ACTES DU X^e CONGRÈS INTERNATIONAL DES LINGUISTES

BUCAREST, 28 AOÛT – 2 SEPTEMBRE 1967

IV

TIRAGE À PART

1969

ÉDITIONS DE L'ACADÉMIE DE LA RÉPUBLIQUE SOCIALISTE DE ROUMANIE

BUCAREST

LARYNGEAL BEHAVIOR, THE SPEECH SIGNAL AND PHONOLOGICAL SIMPLICITY

Notions of adequacy and simplicity are invoked as criteria for preferring one among a set of proposed grammars, all of which meet certain formal requirements. "Simplicity" seems to mean much the same thing as "economy", for which some kind of counting operation provides the measure. Thus one phonology is simpler than another if it posits a smaller inventory of elementary units, if such units can be grouped into a smaller number of classes, if (where the phoneme is the unit) those units show fewer classes of subphonemic variants, and if it requires a smaller variety of units to specify different realizations of the morphemes. The simplicity criteria may also include one requiring minimum departure from certain general postulates concerning the phonological properties of language, among them one which asserts the existence of a small set of distinctive features adequate for all languages. Except for this last requirement it seems that the data for which a phonology has to account consist exclusively of statements embodying the linguist's phonetic impressions. It is reasonable to ask whether such a phonology is not too limited in scope. Certainly the developments in experimental phonetics over the past twenty years have provided us with data that must not be dismissed as irrelevant to linguistic description. To the extent that phonology as currently practiced fails to account for the physiological, acoustical and perceptual aspects of speech communication, it may be said that its simplicity has been bought at the cost of both descriptive and explanatory adequacy.

Attention will be focussed here on a group of feature dimensions — voicing, aspiration, and force of articulation — said to distinguish, singly or in combination, homorganic stop consonants in a number of languages. The implication of most such descriptions is that these are three independent phonetic dimensions; however, the results of phonetic research argue that in many languages these dimensions do not operate independently. The linguist's desire for the simplest description usually leads him to label some one dimension the distinctive feature for a contrast and to assign a subsidiary or redundant role to other dimensions noted. It is to be understood that his choice of such a feature does not lead the linguist to believe that no other features present contribute to the phonemic distinction in speech communication. When it happens that two or more phonetic features clearly carry phonologically relevant information, it is difficult to decide which one is primary. In such an event it should often be possible to achieve simplicity by isolating an underlying mechanism that will account for all the features observed in the speech signal. For the features under discussion here, voicing, aspiration, and force of articulation, physiological and acoustic data, as well as perceptual experiments, support the view that the timing of events at the glottis provides a simplifying explanation. To support this argument, we shall go through a brief survey of our own research as well as that of some other investigators.

In our work we have tried to test the explanatory power of laryngeal timing for languages with two or more categories of stop consonants said to be distinguished by voicing, aspiration or force of articulation. Voicing is usually taken to mean the presence of laryngeal pulsing during the stop occlusion. This is indeed a useful criterion in at least some environments in many languages. For example, the velar stops in such English words as *bugging/bucking* and such French words as *goût/coup* seem to be so differentiated. Now, having shown this for some of the stops of English, we run into difficulty with other allophones of the class /bdg/. For many speakers there will be no laryngeal pulsing during the labial occlusion of such a word as *bugging*, so that, by the usual definition of "voiced stop", /bdg/in absolute initial position are often quite as voiceless as /ptk/.

In some languages, e.g., Cantonese, it is claimed that aspiration is the unique feature distinguishing homorganic stops and that voicing is not involved. In other languages, aspiration is said to participate in the stop system together with voicing, e.g., Thai and Hindi. In English and German, furthermore, aspiration appears to have still another relation to voicing; sometimes, as in a pair like *repel/rebel*, aspiration accompanies voicing as a significant dimension, and sometimes, as in *come/gum*, aspiration serves where, as we have already stated, voicing is not a reliable mark of the phonemic contrast.

Since by an absolute definition of stop voicing the /bdg/ phonemes of English are not invariably voiced and thereby distinguishable from /ptk/, and because aspiration serves the differentiating function where voicing does not, some linguists prefer to invoke the dimension of articulatory force and to speak of tense or fortis vs. lax or lenis consonants. In their view a difference in the force of articulation is more fundamental than voicing or aspiration because it is to be found in *all* positions of contrast. For some linguists it appears that differences in voicing and aspiration are themselves taken to be manifestations of this underlying difference in tension. A similar situation in German and certain other languages is sometimes handled in the same way. Moreover, assertions of a similar kind are found in the literature for languages where the need for positing such a feature is less compelling, e.g. Thai and Spanish.

It is not easy to find detailed statements on what is meant by the tense/lax feature [1], but the general notion seems to be that variations in tension of the speech musculature bring about changes in voicing condition, air pressure level, and perhaps configuration of the vocal tract. Let us consider several attempts made at isolating such a dimension. On the hypothesis that the rather accessible oris obicularis muscle controlling movements of the lips ought to demonstrate differences in tension for English /b/ and /p/, Harris, Lysaught and Schvey [2] did an electromyographic investigation of English labial consonants. In electromyography (EMG), electrodes attached to the skin permit the recording of changes in electrical potential as the muscles under the electrodes contract. The average of the peak EMG amplitudes of all speakers examined shows some tendency for English /p/ to have a higher level of muscle potential than /b/, but the difference is not striking. Indeed, the authors found so much overlap between the two ranges of values that it was clear that no single occurrence of a stop consonant could be identified as to its phonemic category with any degree of confidence on the basis of this information. More EMG data provided by Fromkin [3] force the same conclusion upon us. Malécot [4] used a wafer transducer to examine the possibility that the two categories of English stops are differentiated by the amount of pressure exerted by the articulators during the occlusion. He did find slight difference in favor of the voiceless stops, but just as in the EMG study, Malécot's examination of the distributions of the values forced him to conclude that mechanical pressure is not a reliable index to the distinction.

Perkell's X-ray studies of the vocal tract [5] showed that when laryngeal pulsing is present during the closure of a voiced stop, the width of the pharynx increases. This suggests, he said, that for a voiced stop the pharyngeal walls are lax, thus allowing for this expansion, while for a voiceless stop they are relatively tense. It is to be understood, of course, that if there is not a pressure drop across the glottis, vibration of the vocal folds will stop; therefore, it is helpful to expand the air chamber above the glottis if voicing is to continue during an articulatory closure. It is hard to see, however, that this intriguing finding vitiates the primacy of

[1] See, e.g., R. Jakobson and M. Halle, *Tenseness and Laxness* in *Roman Jakobson: Selected Writings*, I. *Phonological Studies* (The Hague, 1962), p. 550—555.

[2] K. S. Harris, G. F. Lysaught and M. H. Schvey, *Some Aspects of the Production of Oral and Nasal Labial Stops*, "Language and Speech", 8 (1965), p. 135 —147.

[3] V. Fromkin. *Neuromuscular Specification of Linguistic Units*, "Language and Speech", 9 (1966), p. 170—199.

[4] A. Malécot, *Mechanical Pressure as an Index of 'Force of Articulation'*, "Phonetica", 14 (1966), p. 169—180.

[5] J. S. Perkell, *Cineradiographic Studies of Speech: Implications of a Detailed Analysis of Certain Articulatory Movements* in *Proc. 5th Intl. Cong. on Acoustics*, D.E. Commins, ed. (Liège, 1965), Vol. Ia, Paper A32.

instructions to the glottis. In the absence of precise knowledge of the muscular action involved, one could even speculate that the pharynx is actively widened for the voiced stops, thus characterizing *them* as tense, contrary to tradition. In addition, some durations of pulsing observed during articulatory closure are apparently too long to be explained by any advantage gained through pharyngeal widening. Slight velopharyngeal leakage may be the explanation.

Studies of movements and accumulations of air in the vocal tract in connection with the voicing feature have been undertaken by various experimental phoneticians. In a study done recently by one of us [6], supraglottal air pressure was recorded for English stops in initial, medial, and final position in isolated words. In stressed syllables where differences in peak pressure are usually expected, none is observed. In post-stress position, as in the labials in words like *tepid* and *debit*, an environment usually not mentioned in this connection, the voiceless stop shows considerably higher peak pressure. In the latter case, however, and also wherever differences in the *shape* of the pressure pulse are observed, the pressure trace for the "voiced" member of the pair shows signs of glottal modulation, suggesting greater impedance at the larynx. That is to say, for English at least, differences in the course or peak of supraglottal air pressure do not form a separate phonetic dimension with distinctive value, but rather occur in conjunction with the valvular action of the larynx.

Instead of supporting a tense/lax feature, the results of the studies cited seem to point to the timing of laryngeal action. The most useful way of looking at the vocal folds in running speech so far has been transillumination of the larynx. In our version of this technique a fiber optics bundle leading from a light source is passed through the nose and pharynx to a position above the glottis, and light shining between the vocal folds is picked up by a photo cell placed against the neck just below the thyroid cartilage. Limiting ourselves so far to English consonants that exhibit the contrast with which we are concerned, we have examined the resulting glottograms to see whether the glottis is open or closed and whether the vocal folds are vibrating or not. Much remains to be done, but our tentative findings make sense of most of the acoustic observations and air pressure studies too. In general, English /ptk/ are produced with either opening of the glottis or interruption of pulsing, or both. The stops /bdg/ occur without interruption of pulsing in medial position; however, at the beginning or end of an utterance where the closure phase of the stop may well show no laryngeal pulsing at all, the glottis is nevertheless likely to be closed as if for phonation. In addition, then, to the obvious feature of pulsing, such differences in intensity of stop release (burst) or air pressure level as are observed, seem to be well correlated with glottal adjustments. Aspiration in certain voiceless stops

---

[6] L. Lisker *Supraglottal Air Pressure in the Production of English Stops* "in Language and Speech" (In press).

and /h/ can be accounted for by a long enough period of glottal opening to allow for turbulent excitation of the vocal tract.

A few years ago we began to think that the most convenient index to the timing of laryngeal closure relative to articulatory events was simply voice pulsing, although this need not be taken as implying that phonation must start immediately upon closure of the glottis. One can find the onset and offset of this pulsing in spectrograms and oscillograms with a minimum of error. Labelling the dimension "voice onset time" (VOT), we proceeded to examine word-initial stops of eleven languages that seemed typical in this respect of large numbers of languages [7]. These languages are usually said to use voicing, aspiration or force of articulation, or some combination of them, to distinguish two, three or even four categories of stops. Six of the languages had at most two stop categories for a given place of closure; these are Dutch, Spanish, Hungarian, Tamil, Cantonese and English. Three others have three categories each: Eastern Armenian, Thai and Korean. The remaining two, Hindi and Marathi, have four categories each. Three cross-language categories emerge from the data: (1) voicing lead, in which laryngeal pulsing begins somewhat before the release of the initial stop, (2) pulsing beginning at or just after the release, (3) voicing lag, in which pulsing begins somewhat after the release of the stop. In general, the stop categories of each language, no matter how they are described in the literature, are effectively separated by the VOT feature. The only categories clearly not distinguishable on this basis are the so-called voiced aspirates and voiced inaspirates of Hindi and Marathi; to distinguish these categories, it is apparently necessary to invoke another feature of laryngeal behavior, that of a partly opened glottis that allows turbulent air through to accompany the pulsing. As expected, the languages examined vary in the number of categories and in the placement of their boundaries. It is interesting that even in the case of a so-called hole in the pattern, as in the existence of only one velar stop instead of two in Dutch and only two velar stops instead of three in Thai, the phonetic space appropriate to the missing category remains uninvaded by values of VOT from the adjacent categories in the system.

Having demonstrated the considerable analytic power of the VOT dimension, we wished to test the perceptual relevance of the effects of laryngeal timing. We used a speech synthesizer to make equally spaced stop variants along the dimension from a condition of long voicing lead through the moment of release and on to a condition of long voicing lag. For all lag variants the first formant was suppressed [8], as often happens in natural speech when the glottis is open, and the upper formants were

---

[7] L. Lisker and A. S. Abramson, *A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements*, "Word", 20 (1964). p. 384–422.

[8] 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), p. 153–166.

excited by noise to simulate aspiration. In an early study[9] apical stops generated in this fashion were played to native speakers of English, Spanish, and Thai. Aside from certain detailed observations that cannot be discussed here, the main conclusion to be drawn from the data is that speakers of all three languages can sort the synthetic stimuli into categories appropriate to their languages, thus validating the perceptual importance of the physical features manipulated. New experiments for all three major places of articulation have yielded essentially the same results[10], and we have extended our research to the question of the influence of phoneme boundaries on the ability of speakers of various languages to discriminate variants along the VOT dimension[11].

In conclusion, we can say that for the consonant distinctions under discussion, the timing of events at the glottis relative to supraglottal events provides a simplifying explanation for the phonetic features observed. This is generally true even though in certain cases, e.g., one of the Korean boundaries[12], it may unexpectedly fail to be entirely satisfactory and suggest the need of another feature. Presumably other phonological units requiring complex phonetic specification may be referred to relatively simple underlying mechanisms on the basis of detailed matching of the speech signal to the anatomy and physiology of the vocal tract. For such an approach to phonology, however, the linguist must be more willing than he typically seems to be to have recourse to experimental data[13].

DISCUSSIONS

R. Gsell :

La communication de MM Abramson et Lisker montre dans quelle mesure les oppositions consonantiques : occlusive sonore ∼ occlusive sourde ∼ occlusive sourde aspirée, réalisées généralement par un ensemble complexe de différences dans la substance phonique, peuvent s'expliquer à l'aide d'un seul correlat distinctif : la durée de la sonorité et sa situation par rapport au début de l'explosion de la consonne. Le nouveau correlat : « position de la sonorité » permettra certainement de réaliser une grande économie dans la description phonologique des systèmes con-

[9] A. S. Abramson and L. Lisker, *Voice Onset Time in Stop Consonants: Acoustic Analysis and Sythesis* in *Proc. 5th Intl. Cong. on Acoustics*, (Liège, 1965), Vol. Ia, Paper A51.

[10] L. Lisker and A. S. Abramson, *The Voicing Dimension: Some Experiments in Comparative Phonetics* in *Proc. 6th Intl. Cong. Phon. Sci.*, 7—13 Sept. 1967, (Prague, 1970), p. 563—567.

[11] A. S. Abramson and L. Lisker, *Discriminability along the Voicing Continuum: Cross-Language Tests* in *Proc. 6th Intl. Cong. Phon. Sci.*, (Prague, 1970), p. 569—573.

[12] C-W. Kim, *On the Autonomy of the Tensity Feature in Stop Classification (with Special Reference to Korean Stops)*, "Word", 21 (1965), p. 339—359.

[13] We wish to acknowledge support for this research 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.

sonantiques et fera faire un progrès important à la théorie des « distinctive features ». J'aimerais cependant demander aux auteurs :

1. Si les valeurs indiquées sur les schémas sont des valeurs moyennes correspondant à une ou plusieurs langues, ou simplement des échantillons.

2. Si les valeurs du « retard de sonorité » sont stables pour les occlusives des différents lieux d'articulation d'une langue déterminée.

3. Si l'on a pu fixer à l'aide de la synthèse des seuils de discrimination et s'il existe des zones de recoupement.

Interventions de J. GODFREY, J. W. BLACK, A. TĂTARU

A. S. ABRAMSON :

1. The three spectrograms shown with the oral version of our paper were samples of Thai speech. Their values of voice onset time are not averages, but they are good examples of the three modes found in our data.

2. Ranges of voicing lag for the different places of articulation show extensive overlap; nevertheless, dorsal stops tend to have the longest lags.

3. We have used synthetic speech to establish discrimination thresholds for speakers of English and Thai (see footnote 11). We expect soon to have similar data for Spanish. The general result is that the discrimination peaks match the perceptual boundaries between phonemes, where the phonemes occupy relatively distinct ranges along the continuum of voice onset time.