

Voice-timing perception in Spanish word-initial stops*

Arthur S. Abramson† and Leigh Lisker‡

Haskins Laboratories, New Haven, Conn. 06510, U.S.A.

Received 10th July 1972

Abstract:

We have shown in earlier work that speakers of Spanish differentiate voiced /bdg/ from voiceless /ptk/ in word-initial position by means of voice onset time (VOT), the temporal relation between the onset of glottal pulsing and acoustic features of supraglottal articulation. More recently, we demonstrated the perceptual efficacy of VOT for the three categories of Thai and the two of English that lie along the timing dimension. The two Spanish categories differ from those of English in that Spanish voiced stops are produced with VOT values that lead the consonant release, while voiceless stops show VOT upon release or immediately thereafter. In the present study, native speakers of Latin American Spanish identified stops synthesized with VOT varying in small steps. Their responses showed a fairly good fit between production and perception, differing from English in the expected direction. Some listeners also discriminated the variants in a psychoacoustic test format. The latter results, along with the earlier ones for English and Thai, suggest that discriminability is largely determined by language experience, although some subjects reveal considerable sensitivity to changes in the acoustic signal at some remove from the phonemic boundary.

In the general phonetic literature it is commonly stated that languages use such phonetic features as voicing, aspiration, glottalization, implosion, "tensity", etc. to distinguish consonants produced at the same supraglottal place of articulation. In previous work we have argued (Lisker & Abramson, 1971) and to some extent demonstrated (Lisker, Abramson, Cooper & Schvey, 1969; Sawashima, Abramson, Cooper & Lisker, 1970) that some of these features are entirely or largely explainable in terms of laryngeal control. Our view has been that the timing of events at the glottis relative to supraglottal articulation provides a simple description of how this laryngeal control is manifested (Abramson & Lisker, 1970a).¹ In our earlier work on this subject (Lisker & Abramson, 1964), we measured voice onset time (VOT) in word-initial stop consonants across a number of languages. VOT, the interval between the release of the stop and the onset of phonation as shown in spectrograms, was the simplest single measure we could find in the acoustic

*This article is a revised version of a paper given at the 83rd Meeting of the Acoustical Society of America, 18-21 April 1972 in Buffalo, New York. The research was supported in large part by a grant from the National Institute of Child Health and Human Development.

†Also, University of Connecticut, Storrs, Conn. 06268.

‡Also, University of Pennsylvania, Philadelphia, Pa. 19104.

¹Recent electromyographic work on laryngeal muscles lends support to this view for English consonants (Hirose & Gay, 1972). A helpful schematic picture of the temporal relations is given by P. Ladefoged (1971:10).

signal of the timing of laryngeal adjustments. The dimension proved efficacious in acoustically differentiating stop consonants in most of the languages with two and even three phonological categories at each place of articulation.²

In the present study we wanted to determine the nature of the relations between VOT as varied in synthetic speech and the labelling and discrimination behavior of Spanish speakers whose two stop categories differ phonetically from the two of English. This is a continuation of studies reported earlier (Abramson & Lisker, 1965, 1970*b*; Lisker & Abramson, 1970).

To control VOT in measured increments we used the Haskins Laboratories formant synthesizer. Our basic pattern was three steady-state formants for a vowel of the type [a]. Labial, apical and velar stop releases were simulated by means of appropriate formant transitions. We synthesized 37 VOT variants ranging from 150 ms before the release to 150 ms 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 release and onset of the periodic source was excited by hiss alone, with suppression of the first formant to simulate the well-known first-formant "cutback" (cf. Liberman, Delattre & Cooper, 1958). Three conditions of VOT for synthetic labial stops are shown in Fig. 1. The 37 VOT variants thus generated were recorded on tape in eight random orders for each place of articulation and played to a total of twelve native speakers of Latin American Spanish who, using Spanish orthography, were to identify the stimuli with their stop phonemes. Instructions were prepared in Spanish and given to the subjects to help insure that they would apply Spanish categories to the stimuli.

The twelve subjects used in the identification experiments were not dialectally homogeneous, coming from Puerto Rico and some six nations of Central and South America. To the best of our knowledge, there is not enough information about phonetic variation in the Spanish dialects of Latin America with regard to the voicing feature to help explain individual differences in our data.³ For our part, we had too small a sampling of subjects from each of the areas represented to make any dialectological statements based on the results of our experiments. The subjects were all more or less bilingual in Spanish and English, having studied English for some years. At the time of starting the experiment,

²A fourth category examined, voiced aspiration, clearly involves glottal adjustments but not of the kind that is discernible on the VOT dimension. Our current electromyographic work with Hajime Hirose, however, does show that this category is distinguished from the others, at least in part, by temporal factors in the contraction of intrinsic muscles of the larynx.

³A search of the literature, with the much appreciated bibliographical help of Gardiner H. London of the University of Connecticut, yields no statement describing instability of voicing in word-initial /b d g/ or, for that matter, unexpected aspiration in /p t k/. This is true of general works (e.g. Lope Blanch, 1968) and descriptions of varieties of Spanish represented in our sampling of test subjects: Argentinean (Malmberg, 1950; Vidal de Battini, 1964), Colombian (Flórez, 1964), Cuban (Lopez, 1971), Mexican (Lope Blanch, 1964; Harris, 1969), and Puerto Rican (Navarro, 1948). These authors call attention only to dialectal differences in the positional and lexical distribution of stop and fricative allophones. Harris (1969:41) affirms, at least for the cultivated speech of Mexico City, that voicing lead is "clearly audible under good acoustical conditions." Lope Blanch (1964:88) comments that in the Yucatan Spanish of Mexico, stops are glottalized because of the Mayan substratum. It is possible, of course, that certain recent trends in pronunciation have not been documented. Malcah Yaeger of the University of Pennsylvania has observed (personal communication) devoicing of /b d g/ in some regional and social dialects.

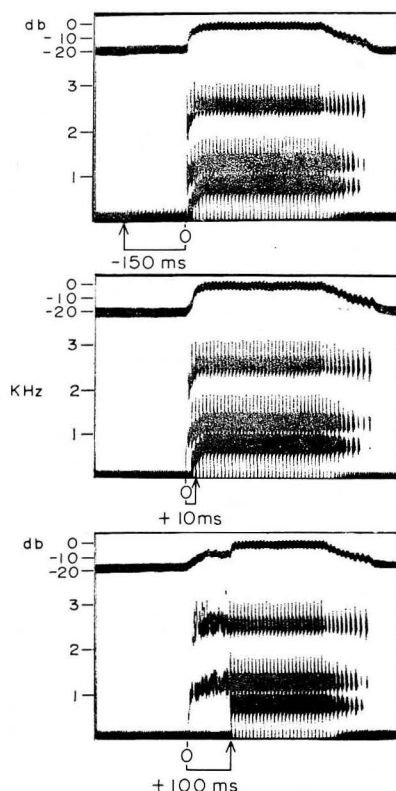


Figure 1

Three conditions of voice onset time in synthetic labial stops. From top to bottom, spectrograms of voicing lead, slight lag, and long lag.

most of them had been in the United States⁴ no more than one year, two of them less than six months, and one for five years. Although they varied considerably in English proficiency, all but one of them showed marked Spanish phonic and syntactic interference in their English. The one exception was an excellent bilingual with a barely detectable Spanish accent and seemingly native English grammar. To help insure against the probability of English interference in the Spanish of our subjects, we chose them with the aid of Spanish language consultants at Queens College of the City University of New York and the University of Connecticut where the tests were run. Our screening of the subjects, done in hiring interviews by our consultants, was perhaps too superficial to rule out entirely the possibility of any phonic interference from their exposure to English, but for the very recently arrived individuals, at least, the likelihood was small.

Figure 2 gives the results of these tests. On the abscissa, negative numbers are assigned to voicing lead and positive numbers to lag, while the moment of stop release is labelled zero. The stimuli varied in 10-ms steps, except for the range of -10 to $+50$, where we made them in 5-ms steps. For each place of articulation, the identification curves are functions of VOT values. The synthetic patterns clearly provided enough cues for two good perceptual categories at each place of articulation. The 50% crossover points are given in

⁴Our single Puerto Rican subject had been on the mainland close to fifteen months. He may well have had much more contact with English than the others, but no effects on his Spanish were discerned.

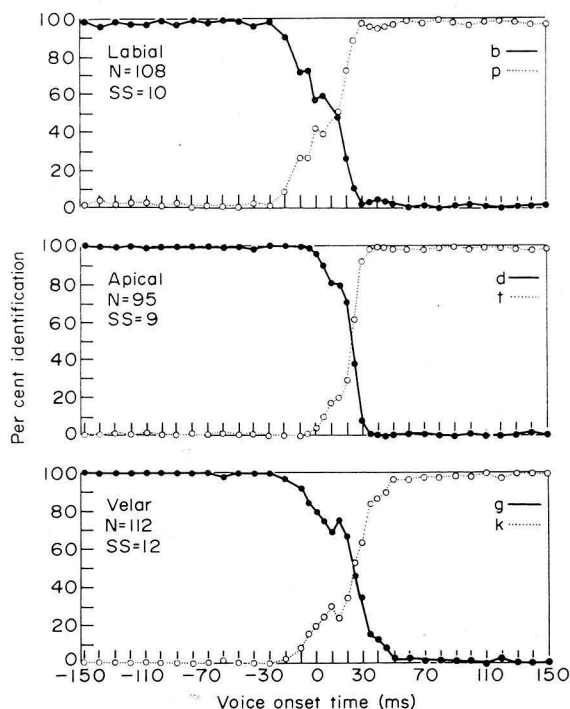


Figure 2 Perceptual identification of VOT variants by native speakers of Spanish. Pooled data.

the table below with the comparable English points, reproduced from our earlier work (Lisker & Abramson, 1970, fig. 2), for comparison.⁵

Spanish and English category boundaries in perception of voice timing (ms)

| | Spanish | English |
|--------|---------|---------|
| Labial | +14 | +25 |
| Apical | +22 | +35 |
| Velar | +24 | +42 |

The Spanish perceptual crossovers have lower VOT values than the English for all places of articulation. This is consistent with the fact that English initial /p t k/ show considerable voicing lag, i.e. aspiration, in stressed syllables while Spanish /p t k/ show little or no voicing lag and are unaspirated; furthermore, Spanish /b d g/ are characterized by voicing lead, i.e. voicing during the occlusion, whereas their English counterparts seem normally to show VOT values of about zero (Lisker & Abramson, 1964: 392, 394).

A long-standing interest in the effects of linguistic experience upon the discriminability of variants along a phonologically relevant continuum (Lieberman, Harris, Hoffman &

⁵The 1970 study also includes VOT identification functions for the three-way voicing distinction of Thai. Perceptual data derived from tests with somewhat similar stimuli have been presented for Dutch (Slis & Cohen, 1969).

Griffith, 1957) has in recent years been investigated across languages (Stevens, Liberman, Studdert-Kennedy & Öhman, 1969). Our own work along these lines, the testing of discriminability of VOT variants in English and Thai (Abramson & Lisker, 1970*b*), has been extended to Spanish in the present study. For this we used 31 of the syllables described earlier, covering the span from -150 to $+150$ ms in steps of 10 ms. We presented these variants in triads as an oddity task. In each triad two stimuli were identical and one was different. The task was to decide whether the odd one was in first, second or third position. The triads were made by pairing stimuli at 2-, 3-, and 4-step intervals along the continuum, thus comparing differences of 20, 30, and 40 ms. Several permutations of the triads and randomizations of the test series were presented to some of the native speakers of Spanish who had taken the identification tests. Very few of the subjects were able to stay with the experiment over a long enough period of time to accumulate a large number of data points for each comparison;⁶ therefore, we have not pooled our data but rather presented them for individual subjects and only for two places of articulation.

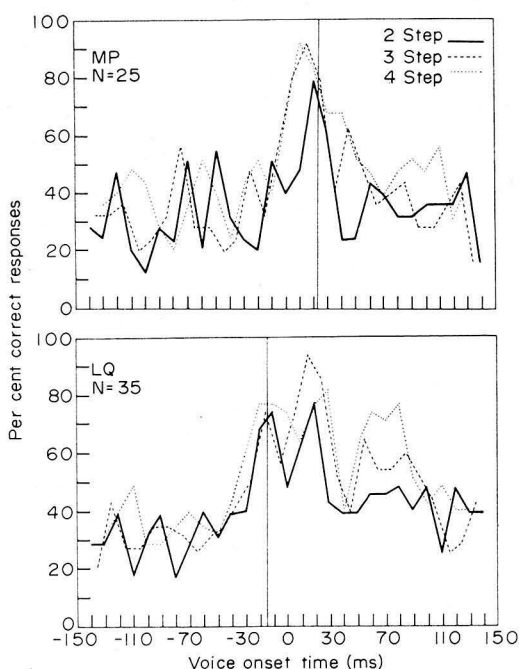


Figure 3

Spanish labial discrimination functions for two individual subjects.

At the top of Fig. 3 we see labial discrimination curves for all three levels of difficulty for Subject MP. Each point on a curve is placed equidistant between the two VOT values being discriminated. The line placed perpendicular to the time axis at $+22$ ms shows MP's 50% perceptual crossover point in the labial identification task. This point is almost

⁶Strange & Halwes (1971) have shown, using our VOT stimuli, that the use of confidence ratings in the oddity task can save much testing time. By the time they had shown this, we were too far along in our Spanish experiments to modify our discrimination procedures. Had we used confidence ratings, we might have been able to salvage another two or three subjects. An important discussion of discrimination procedures in experiments on the perception of speech sounds is found in Pisoni (1971).

precisely at the discrimination peak for all three levels, indicating considerable correlation with the phonological boundary. Note, however, the additional one or two small peaks for higher values of voicing lag.

At the bottom of Fig. 3 we see the discrimination data for LQ. At his 50% crossover point, shown by the vertical line at -15 ms, there is a 4-step discrimination peak of 77%. There are, however, two other large discrimination peaks at +15 ms and +70 ms, and the one at +15 ms is 95%, considerably higher than the peak at the phoneme boundary. Both subjects, then, especially LQ, seem to show effects other than the linguistic.

The discrimination of VOT in velar stops is shown for Subject EL at the top of Fig. 4. His identification crossover at +27 ms is under a discrimination peak that reaches 86%.

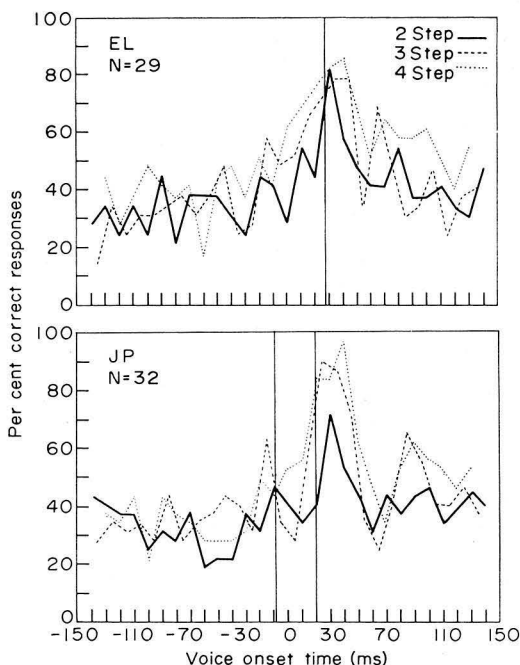


Figure 4 Spanish velar discrimination functions for two individual subjects.

In addition, his voicing lag discrimination is generally quite high with another peak of 70% at about +70 ms. EL, by the way, is the one subject described earlier as an excellent bilingual with hardly any Spanish interference in his English. VOT measurements of his initial /g/ and /k/ in recordings of Spanish words⁷ yield a boundary that corresponds with his identification crossover point. That is, his /g/ and /k/ ranges about at +30 ms. He deviates from previously examined Spanish speakers (Lisker & Abramson, 1964, p. 402) in producing instances of /g/—indeed, 56% of the time—with no voicing lead. His background makes it hard to rule out English interference. He had his elementary and secondary schooling at an American school in Lima, Peru where he studied English for thirteen

⁷We failed to obtain voice recordings of the other subjects. We made a special effort to record EL's speech because of his unusual background. His English stops, recorded in a separate session, look native by and large but seem to show a slight Spanish interference.

years. In addition, he had spent four and a half years in the United States when the experiment began.

The velar data for JP, shown at the bottom of Fig. 4, are somewhat more complicated. His velar identification data do not reveal a single 50% crossover point; rather they show a zone of ambiguity between /g/ and /k/ from -8 ms to +20 ms. We have placed two vertical lines on the time axis to show this span. A discrimination peak reaching 97% straddles the right end of the crossover zone, while a smaller peak straddles the left end; there is a third peak around +90 ms.

The perceptual efficacy of VOT as a sufficient cue for distinguishing the voiced and voiceless stops of Spanish seems established. The possible information-bearing value of other particular acoustic features sometimes associated with voicing distinctions, e.g. pitch (Haggard, Ambler & Callow, 1970; Fujimura, 1971) and F1 transitions (Cooper, Delattre, Liberman, Borst & Gerstman, 1952, p. 600; Stevens & Klatt, 1971), we believe, is also ascribable to the relative timing of events at the larynx and the supraglottal place of articulation. The question of the influence of linguistic categories on the performance of discrimination tasks, at least as far as the present study is concerned, is more complicated. The presence of a phonological boundary certainly has an effect, more with some subjects than others, but there are also discrimination peaks remote from the phonological boundary and indeed always in the lag end of the continuum where spectral variation is somewhat more complex. That is, even though in Spanish and in many other languages the presence or absence of voicing lead is an important cue to a phonological category, stop variants with voicing lag are just easier to discriminate on some psychoacoustic basis. Our earlier work with English and Thai showed similar effects, but they are much more striking here. Of course, there is also the possibility that the psychoacoustic effect is combined with a linguistic one in the sense that large values of lag may sound so aspirated to the Spanish ear that they are considered foreign by the listener and therefore well discriminated from others judged to be more Spanish-like. We have no theoretical rationale for predicting how naive listeners might process a range of speech sounds which lies well outside the norms of their native language, whether they treat these sounds in effect as belonging to a single category of "foreign" speech sounds or as some sort of non-speech continuum.⁸

References

- Abramson, A. S. & Lisker, L. (1965). Voice onset time in stop consonants: acoustic analysis and synthesis. *Proceedings of the Fifth International Congress of Acoustics* (D. E. Commins, Ed.) Liège: Imp. G. Thone.
- Abramson, A. S. & Lisker, L. (1970a). Laryngeal behavior, the speech signal and phonological simplicity. *Actes du X^e Congrès International des Linguistes*, Vol. VI, 123-129. Bucarest: Éditions de l'Académie de la République Socialiste de Roumanie.
- Abramson, A. S. & Lisker, L. (1970b). Discriminability along the voicing continuum: cross-language tests. *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967*, pp. 569-573. Prague: Academia.
- Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M. & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America* 24, 597-606.
- Flórez, L. (1964). El español hablado en Colombia y su atlas lingüístico. *Presente y futuro de la lengua española: Actas de la Asamblea de Filología del I Congreso de Instituciones Hispánicas*, Vol. I, pp. 6-77. Madrid: Ediciones Cultura Hispánica.

⁸For either way of processing these sounds, we do not know what kind of discrimination function the subjects would show. See the discussion in Mattingly *et al.* (1971:152-154).

- Fujimura, O. (1971). Remarks on stop consonants—synthesis experiments and acoustic cues. *Form and substance: phonetic and linguistic papers presented to Eli Fischer-Jørgensen* (L. L. Hammerich, R. Jakobson & E. Zwirner, Eds), pp. 221–232. Copenhagen: Akademisk Forlag.
- Haggard, M., Ambler, S. & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America* **47**, 613–617.
- Harris, J. W. (1969). *Spanish Phonology*. Research Monograph no. 54. Cambridge, Mass.: M.I.T. Press.
- Hirose, H. & Gay, T. (1972). The activity of the intrinsic laryngeal muscles in voicing control: an electromyographic study. *Phonetica* **25**, 129–139.
- Ladefoged, P. (1971). *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.
- Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* **53**, 358–368.
- Liberman, A. M., Delattre, P. C. & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech* **1**, 153–167.
- Lisker, L. & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word* **20**, 384–422.
- Lisker, L. & Abramson, A. S. (1970). The voicing dimension: some experiments in comparative phonetics. *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967*, pp. 563–567. Prague: Academia.
- Lisker, L. & Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language* **47**, 776–785.
- Lisker, L., Abramson, A. S., Cooper, F. S., & Schvey, M. H. (1969). Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America* **45**, 1544–1546.
- Lope Blanch, J. M. (1964). Estado actual del español en México. *Presente y futuro de la lengua española: Actas de la Asamblea de Filología del I Congreso de Instituciones Hispánicas*, Vol. I, 82–91. Madrid: Ediciones Cultura Hispánica.
- Lope Blanch, J. M. (1968). *El español de América*. Madrid: Ediciones Alcala.
- Lopez Morales, Humberto (1971). *Estudios sobre el español de Cuba*. New York: Las Américas.
- Malmberg, B. (1950). *Études sur la phonétique de l'espagnol parlé en Argentine*. Gleeurp, Lund. (Lunds Univ. Årsskrift. N.F.Avd. 1. Bd 45. Nr 7).
- Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology* **2**, 131–157.
- Navarro, T. (1948). *El español en Puerto Rico. Contribución a la geografía lingüística hispanoamericana*. Ed. de la Univ. de Puerto Rico, Río Piedras, P.R.
- Pisoni, D. B. (1971). On the nature of categorical perception of speech sounds. Ph.D. dissertation, University of Michigan. (Also *Supplement to Status Report on Speech Research*, Haskins Laboratories, New Haven, Conn. November 1971).
- Sawashima, M., Abramson, A. S., Cooper, F. S., & Lisker, L. (1970). Observing laryngeal adjustments during running speech by use of a fiberoptics system. *Phonetica* **22**, 193–201.
- Slis, I. H. & Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction, II. *Language and Speech* **12**, 137–155.
- Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M. & Öhman, S. E. G. (1969). Crosslanguage study of vowel perception. *Language and Speech* **12**, 1–23.
- Stevens, K. N. & Klatt, D. H. (1971). The role of formant transitions in the voice-voiceless distinction for stops. *Journal of the Acoustical Society of America* **50**, 146–147 (abstract).
- Strange, W. & Halwes, T. (1971). Confidence ratings in speech perception research: Evaluation of an efficient technique for discrimination testing. *Perception and Psychophysics* **9**, 182–186.
- Vidal de Battini, B. E. (1964). *El español de la Argentina*. Buenos Aires: Consejo Nacional de Educación.