset Time in Stop Consonants: Acoustic Analysis and Synthesis", *Proc. 5th Intl. Cong. on Acoustics*, D. E. Commins, ed. (Liêge, 1965), Vol. Ia, Paper A 51. See also A. M. Liberman, P. C. Delattre, and F. S. Cooper, "Some Cues for the Distinction between Voiced and Voiceless Stops in Initial Position", *Language and Speech* 1 (1958), 153—166.

formant generators with variable frequencies and amplitudes, a choice of buzz or hiss excitation, or a mixture of the two, and control of the overall amplitude and fundamental frequency.

Our basic pattern was built on three steady-state formants for a vowel of the type [a]. Labial, apical and velar stops were made by adding appropriate release bursts and formant transitions to the beginning. We synthesized 37 VOT variants, ranging from 150 msec. before the release to 150 msec. after it. For voicing before the release (voicing lead), we used only low-frequency harmonics of the buzz source. For voice onset after release (voicing lag), the interval between burst and onset of pulsing was excited by hiss alone, with suppression of the first formant to simulate the effects of an open glottis. Calling the release zero time, we gave negative numbers to voicing lead and positive numbers to lag. The stimuli varied in 10 msec. steps, except for the range from —10 to +50, where we made them in 5 msec. steps. Each variant had a fundamental frequency of 114 cps, with a drop toward the end. The 37 variants were recorded in eight random orders, with two occurrences of each on each tape. The test subjects were five native speakers of Latin American Spanish, twelve of American English, and eight of Thai. Using their own orthographies, the subjects identified the stimuli with their stop phonemes.

Figs. 1—3 give the identification curves as functions of VOT values. The bars show frequency distributions of VOT values measured in speech. All the expected categories emerge, but the perceptual crossover zones do not always match very well the zones between the ranges of measured values. In Spanish (Fig. 1), (bdg) are produced with voicing lead, while (ptk) have zero VOT or short lag. The three perceptual crossovers occur to the right of these boundaries, suggesting that some other features, perhaps burst and hiss intensities or formant transitions, were not optimally set. In Fig. 2 the productions of English (bdg) show a small scattering of lead values but a concentration at zero or just after it, while (ptk) all show lag. The boundaries between ranges match the perceptual crossovers well, although there are slight discrepancies in the labials and apicals. For all three places of articulation the English perceptual crossovers have higher VOT values than the Spanish, though the differences are less than expected. Thai was chosen because for two of its places of closure it has three categories, usually called voiced, voiceless unaspirated and voiceless aspirated; the velars have only the latter two. The categories lie in the regions of voicing lead, zero VOT or short lag, and long lag respectively. The match between speech and perception (Fig. 3) is good for the left-hand boundary for the labials and apicals, showing great sensitivity to voicing lead; the right-hand boundary, however, shows the same kind of mismatch as the Spanish and English. For the velars, nevertheless, the match is perfect.

For all three languages the stop categories occupy distinct ranges along the VOT dimension. To be sure, the match between our production and labelling data is somewhat less than perfect, but this is scarcely surprising in view of the severely restricted number of variables involved in the experiment. However, despite the

likelihood that other acoustic features play a role in fixing the category boundaries studied, it seems quite clear that the timing of voice onset is a major factor in de-

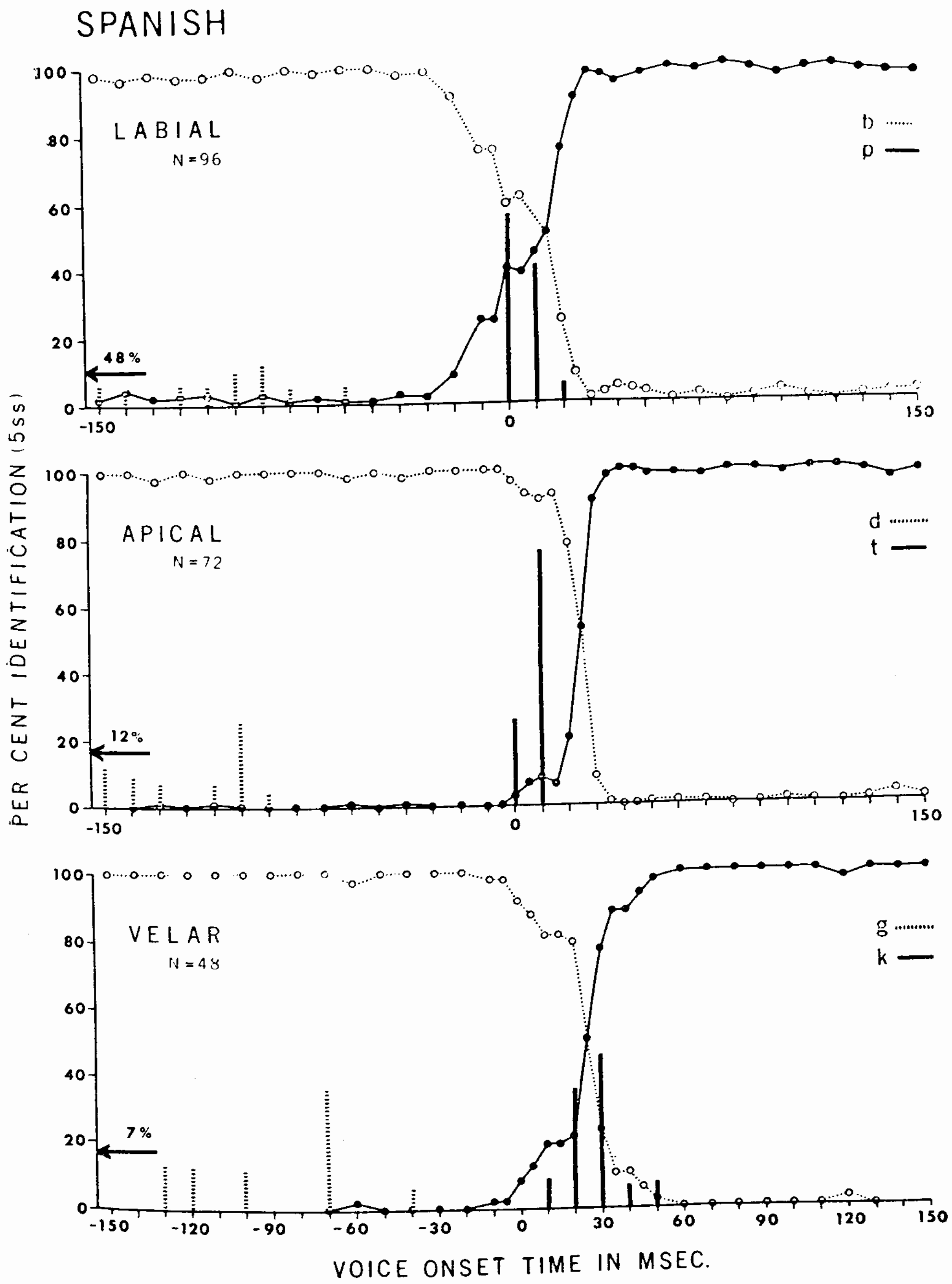

Fig. 1.