

The global character of phonetic gestures

Ignatius G. Mattingly

Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, and Department of Linguistics, University of Connecticut, 341 Mansfield Road, Storrs, CT 06269-1145, U.S.A.

1. Introduction

Articulatory phonology, which takes phonetic gestures as the primitive units of phonological representation, is a development very welcome to those of us who favor the motor theory, according to which phonetic gestures are the primitive units of speech perception (Liberman & Mattingly, 1985). But it is essential to both enterprises that the notion of a phonetic gesture be rich enough to play the theoretical role required of it, yet remain grounded in experimental-phonetic observation.

The term *gesture* is sometimes used by phoneticians simply to mean an observable movement of some articulator. But this usage is hardly satisfactory, if the term is to refer to a primitive unit in phonology or in the psychology of speech production and perception. The movement of an articulator may result from aerodynamic conditions rather than from active motor control, as when the vocal folds vibrate in phonation. Two phonologically equivalent movements may differ physically because of such non-linguistic factors as speaking rate and vocal effort. Movements of two or more articulators may constitute what we would prefer, on phonological grounds, to call a single gesture, as when upper and lower lip collaborate in lip protrusion. Conversely, the movement of a single articulator may consist of components attributable to phonologically distinct gestures, as when the tongue body participates simultaneously in the formation of a vowel and an alveolar consonant. These obvious considerations are enough to remind us that a phonetic gesture must necessarily be something of an abstraction, and that a gestural transcription cannot simply be read off from articulator movement data.¹

2. Defining phonetic gestures

It is therefore not a trivial problem to decide what the set of gestures is, and what movements or movement components are to be assigned to each gesture. Our phonological intuitions, of course, suggest certain solutions and discourage others, as do such traditional articulatory-phonetic terms as "lip rounding", "tongue

¹ In their carefully considered account of gestural patterning, Saltzman & Munhall (1989) use the term *gesture* "to denote a member of a family of functionally equivalent articulatory movement patterns that are actively controlled with reference to a given speech-relevant goal (e.g., a bilabial closure)" (334n.); they use *gestural unit* or *gestural primitive* to refer to the basic action unit that underlies the family of movement patterns. But Browman & Goldstein (1990), Liberman & Mattingly (1985) and most linguists use *gesture* to mean both these things, trusting that context will disambiguate, and I continue this dubious usage here.

raising", and so on. But, obviously, the guidance of neither of these sources is to be accepted uncritically.

In their formulation of articulatory phonology, Browman & Goldstein (1990) take two decisions, justifiable on other grounds, that make the job of defining the gestures easier. First, they stipulate that gestures are to be "dynamically defined". This means that every gesture must be describable as a "coordinative structure" (Turvey, 1977) performing a well-defined task, and that its observed movement trajectories can be modeled according to an abstract, functionally-defined dynamical model, such as task dynamics (Saltzman, 1986; Saltzman & Kelso, 1987; Saltzman & Munhall, 1989). The effect of this decision is to reduce the degrees of freedom. An infinite number of conceivable movement trajectories are ruled out because such a model could not have produced them. However, "coordinative structure" is still a pretty broad concept. Though all phonetic gestures are coordinative structures, not all phonetic coordinative structures are gestures. A coordinative structure may be reducible to two or more simpler coordinative structures which themselves may be further reducible. But I take it that, although a non-primitive coordinative structure (what Browman & Goldstein call a *constellation*) may constitute a higher-level phonetic unit, such as a phone, syllable, or breath group, only a *primitive* (i.e., not further reducible) coordinative structure can be a gesture. Therefore, properties commonly regarded as marking a coordinative structure, such as a systematic phasing relation between movements, are necessary, but probably not sufficient properties of a gesture. The systematic phasing relation that holds between oral closure and glottal opening in a voiceless stop (Kelso, Saltzman, & Tuller, 1986) does not mean that these movements are to be considered part of a single gesture; it is more likely, as Browman and Goldstein suggest, that they form a constellation. The question whether a candidate gesture is really primitive has to be resolved on other grounds.²

Browman and Goldstein's second decision is to follow task dynamics in requiring that the tasks performed by the gestures be defined in terms not of articulator positions but of vocal-tract variables, specifically, location of constriction and degree of constriction. This requirement means that a group of articulator movements must be considered as part of one gesture if they all contribute to the formation of a particular constriction, and it fits well with the fact that articulators that are paired (the lips) or anatomically dependent (the tongue with respect to the jaw) naturally collaborate. Thus, in the production of a bilabial consonant, upper lip, lower lip, and jaw movements are part of one coordinative structure whose task is to make a labial constriction. But more significantly, the implication is that (contrary to what the term "gesture" might perhaps suggest) articulator movements as such are not the primary goals of speech production. Rather, the movements constituting a gesture are the means of producing a particular temporary deformation of the vocal tract.

² It has been suggested to me that, in accepting an action-theoretic view of speech gestures, I am undermining the position taken by Liberman & Mattingly (1985) that phonetic gestures are "special", being produced and perceived by a language module (Fodor, 1983). But I see no difficulty here. Phonetic gestures, though special, are not magical. They must inevitably conform to whatever general physical and physiological principles govern animal movements. Their specialness is manifested in the tasks they perform.

3. Some proposals

These two decisions certainly take the enterprise of articulatory phonology a considerable way. However, it is possible and desirable to go further. I want to offer here three proposals that will make phonetic gestures still more plausible as phonological units and at the same time more phonetically realistic.

3.1. *Globality of phonetic gestures*

The first proposal is that a gesture be allowed to have a global effect on the vocal tract. In articulatory phonology, as currently formulated, the effect of a gesture is purely local: it results in a single constriction at a particular location. In this respect, articulatory phonology is close to traditional phonetic descriptions in which each sound is assigned a particular place of articulation. But such traditional descriptions overidealize the facts of speech production. It is convenient to label a phonetic gesture according to the properties that make it distinctive, e.g., "bilabial plosive", but it is then easy to forget that some other important things may be happening that, though not implied by this label, are essential to the performance of the phonetic task. This is all right for systems of phonology which are comfortably remote from physical detail, but not for articulatory phonology, which claims to "hug the phonetic ground closely" (Hockett, 1955, p. 174).

Consider, for example, that in the phonetic gesture traditionally known as "lip rounding", not only are the lips protruded, but also the larynx is lowered.³ These two movements are surely part of a single gesture which serves to lengthen the vocal tract (Heffner, 1964; Perkell, 1969). This task is obviously global rather than local, despite the fact that it is accomplished by two articulator movements, each of which has a local effect. It would be possible to posit two separate gestures, as current articulatory phonology would seem to require. But then the tasks performed by these gestures would have no phonological motivation in themselves, unless it could be shown that for some language, "larynx lowering" and "lip protrusion" function independently.

As a second example, consider "vowel opening". In the production of an open vowel, the tongue body moves lower, assisted by the jaw. But so does the lower lip, assisted by the same jaw movement, while the upper lip is raised (see, e.g. the photographs in Jones, 1962). These events could, perhaps, be regarded as two gestures one by the tongue and jaw and one by the lips and jaw, each adjusting the width of a local constriction. But surely there is just one task, namely opening up the whole oral cavity, to which movements of all these articulators contribute. If this is conceded, then all these movements must constitute a single vowel-opening gesture, whose effect on the vocal tract is global rather than local.

In fact, vowel opening is even more global than has just been suggested. The account of open-vowel production just given neglects the role of the velum, which lowers somewhat, even for vowels traditionally classified as "non-nasalized" rather than "nasalized", so as to couple the oral and nasal cavities (Henderson, 1984; Krakow, 1989). Thus, in these vowels, the oral cavity is open at the nasal port as

³ It should perhaps be made explicit that in this and other examples, it is the canonical form of the gesture that is being described. Thus the actual amount of larynx lowering and lip protrusion that will be observed in particular instances of lip rounding will depend on the initial position of these articulators, on the demands of other gestures, and on various non-linguistic factors.

well as at the lips and at the entrance to the pharynx. It would seem reasonable, therefore, to include this velum-lowering movement in the vowel-opening gesture as well as the tongue, lip and jaw movements.

Moreover, the velum lowering observed for open vowels is an instance of a more general phonetic pattern in which the position of the velum, and hence the size of the nasal port, is specified for "non-nasal" sounds, being correlated with degree of closeness. Thus, the velum is higher for close vowels and approximants and higher still, so that the nasal port is effectively closed, for obstruents. Just as a velum-lowering movement is part of vowel opening, appropriate velum movements are part of the global gestures for these other sounds. All such velic adjustments are, of course, a quite different matter from distinctive velum lowering in nasal consonants and nasalized vowels, obviously to be regarded as a separate and independent gesture. Thus, it has been observed that the displacement of the velum is greater for low than for high *nasalized* vowels, just as it is greater for low than for high *oral* vowels (Henderson, 1984). In the nasalized vowels, both velum lowering and vocalic gestures that include velic adjustments are active.

The examples that have been considered suggest that the participation of more than one articulator in a gesture is the rule, nor is it limited just to instances in which articulators are paired or anatomically dependent. It is tempting to go a step further. In current articulatory phonology, any articulator not specified in the definition of an active gesture is taken to be in its schwa position. But in a revised, more global articulatory phonology, it should, perhaps, be presumed that every articulator has an explicitly specified role in every gesture. In many cases, of course, the specification would have to be vacuous, merely insuring that the articulator in question keep out of the way. But this presumption would have three advantages: First, it would obviate the need to define a particular, arbitrary "neutral" position for each articulator, in the absence of good reasons to believe that such neutral positions exist. Second, it would discourage the overidealization of the gestures. And finally, though it might seem to complicate description of individual gestures, this presumption would be likely to permit simpler description of their temporal patterning.

Some cases have recently been noted in which an articulator movement, conventionally interpreted as an anticipation of a following "segment" and to be explained by "feature spreading" (Henke, 1966), must actually be attributed to a current segment, even though the movement would not have been predicted from the standard phonological description of the current segment. Thus, some lip protrusion is observed in the production of an apical obstruent, even when there is no rounded vowel in the utterance (Boyce, 1988; Gelfer, Bell-Berti & Harris, 1989). The lip protrusion movement must thus be assigned to the obstruent (though its contribution to this gesture needs to be clarified). Once the movements are properly assigned, coarticulation can be seen as a relatively simple overlapping of periods of gestural activation, rather than a mysterious migration of features from one segment to another (Boyce, Krakow, Bell-Berti & Gelfer, 1990), and the very notion of a "segment" becomes superfluous.

3.2. *Aerodynamic properties of gestures*

The second proposal, really an extension of the notion that phonetic gestures are global, is that the aerodynamic as well as the geometric properties of the vocal tract

be considered relevant to the definition of phonetic tasks, as suggested by Saltzman & Munhall (1989). Browman & Goldstein do not ignore the aerodynamic aspects of speech production, but they apparently regard them as merely derivative from the geometric aspects (see the account of "tube" phonology in Browman & Goldstein, 1989).

But aerodynamic state should be considered part of the gestural goal. As has been mentioned, an oral articulator movement and a velic movement both contribute to the narrowing or closing of the vocal tract that is characteristic of obstruents. But the functions of the narrowing and of the closing are aerodynamic, and velum-raising plays an essential role. For fricatives, high particle velocity is required to produce turbulence and the area of primary constriction is therefore narrow, but if the nasal port were not closed, volume velocity and therefore particle velocity would be lower. For stops, the oral channel is temporarily blocked to enable a rise in oral pressure; this rise in pressure could not be achieved if the nasal port were left open (Catford, 1977). If the specification of these tasks ignored aerodynamics, the raising of the velum would be difficult to interpret.

A further instance in which aerodynamic considerations are crucial to task definition is found in the case of ejective stops. In these sounds, a glottalic and an oral closure are formed (no doubt the velum is raised, too). The larynx is jerked up, so that airflow is rapidly increased from zero, and intraoral pressure rises abruptly. The oral closure is then released, reducing the pressure again (Catford, 1977). An ejective appears to be a constellation of two gestures: one is an oral closure gesture, just as with a pulmonic stop; the other is an initiatory gesture that includes the glottal closure and the larynx raising. In the latter, the glottal and the laryngeal movements are equally essential to initiation of airflow, but would have seemed to be a purely arbitrary conjunction if this aerodynamic function had been overlooked.

Aerodynamics is also relevant to the idea of manner. Browman & Goldstein rightly wish to retain this fundamental phonological concept in their system, and they suggest that the various manner classes may be distinguishable by different values of the dynamic variable "stiffness". While there is something in this, surely aerodynamic properties (to which, of course, variable degrees of stiffness contribute) are more basic to manner distinctions. What plosives have in common is that airflow is temporarily blocked; what fricatives have in common is that particle velocity is increased so as to produce turbulence, and so on. It is the aerodynamic goals of the gestures that distinguish the manner classes.

3.3. *Functional organization of the gestures*

The third proposal is that, given their essentially global character, the gestures should be organized in the phonology functionally rather than anatomically. Browman & Goldstein, quite understandably, wish to incorporate in articulatory phonology some of the advantages of autosegmental phonology (Goldsmith, 1976) and CV phonology (Clements & Keyser, 1983), in which there are several layers of phonological structure, corresponding, e.g., to oral articulation, tone, and nasality. (This multi-layered organization is proposed to account for various inconsistencies of temporal pattern that embarrass earlier phonologies which assumed only one layer.) In their 1987 paper, therefore, Browman & Goldstein define an organization of the "gestural score" into "articulatory tiers":

... the tiers are defined using the notion of articulatory independence. Velic gestures are obviously the most independent, since they share no articulators with other gestures. In this limiting case, velic gestures constitute a wholly separate nasal (velic) subsystem and hence are represented on a separate velic tier. Glottal gestures also participate in an independent subsystem (although other laryngeal gestures, for example, for larynx height, would also participate in this subsystem), and hence are represented on a separate glottal tier. The oral gestures form a third subsystem, with the jaw as a common articulator. Since the oral gestures are distinguished by different combinations of articulators, oral gestures are represented on three distinct oral tiers, one for the lips, one for the tongue body, and one for the tongue tip . . . these articulatory tiers correspond closely to organizations posited by both phoneticians and autosegmental phonologists. The three oral tiers of lips, tongue tip, and tongue body correspond to the traditional grouping of places of articulation into three major sets: labial, lingual, and dorsal . . . (Browman & Goldstein, 1987, p. 4)

This organization of tiers is purely anatomical (note especially the assignment of larynx height gestures to the glottal subsystem), and so runs counter to the essential globality of phonetic gestures that has been stressed above. Thus, we have already seen that the velum is not always independent: various "oral" gestures in fact have velic components. Similarly "labial" gestures may have glottal components and "dorsal" gestures may have labial components. Consider now one more example, in which articulators that, according to current articulatory phonology, are assigned to separate tiers, nevertheless unite in a common task, control of airflow during phonation. It would seem reasonable to regard phonation itself as a constellation of several gestures: perhaps one that initiates and maintains airflow (and is required also for unvoiced sounds), one that approximates the vocal folds, and one that adjusts the tension of the folds. The airflow is primarily the result of pulmonic movements (although Browman & Goldstein do not say so, presumably the lungs constitute a separate tier in current articulatory phonology), but there are other maneuvers by articulators assigned by Browman & Goldstein to other tiers that help to maintain airflow through the glottis during a voiced stop, by reducing supraglottal pressure. Thus, the cheeks may relax and expand or the larynx may be lowered (Westbury, 1983).

What is called for is a functional rather than an anatomical organization, one in which the gestures are grouped according to the nature of their tasks. Thus there might be an airflow tier, a voicing tier, a subsystem of tract-shaping tiers, a nasality tier, and so on. The phenomena that have led phonologists to postulate different layers would then be seen as resulting naturally from the asynchronous execution of these different functions. An articulator would not be assigned to a particular subsystem, but might participate in gestures that belong to different subsystems. Thus, the larynx would participate in the airflow tier in ejectives and the cavity-shaping subsystem in lip rounding; the vocal folds, in the airflow tier in ejectives as well as in the phonation tier; the velum, both in cavity-shaping and in nasalization. Though the motivation would be phonological rather than anthropo-phonetic, the arrangement would be quite similar to the taxonomy of gestures presented by Catford (1977).

Articulatory phonology is a fairly radical proposal, as comparison with other

approaches to phonetic representation described in this issue will make obvious. My comments should be taken as criticisms from the Left, intended to make it still more radical. I have tried to suggest some respects in which the notion of the phonetic gesture in articulatory phonology could be deepened and extended. Gestures are essentially global; they have aerodynamic as well as geometrical goals; they are functionally organized. It will not escape notice that in the course of the argument, the tasks performed by gestures have become more specifically phonetic. Considered from a non-phonetic point of view, these tasks would seem idiosyncratic, bizarre, quite unlike the nonlinguistic tasks performed by the same organs. From the perspective of the motor theory of speech perception, this is just as one might expect.

I thank Catherine Browman and Elliot Saltzman, who made very helpful comments on an earlier draft of this paper. Support from NIH grant HD-01994 to Haskins Laboratories is gratefully acknowledged.

References

- Boyce, S. E. (1988) The influence of phonological structure on articulatory organization in Turkish and in English: vowel harmony and coarticulation. Unpublished Ph.D. thesis, Department of Linguistics, Yale University.
- Boyce, S. E., Krakow, R. A., Bell-Berti, F. & Gelfer, C. E. (1990) Converging sources of evidence for dissecting articulatory movements into core gestures, *Journal of Phonetics*, **18**, 173–188.
- Browman, C. & Goldstein, L. (1987) Tiers in articulatory phonology, with some implications for casual speech, *Haskins Laboratories Status Report on Speech Research*, SR-92, 1–30. Also in *Papers in laboratory phonology I: Between the grammar and the physics of speech* (J. Kingston & M. E. Beckman, editors), pp. 341–376. Cambridge: Cambridge University Press (1989).
- Browman, C. & Goldstein, L. (1989) Articulatory gestures as phonological units, *Phonology*, **6**, 201–251.
- Browman, C. & Goldstein, L. (1990) Gestural specification using dynamically-defined articulatory gestures, *Journal of Phonetics*, **18**, 299–320.
- Catford, J. C. (1977) *Fundamental problems in phonetics*. Bloomington, IN: Indiana University Press.
- Clements, G. N. & Keyser, S. J. (1983) *CV phonology: a generative theory of the syllable*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1983) *The modularity of mind*. Cambridge, MA: MIT Press.
- Gelfer, C. E., Bell-Berti, F. & Harris, K. S. (1989) Determining the extent of coarticulation: effects of experimental design, *Journal of the Acoustical Society of America*, **86**, 2443–2445.
- Goldsmith, J. A. (1976) *Autosegmental phonology*. Bloomington, IN: Indiana University Linguistics Club.
- Heffner, R.-M. S. (1964) *General phonetics*. Madison, WI: University of Wisconsin Press.
- Henderson, J. B. (1984) Velopharyngeal function in oral and nasal vowels: A cross-language study. Unpublished Ph.D. thesis, Department of Linguistics, University of Connecticut.
- Henke, W. L. (1966) Dynamic articulatory model of speech production using computer simulation. Unpublished Ph.D. thesis, Department of Electrical Engineering and Computer Science, MIT.
- Hockett, C. F. (1955) *A manual of phonology*. Memoir 11 of the *International Journal of American Linguistics*. Baltimore: Waverly Press.
- Jones, D. (1962) *An outline of English phonetics*, 9th ed. Cambridge: Heffer.
- Kelso, J. A. S., Saltzman, E. L. & Tuller, B. (1986) The dynamical perspective on speech production: Data and theory, *Journal of Phonetics*, **14**, 29–59.
- Krakow, R. (1989) The articulatory organization of syllables: a kinematic analysis of labial and velar gestures. Unpublished Ph.D. thesis. Department of linguistics, Yale University.
- Lieberman, A. M., & Mattingly, I. G. (1985) The motor theory of speech perception revised, *Cognition*, **21**, 1–36.
- Perkell, J. S. (1969) *Physiology of speech production: results and implications of a quantitative cineradiographic study*. Cambridge, MA: MIT Press.
- Saltzman, E. L. (1986) Task dynamic coordination of the speech articulators: a preliminary model. In *Generation and modulation of action patterns* (H. Heuer & C. Fromm, editors), *Experimental Brain Research Series* 15, pp. 129–144. New York: Springer-Verlag.

- Saltzman, E. L. & Kelso, J. A. S. (1987) Skilled actions: a task dynamic approach. *Psychological Review*, **94**, 84–106.
- Saltzman, E. L. & Munhall, K. G. (1989) A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, **1**, 333–382.
- Turvey, M. T. (1977) Preliminaries to a theory of action with reference to vision. In *Perceiving, acting, and knowing* (R. Shaw & J. Bransford, editors), pp. 311–265. Hillsdale, NJ: Lawrence Erlbaum.
- Westbury, J. (1983) Enlargement of the supraglottal cavity and its relation to stop consonant voicing, *Journal of the Acoustical Society of America*, **73**, 1322–1336.