Journal of Phonetics (1989) **17**, 91–97

On the domain of the quantal theory

Louis Goldstein

Department of Linguistics, Yale University, Box 1504A, Yale Station, New Haven, CT 06520, U.S.A. and Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, U.S.A.

Stevens' quantal theory is examined from the point of view of accounting for the types of gestural structures that are employed phonologically in languages. It is argued that the domain of the theory is both wider and narrower than Stevens proposes. It is wider in that it could be extended to explain, not only *what* gestures languages may employ contrastively, but also *how* those gestures may be organized. It is narrower in that quantal considerations do not seem to actively constrain the precision of talkers in speech production.

Stevens' quantal theory (QT) demonstrates how the acoustic properties of the vocal tract and certain characteristics of the human auditory system could give rise to a set of stable regions within the multidimensional articulatory space. This stability is seen in the fact that the auditory properties associated with such regions are (1) relatively insensitive to small articulatory perturbations along some dimension(s), and (2) qualitatively distinct from the auditory properties of other regions. For Stevens, then, such regions form the basis of possible phonological contrasts in languages, and each is taken to be the basis of some distinctive feature (Jakobson, Fant & Halle, 1969). These features are the basic "atoms" out of which the phonological representations are constructed.

The basic QT is thus an attempt to explain why it is that phonetic contrasts in languages are to be found between particular articulatory regions. The theory does not make any specific claims about how these stable regions and their auditory properties (features) are to be embedded in a theory of either speech production or perception, although such extensions have been addressed in other related work (e.g., Perkell & Nelson, 1982; Stevens & Blumstein, 1981). In this paper, we examine the role of the articulatory/auditory stabilities in the context of an alternative view of phonetic (phonological) structure, one which identifies the basic phonetic units with particular units of speech production—articulatory gestures (Browman & Goldstein, 1986). In this context we will be led to the conclusion that the articulatory/auditory stabilities uncovered by Stevens are broadly relevant to a theory of pattern formation in the domain of articulatory gestures. This includes not only explaining, in certain cases, the kinds of gestures languages employ contrastively (as Stevens argues), but might also be extended to rationalizing the kinds of intergestural organization languages characteristically employ. On the other hand, related considerations lead us to conclude that these stabilities do not critically constrain the processes of talking and listening. The stabilities emerge from these activities, but do not further constrain them.

0095-4470/89/010091 + 07 $03.00/0

Within the framework of an articulatory phonology (Browman & Goldstein, 1986, 1987; Goldstein & Browman, 1986), utterances in a language are represented as organized structures of primitive units, or gestures. Each gesture corresponds to a characteristic pattern of coordinated movement within one of the relatively independent articulatory subsystems. At the coarsest level, these are the oral, laryngeal, and velic systems. Within the oral system, relatively independent lip, tongue tip/blade and tongue body subsystems can be identified. The characteristic patterns of movement can be seen as forming and releasing constrictions within each of these subsystems.

Contrasting utterances in a language can be formed by employing gestures from these different subsystems. For example, the syllable /bi/ would be represented as composed of two gestures: a lip gesture (bilabial closure) and a tongue body gesture (palatal constriction). The syllable /di/ contrasts with /bi/ in having a tongue tip gesture (alveolar closure) instead of the lip gesture. /pi/ constrasts with /bi/ in having an additional laryngeal gesture (glottal opening–closing). These differences are inherently discrete (and potentially contrastive, therefore) within the context of a gestural description of speech —different structures are moving. The fact that there may be relatively invariant auditory properties that correlate with such gestural differences (as Stevens argues) may contribute to their stability (and to the frequency of contrasts of this type in the languages of the world), but the discrete nature of these differences does not depend on such correlations.

However, in cases where contrasts involve the same articulatory system moving in different characteristic ways, quantal articulatory–auditory relations may have a major role to play in accounting for the values of the movement parameters that are used contrastively. For example, /di/ and /zi/ both begin with gestures of the tongue tip system. The difference is in the degree of constriction produced in the two cases (complete closure *vs.* a "critical" degree of constriction for turbulence-production). Likewise, /si/ and /ʃi/ involve the same tongue tip/blade system, but different precise locations of constriction (in addition, shaping of the tongue behind the constriction may differ, Ladefoged and Maddieson, 1986). Thus, gestures of the same articulatory subsystem may contrast in degree and location of constriction. [In the computational model of phonetic structure that we are developing, together with colleagues at Haskins Laboratories (Browman, Goldstein, Saltzman & Smith, 1986; Saltzman, Rubin, Goldstein & Browman, 1987), the dynamical equations that describe the motion of a given gesture include constriction location and degree as parameters.]

If we view a particular articulatory subsystem strictly in terms of its anatomy, without any regard to its acoustic output, then the parameters of constriction location and constriction degree would define a completely continuous space. There would be no reason for languages to contrast gestures with particular values of these parameters. The quantal relations can be seen as partitioning these continuous dimensions into the discrete regions that languages employ contrastively. Given the articulatory–auditory relations Stevens presents, a more or less random (uniform density) sampling of parameter space would give rise to an auditory space with relatively discrete density concentrations. Discrete gestural parameterizations could then develop (phylogenetically or ontogenetically) from these concentrations. Thus, quantal considerations could help account for selection of contrastive parameter values for articulatory gestures, and therefore be part of a more general theory of phonological selection (e.g., Lindblom, 1986; Lindblom, MacNeilage & Studdert-Kennedy, 1983).

Stevens' own view of the role of QT in accounting for systems of phonological contrast makes an additional claim. For him, selection is viewed in terms of the development of

an aspect of communicative efficiency: evolution so as to minimize the precision required of a talker. This can be seen as he assesses the value of phonological features based on quantal considerations: they have "the desirable attributes that. . . only limited articulatory precision is required to obtain a desired auditory response when a feature is implemented" (Stevens, 1989, p. 40). If this view were correct, we would expect to see a reflection of this in speech production behavior: the degree of precision in speech production should be constrained by the requirement to produce specific, local auditory responses in the listener. In other words, speech gestures are predicted to be free to vary, as long as they stay within a particular quantal region—they should involve no more and no less precision than this would require. This also suggests a view of speech perception, in which the listener is critically dependent on certain kinds of (relatively) local auditory properties.

A number of observations lead us to reject this view of precision in speech production. In particular, it is clear that speech gestures are regularly produced with either greater or less precision than this view would suggest. As an example of greater precision, consider the fact that there are regular between-language phonetic differences that do not seem to occur as contrasts within any one language (Ladefoged, 1980). For example, Stevens cites the data presented by Disner (1983) showing that the high front unrounded vowel /i/ is higher and more fronted in some languages than in others. Stevens accounts for these differences by hypothesizing that the same features (e.g., [+ high], [− back]) can be implemented in different ways from one language to another. However, this would mean that the speakers of these languages are producing gestures with greater precision than the quantal theory would suggest is necessary, because both "versions" of the vowel fall within the same quantal feature region. The QT provides no rationale for this extra degree of precision.

In other cases, the production of speech gestures seems to show less precision than a quantal analysis would suggest. Of particular interest here are the frequent examples of "weakening" in casual speech. For example, stops may be weakened so that they do not achieve complete closure: [ɸə'lItIk̟] 'political'; ['spəUxsmən] 'spokesman' (examples from Brown, 1977). Such examples can be considered part of a general process of reduction of gestural magnitude in fast, unstressed, casual speech (Gay, 1981; Lindblom, 1983; Munhall, Ostry & Parush, 1985). Note, however, that these reductions do not seem to be constrained so as to keep some quantal acoustic property intact—the acoustics of closure and release burst are replaced with acoustics more characteristic of fricatives —continuous (but weak) turbulence. These gestures are surely different from clear (careful speech) fricative gestures, and listeners readily identify the intended lexical items. Yet in terms of some criterial quantal property they have crossed a boundary. The problem posed by such reductions for a quantally-based theory of precision in speech production have, in fact, been discussed by Perkell (1980), who attributes the phenomenon to the availability of "higher-level" information in the message. Such appeals to other sources of information available to a listener do not, however, rescue the strong form of the view that the talker's precision is quantally constrained.

Perkell & Nelson (1982) argue that (at least some) aspects of speech production are sensitive to quantal relationships. They show that, in production of the vowels /i/ and /a/, substantially more variability is found in the positioning of the tongue along the dimension corresponding to constriction location of a particular vowel than along the dimension of constriction degree. Since formant frequencies are more sensitive to small perturbations in constriction degree than in constriction location (within quantal

regions), this result is taken as support for the quantal theory in speech production. However, as they go on to show, these effects can also be seen as a consequence of the kinds of characteristic muscular contractions employed in these articulations, as was earlier suggested by the model of Fujimura & Kakita (1979). Thus, the observed differences in variability do not necessarily reflect active ("real-time") constraints on the precision of movement. The assumption that they do *not* reflect such active constraints fits much better with the widespread occurrence of reduction phenomena, as discussed above.

In addition to gestures being regularly produced with greater and less precision than would be predicted on the basis of quantal considerations, there is another phenomenon that argues against quantal constraints being actively employed in speech production. Under certain conditions, speech gestures can be *uncoupled* from their acoustic consequences. Such cases have been described in Browman & Goldstein (1987) as "hiding" of gestures due to the increase in overlap of gestures in casual speech. A gesture may be (normally) produced, without producing any local acoustic or auditory consequences (its quantal characteristics), because the entire constriction and release occurs while some other closure is in place.

Some instances of "hiding" from the AT & T X-ray data base (collected in Tokyo) are presented in Browman and Goldstein (1987). For example, the words "perfect memory" were produced both as a phrase in the context of a fluent sentence, and as a sequence of two words within a longer word list. In the sentence version, the final /t/ of "perfect" is apparently (auditorily) deleted—careful listening reveals no /t/: its phonetic transcriptions is ['pɚfɘk'mɛmɚri]. Yet the lead pellet attached near the blade of the subject's tongue reveals the presence of the alveolar closure gesture for the final /t/—there is a movement toward the alveolar ridge, a movement remarkably similar to that seen in the word list version in which the /t/ is clearly audible. The articulatory difference between the phrases can be seen in the pattern of gestural overlap—in the sentence version, the alveolar closure gesture is completely overlapped by the preceding velar closure gesture (for the /k/) and the following bilabial closure gesture (for the /m/). Its acoustic consequences are, therefore, completely hidden. Other examples of hiding involved apparent assimilations rather than deletions: for example "seven plus" being produced as ['sɛvm̩plʌs], where again the tongue tip gesture for the final alveolar is present, but overlapped.

Browman & Goldstein (1987) argue that such hiding of gestures could be a general phenomenon that accounts for a variety of the kinds of assimilations and deletions reported in narrow phonetic transcriptions of fluent speech in English (e.g. Brown, 1977; Shockey, 1974). If so, this phenomenon represents a strong challenge to the view that speech production is actively controlled so as to achieve certain quantal auditory properties. The gestures in these hiding examples not only fail to have their criterial properties, they do not have any (local) acoustic properties at all. Thus, the gestures remain a (relatively) invariant part of a word's production, even when there are no correspondingly invariant auditory properties associated with them.

The possibility of a gesture being "hidden" by other gestures highlights the fact that the acoustic output associated with a given gesture will be determined not only by that gesture, but by the entire ensemble of concurrent gestures. Thus, a theory that attempts to account for the selection of phonological contrasts on the basis of articulatory–auditory relations must also take into account the possible ways in which gestures may be coordinated with one another. Within the articulatory phonology framework, gestures are

organized into larger structures (represented in gestural "scores") that explicitly specify the relations among the gestures of some lexical item. In general, gestural complexes for different words may contrast not only in the set of gestures involved, but also in their organization. For example, contrasts among prenasalized, nasal, and postnasalized stops (Anderson, 1976) all involve the same set of gestures (an oral constriction gesture and velum lowering and raising gestures), but different relationships among the gestures.

Within the computational model we are developing, the relative timing between any two gestures is specified by means of phase: each gesture is treated as a 360° cycle (including constriction and release), and two gestures are phased by synchronizing some point in the two gestures' cycles: for example, the phase corresponding to the attainment of target of a gesture could be synchronized with the release of another (preceding) gesture. As with the parameters of constriction degree and location, phasing can, in principle, refer to a complete continuum of values. Again, however, it seems that there are certain characteristic phasings that occur repeatedly (and contrastively) in languages. While it may well turn out that the selection of certain values of phasing follows from general constraints on motor coordination, here we consider how QT might profitably be extended to partition the phasing continuum into discrete categories.

If we were to examine the relationship between the relative phasing of two gestures and their acoustic output, we would likely see a pattern much like that in Stevens' Fig. 1: broad, stable plateaus separated by intervals of rapid change. As an example, consider two oral constriction gestures: bilabial closure and alveolar fricative. If the alveolar fricative gesture precedes the bilabial closure substantially, the fricative will be released before the bilabial is formed: we have a sequence that would be transcribed as [zəb], as in the middle of the word "Elizabeth". As we decrease the intergestural interval (continuously) the acoustic output will not change very much (this is a broad plateau), until "suddenly" the release of the fricative gesture overlaps (and is hidden by) the stop closure. This would be the case in the word "Lisbon" [zb].

Further "sliding" of the fricative gesture will again produce little change in the acoustic output (another plateau) until the two gestures are virtually synchronous. At this point, the output changes dramatically: the fricative will be completely hidden by the more extreme constriction. This is not likely to be a broad or stable region, because sliding the fricative just a bit more will result in some fricative noise *following* the oral constriction. (The relative stability of this state will depend on the relative duration of the two gestures). This would not be a useful organization for a language to use for another reason: there will be little acoustic difference between the pair of gestures lined up in this way and the bilabial closure alone, because the tongue tip gesture is hidden. The importance of such considerations in constraining the relations among gestures has been discussed by Mattingly (1981), who finds in them the basis for the sonority hierarchy. In general, more extreme constrictions will "mask" less extreme ones, so the less extreme ones will have to be produced, at least in part, outside of the constriction interval of the more extreme ones.

Finally, sliding the fricative still more will result in [bz], as in "cabs", and then [bəz] as in "bazooka". Changing the relative timing of these two gestures in a continuous fashion results in four relatively discrete, stable regions, and one relatively unstable one.

Thus, quantal considerations can partition the phasing continuum for two oral constrictions into four stable regions. These regions correspond to possible contrastive organizations in English, as exemplified above. The fifth region, when the two gestures are relatively simultaneous, is not employed (in English), and this might also be

accounted for by a QT-type account. Interestingly, the simultaneous arrangement of two gestures *is* possible as as contrastive organization when one of the two gestures is not an oral constriction,[1] but rather a velic or glottal gesture (as discussed in Browman & Goldstein, in press). In these cases, hiding does not occur. Thus, nasal stops (simultaneous organization) can contrast with both prenasalized and postnasalized stops. Voiceless unaspirated stops (simultaneous organization) can contrast with both postaspirated and preaspirated stops. Kingston (1988) also discusses some aspects of glottal-oral gesture timing from a point of view similar to this and concludes that some constraints on gesture timing depend on the degree of oral constriction: for stops and fricatives, glottal gestures influence not only voicing, but pressure levels that support fricative and release-burst characteristics.

While the relevance of quantal relations to a theory of gestural organization needs to be worked out in detail, it seems likely that such relations might have a role to play in accounting for the selection of certain gestural patterns. Thus, the domain of the quantal theory may, in fact, be wider than Stevens proposes—contrastive *organization* of gestures, as well as contrastive gestures themselves, may be accounted for. However, in this new domain as well, the role of quantal considerations is to partition a continuum into natural regions—not constrain the precision required of speakers. Thus, although languages will tend not to have contrastive organizations with two simultaneous oral obstruent gestures (because one of the gestures will be hidden), such structures regularly do occur in casual speech.

This work was supported by NSF grant BNS 8520709 and NIH grants HD-01994 and NS-13617 to Haskins Laboratories.

References

Anderson, S. R. (1976) Nasal consonants and the internal structure of segments, *Language*, **52**, 326–344.

Browman, C. P. & Goldstein, L. (1986) Towards an articulatory phonology, *Phonology Yearbook*, **3**, 219–252.

Browman, C. P. & 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 (1988) *Papers in laboratory phonology I: Between the grammar and the physics of speech* (J. Kingston & M. E. Beckman, editors) Cambridge: Cambridge University Press.

Browman, C. P. & Goldstein, L. (in press) Gestural structures and phonological patterns. In *Modularity and the Motor Theory of Speech Perception* (I. G. Mattingly & M. Studdert-Kennedy, editors). Hillsdale, NJ: Lawrence Erlbaum Associates.

Browman, C. P., Goldstein, L., Saltzman, E. & Smith, C. (1986) GEST: A computational model for speech production using dynamically defined articulatory gestures. *Journal of the Acoustical Society of America*, **80**, S97. (Abstract)

Brown, G. (1977) *Listening to spoken English*. London: Longman.

Disner, S. F. (1983) Vowel quality: The relation between universal and language specific factors. *UCLA Working Papers in Phonetics*, **58**, Los Angeles, CA.

Fujimura, O. & Kakita, Y. (1979) Remarks on the quantitative description of lingual articulation. In *Frontiers of Speech Communication Research* (B. Lindblom & S. Öhman, editors). London: Academic Press.

Gay, T. (1981) Mechanisms in the control of speech rate, *Phonetica*, **38**, 148–158.

Goldstein, L. & Browman, C. P. (1986) Representation of voicing contrasts using articulatory gestures, *Journal of Phonetics*, **14**, 339–342.

Jakobson, R., Fant, C. G. M. & Halle, M. (1969) *Preliminaries to speech analysis: the distinctive features and their correlates*. Cambridge, MA: The MIT Press.

[1]Some particular combinations of two oral constrictions can be organized in simultaneous fashion—those involving labial and velar gestures. Here it can be argued that the acoustic properties of these two constrictions reinforce one another (Ohala & Lorentz, 1977).

Ladefoged, P. (1980) What are linguistic sounds made of? *Language*, **56**, 485–502.
Ladefoged, P. & Maddieson, I. (1986) Some of the sounds of the world's languages, *UCLA Working Papers in Phonetics*, **64**, 1–137.
Lindblom, B. (1983) Economy of speech gestures. In *The production of speech* (P. F. MacNeilage, editor). pp. 217–245. New York: Springer-Verlag.
Lindblom, B. (1986) Phonetic universals in vowel systems. In *Experimental phonology* (J. J. Ohala & J. J. Jaeger, editors), pp. 13–44. Orlando, FL: Academic Press.
Lindblom, B., MacNeilage, P. & Studdert-Kennedy, M. (1983) Self-organizing processes and the explanation of phonological universals. In *Explanations of linguistic universals* (B. Butterworth, B. Comrie & Ö. Dahl, editors) pp. 181–203. The Hague: Mouton.
Mattingly, I. G. (1981) Phonetic representation and speech synthesis by rule. In *The cognitive representation of speech* (T. Myers, J. Laver & J. Anderson, editors) pp. 415–420. Amsterdam: North-Holland.
Munhall, K. G., Ostry, D. J. & Parush, A. (1985) Characteristics of velocity profiles of speech movements, *Journal of Experimental Psychology: Human Perception and Performance*, **11**, 457–474.
Ohala, J. J. & Lorentz, J. (1977) The story of [w]: An exercise in the phonetic explanation for sound patterns. *Proceedings of the third annual meeting of the Berkely Linguistics Society*, pp. 577–599. Berkeley, CA.
Perkell, J. S. (1980). Phonetic features and the physiology of speech. In *Language production* Vol. 1, *Speech and Talk* (B. Butterworth, editor) pp. 337–372. New York: Academic Press.
Perkell, J. S. & Nelson, W. (1982) Articulatory targets and speech motor control: A study of vowel production. In *Speech Motor Control* (S. Grillner, B. Lindblom, J. Lubker & A. Persson, editors). pp. 187–204. Oxford: Pergamon Press.
Saltzman, E., Rubin, P. E., Goldstein, L. & Browman, C. P. (1987) Task-dynamic modeling of inter-articulator coordination. *Journal of the Acoustical Society of America*, **82**, S15. (Abstract)
Shockey, L. (1974) Phonetic and phonological properties of connected speech, *Ohio State Working Papers in Linguistics*, **17**, 1, 143.
Stevens, K. N. (1989) On the quantal nature of speech, *Journal of Phonetics*, **17**, 3–45.
Stevens, K. N. (1981) The search for invariant acoustic correlates of phonetic features. In *Perspectives on the Study of Speech* (J. Miller & P. Eimas, editors) pp. 1–38. Hillsdale, NJ: Lawrence Erlbaum Associates.