Representation and reality: physical systems and phonological structure

Catherine P. Browman and Louis Goldstein*

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

Pierrehumbert's contribution lays out, in lucid fashion, the problems inherent in relating different kinds of representations of speech, in particular, representations that have as their goal elucidating systems of contrast and combination of units (the phonological), and representations that have as their goal precise physical descriptions (the phonetic). She argues that, in resolving these problems, first, phonetics and phonology cannot be treated in isolation from each other; second, properties of the real world must be seen as constraining the relation between phonetics and phonology; and third, a one-to-one correspondence between quantitative and phonetic, on the one hand, and abstract and phonological, on the other, cannot be maintained—mental representations can be quantitative, and physical representations abstract. On all these points, we agree with Pierrehumbert.

Pierrehumbert also maintains that phonology and phonetics involve disparate representations, drawing upon several dichotomies to help characterize the differing kinds of terrain for which these different representations are assumed to be suitable guides. One of these dichotomies is *cognitive vs. physical.* Another is *qualitative vs. gradient.* The phonetic domain is seen as physical and gradient and is described by "continuous" mathematics (the calculus). The phonological is seen as cognitive and (for the most part, but not completely) qualitative and is described by formal languages, rather than calculus.

While these dichotomies are superficially plausible, recent work on self-organization in complex biological and physical systems (e.g., Haken, 1977) can be taken as a lesson that the rich structuring shown by these systems can be understood when the importance imparted to such dichotomies is abandoned. In the paragraphs below, we first give some examples (from speech and elsewhere) that incline us against enforcing these kinds of dichotomies. We then introduce an alternative suggestion for speech representations that can serve the different goals that Pierrehumbert identifies. In this alternative, the goals are served by macroscopic and microscopic properties of one and the same system. We conclude by exemplifying this approach using speech error data.

1. Cognitive and physical need not be distinct

Pierrehumbert argues that phonology is, *prima facie,* cognitive, because phonological contrast is the basis for the association of form and meaning, which in turn must

be part of an individual's cognitive structure. However, because she also assumes that the cognitive component is qualitative and discrete, and further that the physical cannot be qualitative (except as "analyzed" by a cognitive system), the cognitive and the physical must be different (and representationally incommensurate). However, as we will show below, the physical world is both gradient and qualitative, and there is no reason why the cognition cannot be attuned to those qualitative aspects of the real world (in this case the activity of talkers and listeners), rather than imposing a discrete, qualitative order on an otherwise homogeneously continuous world. Indeed, the entire research program of "direct realism" (e.g., Gibson, 1966, 1979; Turvey, 1977; Fowler, 1980; Fowler, Rubin, Remez & Turvey, 1980) proceeds from the assumption that human beings' perceptions of the world are, in fact, true of the real world, and investigates how the perceptual information specifies properties of the real world.

Recent research on coordinated movement provides a number of challenges to the view that cognitive and physical systems are distinct and incommensurate. This research (e.g., Kelso & Tuller, 1984; Kugler & Turvey, 1987; Turvey, 1990) shows that in order to account for coordinated movement, the physical properties of an organism must play a central role, and that coordination is accomplished through an interaction of cognitive and neural activity with the principles of physical self-organization. A particularly striking demonstration of this can be seen in the nature of phase transitions between alternate modes of coordinating rhythmic movements (Kelso, 1984; Kelso & Scholz, 1985). When two index fingers are oscillated at the same frequency, two phasing modes are possible, symmetric and anti-symmetric. If a person starts oscillating in the antisymmetric mode and is asked to speed up, eventually there is an abrupt (and involuntary) shift to the symmetric mode. (The reverse is not true.) This phase shift has been modeled by Haken, Kelso & Bunz (1985) using a potential function that predicts the stability of particular phase relations as a function of the oscillation frequency of the components. A number of empirical details of the phenomenon (such as the appearance of fluctuations in phase as the critical point is reached) support this physical stability analysis and the hypothesis that the sudden change in phase is a physical bifurcation related to increased instability of the antisymmetric mode.

Now in this simple kind of task, the fact that the results can be predicted on the basis of a physical systems analysis may not seem particularly surprising. After all, the human body *is* a physical system (whatever else it is), and it is not hard to convince oneself that the two fingers are mechanically (or at least neurally) coupled to one another. A common intuition here is that it is just somehow "more difficult" to move the fingers in the antisymmetric mode, and at high frequencies, the subject simply gives up. Given such an analysis, this task might not have much to say about the role of cognition. Recently, however, Schmidt, Carello & Turvey (1990) have extended this method to the case where the two limbs being coordinated belong to two different people. The subjects face each other and are instructed to swing their legs in synchrony (either in or out of phase). Exactly the same results were obtained as in the within-person case—an abrupt shift to the symmetric mode as frequency was increased, with all the hallmarks of a physical bifurcation. So what we have is a system that is behaving in a way that is well explained by physical systems, yet there is no mechanical (or hard-wired neural) coupling between the oscillating elements in this case. The coupling must be, as the authors conclude, *informational*. To the

extent that we think of this kind of information as "cognitive" (which in a broad sense it must be; certainly " visual information processing" has traditionally been so considered), cognitive activity actually functions to couple the elements of a (single) physical system. It is hard to imagine a tighter interpenetration of the cognitive and the physical.

Of course, one may want to object that the sense of cognitive here is not the same one that Pierrehumbert is employing. Specifically, she is referring to a kind of introspectively available cognition that is not specifically perceptual or motoric. But there is still another moral lurking here. The coupled swinging legs constitute a physical system, but it is one that is softly assembled. That is, the two people can sit and swing their legs and watch each other without *intending* to synchronize their swings. Without the intentionality, the complex system is never actually assembled from the pieces (although there might be some spontaneous tendency to synchronize anyway). Thus, the (clearly cognitive) act of intending to synchronize provides the boundary conditions under which the self-organization of the physical system takes place (Kugler & Turvey, 1987; Turvey, 1990).

2. Physical systems are simultaneously gradient and qualitative

Perhaps the most dramatic examples of the simultaneously gradient and qualitative nature of the physical world are the instances of "self-organization" that have attracted attention in recent years (e.g., Prigogine & Stengers, 1984; Madore & Freedman, 1987). These demonstrations show that, under the right conditions, complex, qualitatively distinct forms may spontaneously emerge in previously homogeneous, undifferentiated media. For example, with the right concentration of reagents, a quiescent petrie dish will suddenly display colored concentric circles and rotating spirals, which eventually dominate the entire surface and then die out (the Belousov–Zhabotinskii reaction). One can describe what is going on in the petrie dish in different ways. From one point of view, it is clearly gradient—there is spatial and temporal variation in the concentration of particular ions that can be described in terms of the relevant (continuous) differential equations. From another point of view there are a number of distinct observable qualitative forms (e.g., circles and spirals). These are descriptions of the same physical system, differing only in the "grain" employed.

Self-organized systems, like the chemical oscillators described above, provide another window onto the relation between gradient and qualitative properties of physical systems. The reaction described above (and others, see Prigogine & Stengers, 1984) occurs only when the concentration of some of the reagents is within some critical limits. Thus, while concentration is a gradient quantity, the concentration continuum is inherently partitioned into regions which show qualitatively distinct behavior (quiescence *vs.* ring-formation). This partitioning into distinct long-term behaviors, or *bifurcation* as it is sometimes called, is a common property of physical systems. Even an equation as simple as (1), used to describe population dynamics (May & Oster, 1976, in Gleick, 1987; Hofstadter, 1981), shows such properties.

$$x_{\text{next}} = rx(1 - x) \tag{1}$$

(1) is iterated to get the population of each successive generation. Depending on the value chosen for *r*, the long-term behavior of the system will differ. At low values, the system settles down to a single value for *x* in successive iterations. As *r* increases, the behavior abruptly changes qualitatively, yielding a stable alternation of two values for *x*. This period-doubling bifurcation will occur again at some point as the value of *r* is increased further, yielding a cycling among four values. The doubling continues until at some point the system becomes chaotic, and does not show any predictable pattern of repetition. Thus, while equation (1) is defined in the world of continuous mathematics, it provides a landscape of discretely different behaviors as a function of the (gradient) value of its parameter. The potential for such discrete, qualitative behavior is always lurking in systems that include non-linear terms (cf. Thompson & Stewart, 1986), even quite simple ones, such as (1). Only in strictly linear systems is the gradience of the parameter space mirrored by a relatively undifferentiated landscape of system behavior (and even then, not always). Complex systems in the real world, however, involve non-linearity and qualitatively distinct behaviors.

There are various ways in which speech, as a physical system, can be seen as simultaneously gradient and qualitative. One familiar example is Stevens's quantal theory (Stevens, 1972, 1989). This theory holds that while it is possible to describe constrictions within the vocal tract in terms of continuous (geometric) parameters, the acoustic (and auditory) properties of the sound produced by the vocal tract are such that continuous parameter space is effectively partitioned into discrete regions. Within each region, the auditory properties are stable (they do not vary greatly as a function of small changes in the articulation), and qualitatively distinct from auditory properties associated with other regions. These regions are thus seen as the basis for contrast (distinctive features, for Stevens). To the extent to which this theory is correct, then speech, like some of the other examples of complex systems described above, is a physical system that is intrinsically both gradient and qualitative.

A second example involves the characterization of articulatory trajectories during speech. As Pierrehumbert notes, the articulators are constantly in motion, and thus the trajectories can be described in a gradient fashion. However, as a number of researchers have hypothesized and found, it is possible to model the time-varying motion of the articulators (during a particular speech gesture) using an invariant dynamical specification (Fowler *et al.*, 1980; Browman & Goldstein, 1985, 1990; Kelso, Vatikiotis-Bateson, Saltzman, & Kay, 1985; Ostry & Munhall, 1985; Vatikiotis-Bateson, 1988; Beckman, Edwards, & Fletcher, in press). That is, even though the articulators are moving, the underlying dynamical system that gives rise to this motion is not varying over time. There is a discrete interval of time during which this (temporally invariant) regime for a particular gesture is active. Thus, once again, the speech system exhibits behavior that is at one and the same time gradient and qualitative.

3. Macroscopic and microscopic properties of phononological structure

While the cognitive-physical and qualitative–gradient dichotomies do not seem to us to be useful, the differing representational goals that Pierrehumbert identifies as phonological *vs.* phonetic must be satisfied in some way. We would like to suggest

that these can be construed as macroscopic and microscopic properties of a single complex system. Thus, properties such as contrast and paradigmatic and syntagmatic relations are coarse-grained, macroscopic, relational properties that hold among the system's units. The precise articulatory and acoustic characterizations of these units (and their variation) are fine-grained, microscopic properties.

This macro–micro perspective does more than just provide a different name for a familiar distinction. The fact that macroscopic and microscopic properties simultaneously characterize the same complex (physical) system has substantive consequences. Recent work has begun to explore the general properties of cooperativity among system components, and of the linkage between patterns at different "scales" (Kugler & Turvey, 1987; Schoner & Kelso, 1988). An important characteristic of complex physical systems showing "self-organization" is that there is an interaction (or reciprocity) between the microscopic and macroscopic properties of the system (Kugler & Turvey, 1987). The nature of the microscopic units, or atomisms, affects the possible stable macroscopic structures, and the macroscopic organization affects the behavior of microscopic units.

Kugler & Turvey (1987) present examples of such micro–macro cooperativities. For example, they show how nest-building in insects can be analyzed in this way. Macroscopic chemical gradients *originate in* the behavior of individual insects (whose deposits contain pheromones), but these gradients also *constrain* the activities of the insects (causing them to deposit in high-pheromone density locations which results in the formation of macroscopic pillars and arches). This example was also used by Lindblom, MacNeilage & Studdert-Kennedy (1983), in order to show how certain phonological units could be derived from properties of speech. However, it is important to see that the micro–macro constraints are typically *reciprocal*.

For example, Kugler & Turvey present experiments that demonstrate micro–macro reciprocity in human coordinated movement. When individuals are asked to swing a pendulum from the wrist, the preferred "comfort mode" frequency depends on the size of the pendulum. When asked to swing two differently-sized pendula (one in each wrist) in absolute coordination, the observed frequency does not correspond to either single frequency, but it turns out to correspond to the frequency of a macroscopic virtual system, that treats that the two pendula as rigidly coupled. The macroscopic properties of the system are determined by the microscopic properties of the components (their natural frequencies), but the coupled system constrains, in turn, the individual wrist-pendulum systems, pushing them away from their own natural frequencies (see also Turvey, Rosenblum, Schmidt & Kugler, 1986). Schmidt (1988, discussed in Turvey, Saltzman & Schmidt, in press) shows that such effects hold even when it is two different individuals swinging the two pendula, so that the coupling is informational.

Applying this macro–micro perspective to phonological structure would involve showing that the contrastive and combinatoric properties (the ones Pierrehumbert calls phonological) arise out of the microscopic (articulatory and acoustic) properties of the individual atomisms (such as articulatory gestures), and, in turn, that the macroscopic properties constrain the details of the units. This perspective predicts that reciprocity between the grains of description should be the rule, not the exception. Such reciprocity, or mutual constraint, would come as a surprise to the phonological and phonetic "imperialists" that Pierrehumbert attacks, but fits more

comfortably with her outlook that a theory encompassing both domains, as well as their relations, is necessary to a full understanding of phonology and phonetics. Likewise, the syntactic view of phonetic/phonology relations, that Pierrehumbert rejects, runs afoul of such reciprocities. However, it seems to us that the semantic type of mapping which she proposes is also not a good analogy here. As Pierrehumbert notes, one mapping called "semantic", the lexical sound-meaning correspondence, is too arbitrary to capture what is going on in the phonology-phonetics relation. Even if the semantic mapping in question is that between a concept and its real-world extensions (such as the concept DOG and the set of dogs in the real world), it differs substantively in its possibilities for macro–micro reciprocities from the phonology-phonetics relation. While both concepts and phonological structure may be dependent on real-world conditions, their potential for affecting real world properties differs considerably. That is, although one's concept of DOG may affect one's relation with a real-world dog (e.g., patting or running from it), the concept does not have the same potential to constrain the nature of that dog that phonological structure does to constrain the properties of speech.

While macro–micro reciprocity has not been demonstrated conclusively for language at this point, it is possible to find a variety of examples of such reciprocity. One such example can be seen in the organization of English vowels into the paradigmatic system revealed in a series of "chain shifts" identified by Labov, Yaeger & Steiner (1972). In these sound changes in progress, subsets of vowels show coordinated patterns of movement along particular "tracks" in the vowel space. In particular, tense (and ingliding) vowels tend to raise, so that low vowels become mid, and mid vowels become high. These constrained (and apparently universal across dialect) patterns of movement and the paradigmatic relations that they reveal clearly constitute a macroscopic property of the English vowel system (and one that would be captured as part of the phonology, in most accounts.) It is possible, however, to see reciprocity between this structure and the microscopic properties of the vowel units themselves.

First, Goldstein (1983) has shown that it is possible to derive the tracks along which vowels move during shifts on the basis of articulatory–acoustic relations. Random articulatory perturbations of model vowels results in acoustic variation that is directed along the dimensions that are involved in chain shifts (principally, the high–low dimension). Thus, the microscopic properties of these vowels gives rise to the layout of the macroscopic tracks.

Second, there is also evidence that the macroscopic state of the vowel system also constrains the microscopic properties of individual vowels. Labov (1972) showed that while part of an "active" (macroscopic) shift (which might span 50–75 years), a vowel shows large, regular (microscopic) context effects in speakers' productions. Before and after the shift, however, the context effects are much smaller. Thus, the macroscopic dynamical state of the system (actively moving *vs.* static) constrains the microscopic properties of the vowels. This difference in amount of contextual variation at different macroscopic stages provides another kind of support for the complex self-organized system analysis. When such systems are pushed away from particular stable configurations, they typically show an increase in fluctuations, that is, variations in behavior that differ from the stereotypic pattern (Turvey *et al.*, 1986). As a critical point for a non-equilibrium phase transition is reached (as in the

finger oscillation example discussed earlier), such fluctuations become extremely large, and provide part of the signature of the change in state (Schoner & Kelso, 1988). The context variation of individual vowel productions can be viewed as just such fluctuations. When the vowel system is at a critical "point" (where a point at this time scale can last many years), increased fluctuations (variation) can be observed.

A different kind of macro–micro linkage may be found in the detailed language-particular differences in the physical properties of the same phonological or categorial phonetic structures (Keating, 1985, 1990). In at least some cases, these microscopic language differences are correlated with the macroscopic structure of contrasts. Let us take one of the examples that Pierrehumbert gives, the fact that the precise frontness of high front unrounded vowels may vary from language to language. Wood (1982) has shown that a prepalatal constriction location is preferred for /i/ in languages which contrast /i/ and /y/, whereas a midpalatal location may be found in languages without the rounding contrast. He then presents modeling results that suggest that this difference is functional—a prepalatal location yields a greater acoustic differentiation of /i/ and /y/. Thus it appears that the macroscopic property of contrast is constraining microscopic properties of the units. Ladefoged (1982) presents a number of examples of this kind, in which language differences in details of production can be related to presence or absence of certain contrasts.

Finally, it has been shown that the amount of contextual variation evidenced by a given phonetic unit may vary. This allowable "region" for a given unit has been modeled by Keating (1988) as a "window", and by Manuel (1990) and Manuel & Krakow (1984) as a "target area". These latter two papers have related the size of the target areas for vowels to the number of contrastive vowels in the system, showing that the areas are smaller when there are more vowels in the system. Again, this seems to suggest an interaction between macroscopic and microscopic system properties.

In the last two examples, the effect of contrast on phonetic units does not provide evidence of reciprocity, *per se*. The effects have seemed only to propagate from the macroscopic to the microscopic. However, there have been a number of attempts to show how the qualitative properties of contrast and combination of phonological systems arise from, or at least are constrained by, the articulatory and acoustic properties of speaking (Stevens's quantal theory—Stevens, 1972, 1989; Lindblom's theory of adaptive dispersion theory—Lindblom *et al.*, 1983; Lindblom & Engstrand, 1989; Ohala's vocal tract constraints—Ohala, 1983). While none of these attempts is completely successful, it seems clear that the formation of phonological systems is at least partly molded by the articulatory and acoustic properties of talking. Thus, this is another instance of macro–micro reciprocity: systems of contrast are founded on the microscopic properties of talking, but they also constrain microscopic properties.

In the above sections, we have illustrated that treating phonological and phonetic representations as incommensurate, on the basis of dichotomies such as cognitive–physical and qualitiative–gradient, is probably misguided. These distinctions are irrelevant to, or get in the way of, an understanding of complex physical systems. Within the view that phonological structure is a complex, self-organized system, it is possible to acknowledge that different descriptive tools (e.g., symbols *vs.* equations) are appropriate for different classes of phonological phenomena, while treating the

phenomena as differing grains of a single integrated system. In practical terms, when a linguist is describing some regularities for which a single descriptive tool is appropriate, it may be possible (in some circumstances) to ignore other grains of description. However, evidence presented for reciprocity between macroscopic and microscopic properties strongly suggests that when pursuing a complete understanding of phonological structure and the cognitive/physical activity of talkers and listeners, one ignores the complete system at one's peril.

4. Phonological structure, its representation and transcription: the status of the segment

As we have suggested, paradigmatic contrast and syntagmatic combination can be usefully viewed as important macroscopic properties of phonological system structure. As such, they ought be captured in a *representation* of this system structure. While there are many ways in which this could be done, the use of the *segment,* or a *phonemic transcription,* has been very widely employed as a particular hypothesis, identifying the unit of contrast with the unit of combination. However, we would argue that the basis for such units seems to be in their utility as a practical tool rather than in their correspondence to important informational units of the phonological system.

The primarily practical utility of segmental transcriptions is noted by other papers in this volume. For example, Pierrehumbert (1990) sees fine phonetic transcription as "a convenience for the researcher attempting a rough organization of his observations" (p. 390), but finds "no evidence that the elements of fine transcription can be viewed as elements of a discrete representation in the mind" (p. 389). Although Pierrehumbert very clearly distinguishes fine transcription from transcription using phonologically distinctive elements, the pragmatic view towards transcription would appear to be extended to all transcription by Ladefoged (1990). He quotes from Abercrombie's (1964) definition of a phonemic transcription as one in which "the smallest possible number of different letters [symbols] . . . distinguish unambiguously all words of different sound in the language" (p. 341). Given that an alphabetical transcription system is being used, this means the symbols used are "segments".

It is important not to confuse the units of such a practical descriptive tool, however useful, with qualitative, informational units that function in the (cognitive/physical) phonological system, when viewed from a macroscopic perspective. The criteria for a useful practical representation (such as economy) are different from those relevant to representing theoretically significant aspects of the system. It seems to us, however, that the success of segmental symbol strings as practical devices has inclined many to make just this confusion, and to assume an isomorphism between the units of transcription and the units of the system itself. Thus, just as Pierrehumbert suggests that fine phonetic transcription has no real theoretical status in phonetics, we suggest that there is no reason to assume that representations employing segmental transcriptions have any theoretical status in phonology.

From this perspective, then, phonemes become one particular, linear, local and symbol-oriented—"segmental"—solution to the necessity of capturing two related kinds of macroscopic information: distinctive aspects of lexical items, and groupings of allophones (whether alternative pronunciations of the same word, or regular variations, that is, restrictions of distributions). It is, however, not a necessary

solution; these same facts can be captured with other units, including gestures and constraints on gestures. Indeed, for almost half a century, the unit of distinctiveness has usually been considered to be, not the segment, but the feature. Moreover, recent phonological proposals such as feature geometry (Clements, 1985; McCarthy, 1988) have explicitly separated the paradigmatic and syntagmatic properties that have been traditionally conflated in the segment (when viewed as a feature bundle). In particular, root nodes are only syntagmatic units in these proposals, constraining how features combine, but it is the features that convey contrast, and they can align in various ways with respect to the syntagmatic frames.

From this perspective, the segmental hypothesis can be viewed as being primarily a specific (local and linear) hypothesis about featurally cohesive syntagmatic units. Researchers have also attempted to extend the segment to indicate a linear chunking of speech, or to subdivide some larger unit such as the syllable. To at least a first approximation, the segment has been useful in dealing with the acoustic signal. That is, localized linear segmentations of the acoustic signal have real validity, at a coarse level of description. However, it appears that the value of the segment even in characterizing the acoustic signal is limited to the kind of rough organization of observations that Pierrehumbert mentions. The segmental approach runs into trouble, for example, in the syntagmatic world of actual utterances, where it is difficult to find acoustic invariance for any single segment, and where the information associated with a segment might in fact not be temporally localized in the region of that segment, but extend throughout the syllable, or even into other syllables.

Nearey (1990) as well as others has attempted to handle this latter problem by redefining the segment as sensitive to information present in an entire VC (or theoretically, an entire VCV). This of course completely destroys the simple physical definition of a segment as a local linear chunking of the acoustic signal. Nearey explicitly disavowed any featural assumptions in this paper, but investigated the hypothesis of the segment as a subsyllabic unit by comparing segmental and "transssegmental" (i.e., diphone VC or CV) models. While the title of the paper might lead the casual reader to think the paper presents evidence that the segment is a unit of speech perception, in fact Nearey found that a V × C bias component was essential, so that a "pure" segmental model was inadequate. Moreover, Nearey did not compare the segmental hypothesis to other hypotheses of subsyllabic units, such as syllable components (onset, nucleus, coda) or gestures. Thus, his analyses assume acoustic information is distributed transsegmentally and support a transsegmental cognitive component, as well as possibly providing evidence for some kind of subsyllabic cognitive unit. (In fact, it appears to us that his results would be consistent with a gestural analysis.)

Nearey (1990) also cited speech production error data as evidence in favor of the relevance of the segment to speakers' behavior. However, we argue that speech production errors provide no evidence for the behavioral relevance of the segment. (Similar conclusions were reached by Roberts, 1975, and by Boomer & Laver, 1968.)

It has been repeatedly observed that most speech production errors are single feature errors, and therefore an analysis in terms of features (or feature-like entities) is necessary to describe one important aspect of the entire corpus. Beyond this, however, production errors appear to be divided into two categories:

non-interaction and interaction errors (Shattuck-Huffnagel, 1986). These categories differ in two ways: whether a featural description is sufficient, and whether the errors are concentrated in word onsets.

Thirty to forty percent of the errors in the MIT corpus fall into the category of non-interaction errors (30%: Shattuck-Huffnagel & Klatt, 1979; 40% of 1984 count: Shattuck-Huffnagel, 1987). For errors in this category, there is no obvious source for the error in the environment (e.g., "the *D*utch publishers"→"the *G*utch publishers": Shattuck-Huffnagel & Klatt, 1979), and therefore a purely featural analysis is sufficient to describe this subset (assuming it maintains the same featural distribution as the corpus as a whole). The errors tend to occur throughout the word, rather than being concentrated in word-onset position (Shattuck-Huffnagel, 1987).

Sixty to seventy percent of the MIT error corpus consists of interaction errors: anticipatory, perseveratory and exchange errors (e.g., "they cut their *h*air *sh*ort"→ "they cut their *sh*air *h*ort": Shattuck-Huffnagel, 1987). In these errors, the two consonants presumed to be causally interacting are much more similar than would be expected from an interaction of purely independent features, and therefore a purely featural analysis is not sufficient to account for this subset of errors. Rather, some kind of featurally cohesive unit is necessary to describe the interaction errors, unlike the non-interaction errors. Although Shattuck-Huffnagel & Klatt (1979) suggested this featurally cohesive unit was the segment, they had not at that time considered other possibilities, such as syllable or word onsets, which have been shown to account for more data in studies that have compared segmental and onset hypotheses (e.g., Shattuck-Huffnagel, 1983; Vitz & Winkler, 1973).

The interaction errors tend to occur in word onsets. In fact, 82% (and as much as 91% of the exchange error subset) occur in word onsets (Shattuck-Hunffnagel, 1987). Shattuck-Huffnagel (1987) argued that when a word-initial consonant participates in an interaction error, "it usually does so by virtue of the fact that it is a word onset (p. 37)", even when it is a single consonant. The featurally cohesive unit for interaction errors, then, appears to be the word onset, not the segment. For example, out of 40 word-onset consonant clusters participating in exchange errors, 36 involved the cluster as a whole (e.g., "*br*eathing and *sm*oking"→"*sm*eething and *br*oking": Shattuck-Huffnagel, 1987).

The validity of the analysis of the structural importance of the lexical item in interaction errors is supported by the striking similarities between production interaction errors and lexical retrieval errors. Brown & McNeill (1966) showed that word onsets (and endings) differ from the rest of the word in being recalled more often in the tip-of-the-tongue state. Browman (1978) showed more specifically that, in addition to a general tendency for "gregariousness" (or "stickiness", in Nearey's terms p. 348), word-initial onsets, word-final VCs, and pre-stressed onsets are prominent in lexical retrieval errors. The two onset categories are thus prominent in both lexical retrieval errors (Browman, 1978) and production interaction errors (or at least exchange errors: Shattuck-Huffnagel, 1987). In addition, production interaction errors and lexical retrieval errors are similar in apparently having separable item and order (or filler and slot) components. Finally, Browman (1978) has argued that the prominence structure in lexical retrieval errors is an attribute of the retrieval process rather than of the lexical entry, which (in conjunction with the other similarities) suggests a possible identity between this process and Shattuck-

Huffnagel's (1987) first stage process during which the interaction errors are posited to occur. Thus, it seems likely that the patterns observed for the production interaction errors are attributable to the same process that is observed in lexical retrieval errors (except for the difference in word-final prominence), lending support to the analysis of these errors in terms of lexical units.

To recapitulate: word onsets plus independent features (or feature-like entities) are necessary and sufficient to account for most interaction errors, while features (or feature-like entities) are sufficient to account for the non-interaction errors. That is, cohesion in speech production errors appears to be defined with respect to the word. The featurally cohesive units are not the same everywhere in the word, nor are they segments, or even onsets of syllables. Rather, the cohesive units are the onsets of words. Arguments based on cohesion do not support the segment in production errors.

However, upon occasion errors involving word onsets break the onset into components, both in production interaction errors and in lexical retrieval errors. While such divisions might appear to be evidence for segments, we suggest that instead they are evidence for articulatory gestures. Like segments, gestures have the potential to be independent movable entities, and can combine into higher level units such as onsets, words, etc. Although a complete analysis in terms of gestures would need to be performed for a more nearly definitive statement, it is nevertheless suggestive that of the 40 errors listed in the Appendix of Fromkin (1973) under "Division of Consonant Clusters", approximately 35 can be analyzed as the movement of a single gesture. Moreover, at least two of the archisegments argued for by Stemberger (1983) as psychologically real units can be equated with gestures.

Many of the similarities between targets and errors captured in earlier analyses as featural similarities can also be captured using gestures. Using the pseudo-gestural analysis in Fig. 1 of the confusion matrix from Table 2 in Shattuck-Huffnagel & Klatt (1979), the distribution of numbers of gestures differing between the target and error is suggestive of gestural independence (81% one gesture, 18% two gestures, 1% three gestures). And at least 48 (and possibly up to 54) out of 54 "single feature" errors listed in the Appendix of Fromkin (1973) are also single gesture errors (e.g., "*pe*destrian" → "*te*bestrian", exchange of oral gestures). Note that if gestures are indeed a basic unit, then higher level phonological units such as words, onsets, etc., are associations of gestures (gestural "constellations"), although not necessarily in segmentally-sized units as assumed by Nearey (1990).
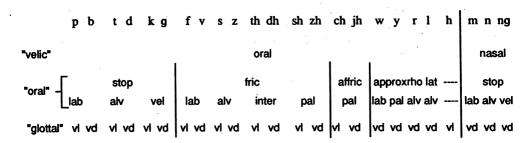
```
           p b   t d   k g   f v   s z   th dh   sh zh   ch jh   w  y  r  l  h | m n ng

"velic"                              oral                                        | nasal

        ┌          stop                 fric              affric | approxrho lat ---- | stop
"oral" ─┤
        └ lab     alv     vel | lab    alv    inter    pal | pal | lab pal alv alv ---- | lab alv vel

"glottal" vl vd  vl vd  vl vd | vl vd  vl vd  vl vd    vl vd | vl vd | vd vd vd vd  vl | vd vd vd
```

Figure 1. Pseudo-gestural analysis of segments in Table 2 of Shattuck-Huffnagel & Klatt (1979).

The hypothesis that gestural primitives are critically involved in speech errors makes an additional claim. As defined in articulatory phonology (Browman & Goldstein, 1986, 1989), gestural units are simultaneously discrete units of contrast and quantitatively specified units of articulatory action. The analyses in the previous paragraphs made use of the discrete properties of gestures, since most work on speech errors has assumed that they involve misorderings of discrete, qualitative units in the plan for an utterance. In fact, it has been argued that such misorderings occur before the units receive any articulatory instantiation (Shattuck-Huffnagel, 1987). Gestures, however, "always" have quantitative (gradient) articulatory properties, and if they are the units implicated in speech errors, then it should be possible to find evidence of these properties.

Such evidence is provided in a recent study by Mowrey & MacKay (1990). In this study, they recorded muscle activity during experimentally elicited speech errors, where the errors were induced using tongue twisters such as "Bob flew by Bligh Bay". The electrode placement allowed them to examine activity for [l]. For one recording session of this particular tongue twister, 48 of 150 tokens were produced with anomalous tongue muscle activity (outside the normal range of variation). These anomalies involved the insertion of [l] activity at time points where it was not appropriate (e.g. in "Bob" or "Bay") and the diminution of [l] activity in positions where it was expected. Crucially, these errors were graded. The inserted [l] activity showed a continuum from small amounts of activity to a level consistent with an intended [l]. Likewise, diminution of [l] showed a continuum of reduced activity. Overall, only five tokens showed an "all-or-none" change. Such graded activity is consistent with the quantitative characterization of gestures (in fact, reduction of magnitude has been proposed as a general property of gestures in casual speech, Browman & Goldstein, 1987), but is not consistent with a purely discrete pre-articulatory view of these errors.

Yet in another sense, the errors did show qualitative or discrete behavior. The inserted [l] activity was not smeared throughout the sentences, but was localized at very specific points with temporal profiles comparable to those of an intended [l], even though reduced in magnitude. This is also consistent with the discrete nature of gestures. A gesture is a dynamical system, with an invariant parameter set, that is active for a finite interval of time. As far as it is possible to tell from the muscle activity, the anomalous activity involved an inserted discrete unit of this kind, but with variably reduced magnitude. Thus, it appears that dual nature of gestures—discrete and quantitative—may well be crucial in accounting for speech errors.

References

Abercrombie, D. (1964) *English phonetic texts*. London: Faber & Faber.

Beckman, M. E., Edwards, J. & Fletcher, J. (in press) Prosodic structure and tempo in a sonority model of articulatory dynamics. In *Papers in Laboratory Phonology II* (G. Docherty & D. R. Ladd, editors). Cambridge: Cambridge University Press.

Boomer, D. S. & Laver, J. D. M. (1968) Slips of the tongue, *British Journal of Disorders of Communication*, 3, 1–12.

Browman, C. P. (1978) Tip of the tongue and slips of the ear: Implications for language processing, *UCLA WPP*, **42**.

Browman, C. P. & Goldstein, L. (1985) Dynamic modeling of phonetic structure. In *Phonetic linguistics* (V. A. Fromkin, editor), pp. 35–53. New York: Academic Press.

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*, **92**, 1–30.

Browman, C. P. & Goldstein, L. (1989) Articulatory gestures as phonological units, *Phonology*, **6**, 201–251.

Brown, R. & McNeill, D. (1966) The 'tip of the tongue' phenomenon, *Journal of Verbal Learning and Verbal Behavior*, **5**, 325–337.

Browman, C. P. & Goldstein, L. (1990) Gestural specification using dynamically-defined articulatory structures, *Journal of Phonetics*, **18**, 299–320.

Clements, G. N. (1985) The geometry of phonological features, *Phonology Yearbook*, **2**, 225–252.

Fowler, C. (1980) Coarticulation and theories of extrinsic timing control, *Journal of Phonetics*, **8**, 113–133.

Fowler, C. A., Rubin, P., Remez, R. E. & Turvey, M. T. (1980) Implications for speech production of a general theory of action. In *Language production* (B. Butterworth, editor). New York: Academic Press.

Fromkin, V. A. (editor) (1973) *Speech errors as linguistic evidence*. The Hague: Mouton & Co.

Gibson, J. J. (1966) *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.

Gibson, J. J. (1979) *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

Gleick, J. (1987) *Chaos: making a new science*. New York: Viking.

Goldstein, L. M. (1983) Vowel shifts and articulatory-acoustic relations. In *Abstracts of the tenth international congress of phonetic sciences* (A. Cohen & M. P. R. van den Broeke, editors), pp. 267–273. Dordrecht: Foris Publications.

Haken, H. (1977) *Synergetics: an introduction*. Heidelberg: Springer Verlag.

Haken, H., Kelso, J. A. S. & Bunz, H. (1985) A theoretical model of phase transitions in human hand movements, *Biological Cybernetics*, **51**, 347–356.

Hofstadter, D. R. (1981) Strange attractors: mathematical patterns delicately poised between order and chaos, *Scientific American*, **245**, 22–43.

Keating, P. A. (1985) Universal phonetics and the organization of grammars. In *Phonetic linguistics* (V. A. Fromkin, editor). New York: Academic Press.

Keating, P. A. (1988) The window model of coarticulation: articulatory evidence, *UCLA WPP*, **69**, 3–29.

Keating, P. A. (1990) Phonetic representations in a generative grammar, *Journal of Phonetics*, **18**, 321–334.

Kelso, J. A. S. (1984) Phase transitions and critical behavior in human bimanual coordination, *American Journal of Physiology*, **246**, R1000–R1004.

Kelso, J. A. S. & Scholz, J. P. (1985) Cooperative phenomena in biological motion. In *Complex systems: operational approaches in neurobiology, physical systems and computers* (H. Haken, editor). Berlin: Springer-Verlag.

Kelso, J. A. S. & Tuller, B. (1984) A dynamical basis for action systems. In *Handbook of cognitive neuroscience* (M. S. Gazzaniga, editor), pp. 321–356. New York: Plenum.

Kelso, J. A. S., V.-Bateson, E., Saltzman, E. & Kay, B. (1985) A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling, *Journal of the Acoustical Society of American*, **77**, 266–280.

Kugler, P. N. & Turvey, M. T. (1987) *Information, natural law, and the self-assembly of rhythmic movement: theoretical and experimental investigations*. Hillsdale, NJ: Erlbaum.

Labov, W. (1972) The internal evolution of linguistic rules. In *Linguistic change and generative theory* (R. P. Stockwell & R. K. S. Macaulay, editors), pp. 101–171. Bloomington: Indiana University Press.

Labov, W., Yaeger, M. & Steiner, R. (1972) *A quantitative study of sound change in progress*. Philadelphia, PA: The US Regional Survey.

Ladefoged, P. (1982) *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.

Ladefoged, P. (1990) Some reflections on the IPA, *Journal of Phonetics*, **18**, 335–346.

Lindblom, B. & Engstrand, O. (1989) In what sense is speech quantal?, *Journal of Phonetics*, **17**, 107–121.

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 & O. Dahl, editor), pp. 181–203. Mouton: The Hague.

Madore, B. F. & Freedman, W. L. (1987) Self-organizing structures, *American Scientist*, **75**, 252–259.

Manuel, S. Y. (1990) The role of contrast in limiting vowel-to-vowel coarticulation in different languages, *Journal of the Acoustical Society of America*, **88**, 1286–1298.

Manuel, S. Y. & Krakow, R. A. (1984) Universal and language particular aspects of vowel-to-vowel

coarticulation, *Haskins Status Report on Speech Research,* SR-77/7, 69–78.

May, R. & Oster, G. F. (1976) Bifurcations and dynamic complexity in simple ecological models, *The American Naturalist,* **110,** 573.

McCarthy, J. J. (1988) feature geometry and dependency: A review, *Phonetica,* **45,** 84–108.

Mowrey, R. A. & MacKay, I. R. A. (1990) Phonological primitives: Electromyographic speech error evidence, *Journal of the Acoustical Society of America,* **88,** 1299–1312.

Nearey, T. M. (1990) The segment as a unit of special perception, *Journal of Phonetics,* **18,** 347–373.

Ohala, J. J. (1983) The origin of sound patterns in vocal tract constraints. In *The production of speech* (P. F. MacNeilage, editor), pp. 189–216. New York: Springer-Verlag.

Ostry, D. J. & Munhall, K. (1985) Control of rate and duration of speech movements, *Journal of the Acoustical Society of America,* **77,** 640–648.

Pierrehumbert, J. (1990) Phonological and phonetic representation. *Journal of Phonetics,* **18,** 375–394.

Prigogine, I. & Stengers, I. (1984) *Order out of chaos.* New York: Bantam Books.

Roberts, E. W. (1975) Speech errors as evidence for the reality of phonological units, *Lingua,* **35,** 263–296.

Schmidt, R. C. (1988) Dynamical constraints on the coordination of rhythmic limb movements between two people. Ph.D. Dissertation, University of Connecticut.

Schmidt, R. C., Carello, C. & Turvey, M. T. (1990) Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people, *Journal of Experimental Psychology: Human Perception and Performance,* **16,** 227–247.

Schoner, G. & Kelso, J. A. S. (1988) Dynamic pattern generation in behavioral and neural systems, *Science,* **239,** 1513–1520.

Shattuck-Hufnagel, S. (1983) Sublexical units and suprasegmental structure in speech production planning. In *The production of speech* (P. F. MacNeilage, editor), pp. 109–136. New York: Springer-Verlag.

Shattuck-Hufnagel, S. (1987) The role of word-onset consonants in speech production planning: new evidence from speech error patterns. In *Motor and sensory processes of language* (E. Keller & M. Gopnik, editors), pp. 17–51. Hillsdale, NJ: LEA.

Shattuck-Hufnagel, S. & Klatt, D. (1979) The limited use of distinctive features and markedness in speech production: evidence from speech error data, *Journal of Verbal Learning and Verbal Behavior,* **18,** 41–55.

Stemberger, J. (1983) *Speech errors and theoretical phonology: a review.* Bloomington, Indiana: Indiana University Linguistics Club.

Stevens, K. N. (1972) The quantal nature of speech: evidence from articulatory-acoustic data. In *Human communication: a unified view* (E. E. David & P. B. Denes, editors), pp. 51–66. New York: McGraw-Hill.

Stevens, K. N. (1989) On the quantal nature of speech, *Journal of Phonetics,* **17,** 3–45.

Thompson, J. M. T. & Stewart, H. B. (1986) *Nonlinear dynamics and chaos.* New York: John Wiley & Sons.

Turvey, M. T. (1977) Preliminaries to a theory of action with reference to vision. In *Perceiving, acting and knowing: toward an ecological psychology* (R. Shaw & J. Bransford, editors). Hillsdale, NJ: LEA.

Turvey, M. T. (1990) Coordination, *American Psychologist,* **45,** 938–953.

Turvey, M. T., Rosenblum, L. D., Schmidt, R. C. & Kugler, P. N. (1986) Fluctuations and phase symmetry in coordinated rhythmic movements, *Journal of Experimental Psychology: Human Perception and Performance,* **12,** 564–583.

Turvey, M. T., Saltzman, E. & Schmidt, R. C. (in press) Dynamic and task-specific coordinations. In *Making them Move* (Badler, N. I., Barsky, Barsky. B. A. & Zeltzer, D., editors), pp. 157–170. Mountain View, CA: Morgan Kaufmann.

Vatikiotis-Bateson, E. (1988) *Linguistic structure and articulatory dynamics.* Bloomington: Indiana University Linguistics Club.

Vitz, P. C. & Winkler, B. S. (1973) Predicting the judged "similarity of sound" of English words, *Journal of Verbal Learning and Verbal Behavior,* **12,** 373–388.

Wood, S. (1982) X-ray and model studies of vowel articulation, *Working Papers, Lund Universite,* **23.**