

Catherine P. Browman

Louis Goldstein

Haskins Laboratories  
and Department of Linguistics,  
Yale University,  
New Haven, Conn., USA

## Response to Commentaries

We thank the commentators. Thinking about the problems they raised has helped sharpen the issues for us. In the response that follows, we will first briefly attempt to clarify some of general properties of our approach that seem to have been unclear, and then respond in more detail to various of the specific examples brought up in the commentaries.

### 1. General Properties of the Approach

#### 1.1. Articulation and Acoustics

Scully asserts that the task of the speaker is to achieve an auditory goal. Articulatory phonology does not take the goal to be auditory. Rather, it hypothesizes that the goal of the speaker's behavior is a particular organized ensemble of articulatory gestures. The acoustics (and auditory patterns) result lawfully from this gestural pattern. Indeed, a major point of the approach is to pursue the theoretical consequences of this hypothesis. Gestures, or to be more precise, tract variables, are purely articulatory in their definition, and, as currently defined, refer to local constrictions. Gestural terminology has been carefully se-

lected to reflect the purely articulatory nature of the gestures. Thus, an [f] is described as having constriction degree of {critical}, meaning that degree of constriction which, given the appropriate aerodynamic conditions, will typically result in the output of frication. Note that, if the appropriate aerodynamic conditions are not met, frication will not result. This is one of the clarifying features of the gestural approach – that it separates articulatory, aerodynamic, and acoustic descriptions of the speech event, and is explicit in its hypothesis of where the goals are specified. Perceptual and quantal effects may help to select the goals (the particular values of a gesture's dynamic parameters) that are lexicalized without changing the fact that what is lexicalized is purely articulatory.

In an articulatory phonology, as in most phonologies, structure and process are distinguished: in a sense structure might be thought of as the 'phonology' and process as the 'phonetics'. From this perspective, articulatory phonology can be seen as recasting the description of the structure in terms of units that are commensurate with a description of the process. That is, in the lexicon, gestures

and their organization/phasing/overlap are discrete and categorical (while also fully specified metrically). Gradience becomes relevant during speech production when the metric properties of these gestural events are subject to lawful variation. As argued in the position paper, this gradience can be seen in instances such as (some types of) positional allophonic differences and also in the reduction observed in casual speech. The movements associated with the gestures are affected, by definition, by processes occurring during the act of talking. These production processes as well as other sources of variability can lead to acoustic output that is then interpreted by the listener as one or another categorical pattern of gestures (possibly not the same pattern as the one produced by the speaker). As argued elsewhere, such an analysis permits a single phonological structure to characterize all instances of a given item, with casual differing from canonical utterances, and (some) positional and other allophones from each other, not in their discrete representations, but rather in the set of speech production processes that have applied to them. One way of thinking about this relationship, perhaps metaphorically, is that within a gestural framework articulatory and acoustic characterizations of speech are related as 'underlying' and 'surface' representations by constrained and general principles of the vocal tract structure and acoustics [see Liberman, 1970].

### *1.2. Duration; Higher-Level Units*

We agree with the conclusion implicit in some of Clements' commentary that the distinctive use of duration in articulatory phonology has not yet been adequately addressed. However, we do not agree that the restriction on possible contrasts of duration which is observable in languages is represented more naturally in a featural model than in a gestural

model. The convention in nonlinear versions of feature models that most length contrasts can be represented by associating one or two units of timing with a single feature bundle reflects a fact about how languages work, but it is not intrinsic to the particular representational scheme used. That is, it is perfectly possible logically to have associations between feature bundles and any number of timing units. Thus, just as possible contrasts currently need to be stipulated in a gestural phonology (although we hope that ultimately they would follow from dynamical and coordinative principles), constraints need to be stipulated in the featural approach, for example by ruling out structures with more than two associations or by declaring any structures with more than one association to be phonologically equivalent [see Steriade, 1990].

We would like to clarify the notion of 'internal duration' in the gestural approach. Each gesture is modeled as a dynamical system which is parameterized both spatially (equilibrium position) and temporally (stiffness). The stiffness parameter controls or characterizes the internal duration, or time base, of the gestural event as a whole. Thus, the internal duration is a unitary property, rather than a collection of 'temporally ordered points', although as Clements observes, various points (in the cycle defined by the dynamical system) can be identified 'for the purposes of stating phasing relations: onset of movement toward the target, achievement of the target, and beginning of movement away from the target'.

One difference between featural and gestural structures that Clements observes is that while there is just one way to link two features (they are linked or they are not linked), two gestures may be associated in a variety of ways, with different phases of the two gestures being synchronized. As Steriade [1990] discusses, the possibilities allowed in the ges-

tural model may, in fact, be phonologically relevant. For example, Steriade discusses one class of alternations that can be described very simply in terms of changes in relative timing of two gestures (which would presumably reflect some change in how they are phased), but which would be less satisfactorily described in terms of standard nonlinear featural structures.

In addition, it is not the case that, as Clements asserts, a 'gesture-based system has no way of expressing patterns of cohesion among different gestures'. The association of gestures into constellations exactly expresses patterns of cohesion among gestures. Omitting hierarchical nodes in the representation should definitely not be taken to mean that higher-level units are not permitted in the gestural approach; indeed, the discussions in Sections 2.2 and 2.3 of the position paper are either directly or indirectly about higher-level units. What a lack of hierarchical nodes does mean is that all higher-level units must always be accessed in terms of primitive units, e. g. gestures, and the relations among them.

### *1.3. Computational Model*

Various commentators pointed out missing components in the computational model. Sometimes these limitations are intentional, and sometimes they are the result of work being in progress. For example, the lack of an aerodynamic component, whose importance was stressed by Scully, is very much not an intentional limitation, but rather due to the fact that the computational model is still under development. The eventual incorporation of an aerodynamic component should allow the kinds of positional reductions in aerodynamic variables described by Kohler to be modeled.

An example of an intentional limitation is that pointed out by Kingston and Cohen, namely that the task dynamic model does not have a biomechanical component. While we

agree with Kingston and Cohen that a complete mechanical account of speech production would incorporate information about muscle activity and its neural substrate, the existing model has different goals (as they acknowledge). Indeed, we are not proposing a complete mechanical model of speech production, but rather are developing a functional model of those aspects of the speech production process that are directly relevant to specifying the phonological structure of utterances, and to accounting for the linguistically relevant contextual variation that can be observed for such structures. We believe that task dynamics is operating at the right level of description for this. Specifically, task dynamics provides (1) a dynamical definition of the hypothesized invariant speech units in an appropriate goal space, (2) a set of independent articulatory degrees of freedom that can be harnessed to achieve the task goals, and a model of how they are coordinated, and (3) a model of gestural interaction, or blending, that serves to predict the articulatory consequences of gestural overlap in time. While the adequacy and utility of the model might well be improved by using knowledge of the tongue's musculature to refine the set of independent articulatory degrees of freedom, any more detailed biomechanics would probably not be useful in pursuing the current goals (and would be untractable, for all the reasons Kingston and Cohen discuss). Moreover, we do not view this decision as limiting the model's explanatory potential. Task dynamics is embedded within an approach to coordinated action that emphasizes the importance of discovering the macroscopic, low-dimensional dynamical properties of an act [e. g., Turvey, 1990]. In the context of this approach, 'the developing nervous system is understood as harnessing dynamical laws underlying behavioral patterns, rather than being the ultimate cause of such patterns' [Beek and Bingham, 1991, p. 38].

## 2. Examples

### 2.1. Vowel Nasalization

The account of vowel nasalization in English that Kingston and Cohen attribute to Browman and Goldstein is not ours; here we will attempt to present a possible account.

Following Cohn's [1990] analysis, Kingston and Cohen suggest the possibility that vowel nasalization in English is sometimes phonological (in words such as 'sent') and sometimes phonetic (in words such as 'pen'). As will be seen below, an articulatory phonology account treats vowel nasalization as a unitary phenomenon, to be treated in the same way whenever it occurs, with the differences between 'sent' and 'pen' resulting from the differences in voicing. Here, we will simply note that we are puzzled by Kingston and Cohen's apparent identification of an articulatory phonology account with a phonetic account. Their reasoning appears to be that since gestures can be affected by gradient processes, they must perforce be 'phonetic' and not 'phonological'. However, articulatory phonology posits that gestures can be both 'phonetic' and 'phonological' [akin to SPE's positing that the same features occur in phonetic, phonological, and lexical representations, Chomsky and Halle, 1968]. As we have pointed out in the position paper and elsewhere, gestural structures have categorical properties in addition to the metric properties that allow them to be quantitatively scaled by processes ongoing during speaking (and by positional and stress variation). In fact, as discussed in the position paper, aspects of vowel nasalization are determined by categorical differences among gestural structures as a function of word, or perhaps syllable, position.

A plausible articulatory phonology account of the differences in vowel nasalization observed in 'sent' and 'pen' would combine the analysis proposed by Fujimura [1981] with the

generalizations uncovered by Krakow [1989]. Fujimura found, in comparing utterances such as 'pint' and 'pine', that there was very little difference between the two cases in the final alveolar closure gestures and even in the relationship between the final alveolar closure and vocalic gestures, despite acoustic differences such as the lack of a nasal murmur in 'pint'. The two articulatory differences reported (a glottal closure and a shorter temporal distance between onset and coda consonants in 'pint') were assumed by Fujimura to reflect general differences in gestural organization associated with tense vs. lax syllable codas (i.e., voiceless vs. voiced), independently of whether there is any nasalization in the coda. Krakow's generalizations, discussed in the position paper, predict that in both 'pint' and 'pine' the offset of velum lowering should be coordinated with onset of tongue tip/blade movement for the final alveolar closure; this is confirmed by Fujimura's displays. Since the interval between the release of the initial labial gesture and the onset of the tip movement is much shorter in 'pint' than in 'pine', due to the voicing difference, it automatically follows that full velum lowering is achieved much earlier in the acoustic vowel interval for 'pint'. This is consistent with Cohn's observations of earlier maximum nasal flow in similar utterances.

Thus, in this articulatory phonology analysis of the observed difference in vowel nasalization between words such as 'sent' and 'pint', on the one hand, and words such as 'pen' and 'pine', on the other hand, the difference requires neither a categorically different statement of velic coordination nor a gradient process. It is also not the case that one is 'phonological' and the other 'phonetic'. Rather, the difference follows automatically from two independently motivated principles, those governing how velic lowering gestures are coordinated with word-final oral constrictions [Krakow, 1989] and how initial and final

consonant gestures are coordinated as a function of final voicing or tenseness [Fujimura, 1981].

## 2.2. *Aspiration*

A major example considered by Kingston and Cohen involves aspiration. Couching aspiration and voicing in gestural terms has many advantages, as discussed in the position paper and elsewhere. In addition, developing gestural analyses facilitates drawing upon general principles of motor coordination to serve as potential explanations for the structure of phonological inventories. As regards aspiration, such movement-based explanations might serve to capture the widest generalization, which is that unaspirated voiceless stops are by far the commonest kind of stop [Maddieson, 1984; Keating et al., 1983]. In those cases in which voiceless unaspirated stops occur with a glottal opening-and-closing gesture, the timing of the glottal and supra-glottal gestures can be effectively synchronous [e. g., Benguerel et al., 1978, for French]; given roughly equal durations of the glottal and supra-glottal gestures, this would mean  $VOT = 0$ . In at least some low-degree-of-freedom systems, synchronicity is one of the most stable kinds of movement phase-locking [e. g., Kelso et al., 1987; Yamanishi et al., 1980]. Thus, to the extent that glottal opening-and-closing gestures do indeed occur in voiceless unaspirates (the relevant data are usually not available), the stability of synchronicity in motor behavior might explain the commonness of voiceless unaspirates.

For the less common cases in which the glottal and supra-glottal gestures are asynchronous, there is so far nothing in our understanding of human movement that explains the directional asymmetries (post-aspiration being more frequent than pre-aspiration). It is of course not clear whether our inability to provide a movement-based explanation reflects

inadequacies in our understanding or the lack of explanatory power of movement-based principles in this instance. Nevertheless, it could also be true that differential acoustic effects may lead to a preference for one kind of organization over the other, even though at least one hypothesis based on the primacy of bursts – Kingston's articulatory binding hypothesis – has been shown to be wrong by Kingston [1990] and Goldstein [1990].

Regardless of whether the preference for post-aspiration is acoustically or articulatorily motivated, it is still necessary to determine the nature of any points that are being coordinated. Discrete points occurring at or close to consonantal release, sometimes called 'landmarks', have been identified both articulatorily and acoustically [Stevens, in press]. Determining whether an acoustic or articulatory description of the landmarks is preferable can be difficult, since speech articulation and acoustics are lawfully related and therefore in most cases there will be no testably different predictions. However, since articulatory gestures can change their overlap, the relation between the articulatory and acoustic descriptions can change as the overlap changes. In this way, it becomes possible to separate articulatory and acoustic accounts. We have argued in the case of casual speech that, because gestures can occur even though they are imperceptible (i. e. 'hidden'), lexical representation is more invariably described in gestural terms. In the case of landmarks, if two gestures retain their pattern of coordination despite the presence or absence of a third gesture that changes (or eliminates) a potential acoustic landmark such as a release burst, then the landmark could only be described articulatorily, since the acoustic landmark would be missing 'or changed. A potential example of this kind, from the domain of velic coordination, was described in the previous section, in which for 'pine' and 'pint' the coordination of the velic

lowering and the alveolar closure gestures is invariant, despite other changes that result in the presence vs. absence of an acoustic nasal murmur.

### 2.3. *Retroflexes*

Kingston and Cohen argue that the data reported in Bhat [1973], showing interactions between the backness of a vowel and the retroflexion of a neighboring consonant, require gestural replacement as a form of allophonic variation. They further argue that the basis for the interactions is the acoustic similarity between adjacent consonants and vowels in the relevant languages. Here we attempt to show that neither of these conclusions is warranted. First, in general Bhat's examples appear to reflect not true allophonic variation, but rather selective (lexical) neutralization of contrast. The bulk of cases reported by Bhat fall into three categories: (1) backing (or centralization) of front vowels before retroflex consonants, (2) loss of a consonant retroflexion (i.e., retroflex consonants becoming alveolars or dentals) before front vowels, and (3) addition of retroflexion to consonants before back vowels. Most cases of types 1 and 2, and possibly the type 3 cases, appear to involve selective neutralization of contrast, rather than allophonic variation. That is, in these cases, the contrast between front and back vowels, or between retroflex and non-retroflex coronal consonants, is neutralized in the relevant environments. Within articulatory phonology, such neutralizations (occurring within lexical units) can be expressed as lexical constraints on gestural distribution, i.e., constraints on how gestures may combine into larger constellations. There is no question of one gesture being replaced by another.

Bhat's [1973] description of the type 3 cases might alternatively be consistent with viewing them as genuine allophonic processes, although it is impossible to tell without

more details. In such an analysis, the quantitative values of a tongue tip gesture's parameters (in particular the values of its constriction location and orientation) would be changed when overlapped by a back vowel gesture. Note that even this analysis, in which gestural parameters are changed, is far from being an arbitrary gestural replacement. That is, the gesture is still defined with respect to the same set of tract variables (and articulators); only the attributes have changed. If this analysis were shown to be necessary, however, it would indicate that one of our particular hypotheses is too strong, namely that a gesture is invariantly specified across all gestural contexts.

No matter how these facts are synchronically expressed, it is of considerable interest to determine why such patterns have developed in languages. Kingston and Cohen argue that the patterns have developed due to an acoustic interaction between the consonant and the vowel, because they find their proposed articulatory account to be inadequate. However, when considering the actual tongue movements involved in these gestures, and the articulatory degrees of freedom available to form these movements, a very natural articulatory account emerges.

Consider on the one hand that front vowels are produced by advancing the mass of the tongue through action of the genioglossus muscle. The contraction of the anterior fibers of this muscle (GGA) serves to front the central part of the tongue dorsum while simultaneously depressing the very front-most part of the tongue, including the tip. Such GGA activity can be seen in all front vowels and is not seen in back vowels [Baer et al., 1988]. On the other hand, retroflex consonants involve an extreme raising and retraction of the tip of the tongue, likely due to contraction of the superior longitudinal muscle (SL) [Hardcastle, 1976]. That is, the activities of the GGA and the SL are antagonistic in their effects on the

tip of the tongue and therefore the articulatory demands of front vowels and retroflex consonants are incompatible in this respect. Since consonant and vowel gestures typically overlap in time (and are so modeled in articulatory phonology), such incompatibility would be problematic, and could be remedied by avoiding overlapping front vowels and retroflex consonants (as would be the case in the neutralization of contrast described by Bhat).

#### 2. 4. Place Assimilation

Arguments made by Kohler and Clements identify a need for a less simplistic approach to assimilation. One type of example that both Kohler and Clements use involves assimilation of place for nasals in consonant sequences. Both examples question whether all assimilations can be due to sliding of unitary gestures. Kohler's examples question the temporal integrity of gestural units, while Clements' question the integrity of their spatial specification for both constriction location and degree. We will consider Kohler's data first, and then Clements'.

Kohler presents a number of examples in which the German suffix '-en' is assimilated, in that it is perceived as using the same articulators (or place of articulation) as the preceding stop. He argues that, since electropalatograph data show that there is no coronal closure in these cases, they cannot result purely from a change in gestural overlap. It appears from these examples, and also those of Barry [1991], that assimilation, viewed gesturally, cannot be restricted to changes in overlap and reductions in gestural magnitude. However, it seems as though these cases can be handled by hypothesizing that (durational) extension of gestures may also occur. In Kohler's example of German 'Beamten' pronounced [...mpm], this could mean that the labial closure gesture is extended in time and co-occurs with the following velic raising and lowering gestures.

Note that extension has been observed even when the neighboring (assimilated) gesture is *not* deleted [Barry, 1991]. Further research is required to determine if, like overlap and reduction, extension can be both categorical and gradient, and also to determine if there are specific conditions under which it can occur.

More generally, a growing body of data indicates that a number of quite different phenomena have apparently been lumped together under the rubric of 'assimilation', and therefore any framework that does not distinguish them fails a minimal criterion of descriptive adequacy. The phenomena can be described gesturally as follows: (1) temporal sliding, (2) magnitude reduction (in space and time), (3) temporal extension, or one of two combinations: (4) reduction plus sliding or (5) reduction of one gesture plus extension of another gesture. (The kind of assimilation represented in feature geometry by the technique of linking and delinking could refer to a subset of possibility (5), namely extension of one gesture [linking] and deletion of the other gesture [delinking]. A further possibility, that of gestural addition and/or replacement, could be described in the gestural framework, but does not draw upon gestural structure in any explanatory way). It has now been shown that more than the first two possibilities occur. To the extent that all the various possibilities not only occur but are correlated with other aspects of a language's phonological structure, any approach – either gestural or non-gestural – that does not permit such distinctions to be made is missing important generalizations.

Ohala [1990] has also hypothesized that many of the phenomena labelled 'assimilation' are likely to be different in nature. While viewing assimilation as a number of disparate phenomena might seem at first to be undesirable, we hope to show in the remainder of this section that doing so is descriptively necessary, and/or that it predicts differences in pho-

nological behavior that in fact occur, but are not accounted for in a more simplistic framework. For example, Barry [1991] described two different kinds of assimilatory behavior. In British English, when the utterance 'hand grenade' was produced at increasing speech rates, the coronal gesture at the end of 'hand' remained present but decreased in magnitude while the dorsal gesture at the beginning of 'grenade' extended in duration so as to co-occur with the coronal (type 5 above). In contrast, in Russian nasal-stop sequences, certain utterances showed reduction of magnitude of the coronal closure gesture (possibly accompanied by increased overlap) but no extension of the neighboring dorsal closure gesture (either type 2 or 4, above). Barry considers these differences in behavior between Russian and British English in light of other phonological differences between the languages. If it turns out to be possible to correlate the different types of assimilation with other aspects of phonological structure/behavior, then any approach that does not differentiate the various kinds of assimilations would miss valuable generalizations.

Assimilation examples were used by Clements to argue that place (articulator plus constriction location) and stricture (constriction degree) behave more independently of one another than would be expected on the basis of overlap of unitary gestures. In particular, he presents examples from Yoruba and Shona showing the assimilation of the place of a prefix-final nasal consonant to that of an initial stop or fricative of the following stem. In the Shona example, he shows palatographic evidence that the assimilated nasal consonant is produced with an oral closure, even when the stem begins with a fricative. Clements takes this to be evidence for the independence of articulator (which is shared between nasal and following obstruent) and CD (which is not). Here we turn our attention to how such cases

might be accounted for in a gestural framework, after sounding a precautionary note.

Before outlining a possible gestural account, we wish to emphasize that any proposed gestural analysis of the cases Clements presents is necessarily incomplete in the absence of knowing more about the gestural processes involved. For example, it is important to know whether there is any reduced oral constriction gesture associated with the prefix-final nasal consonant, distinct from the oral gesture beginning the stem, or whether the prefix truly contributes only a velic lowering gesture. In addition, it is important to know whether the oral constriction gesture (contributed by the stem) in the assimilated forms has increased duration compared to 'single' consonants in comparable positions. Such an increase in duration would be consistent with there being two distinct gestures involved, or with some kind of gestural extension.

Even without the above information, however, one possible attribute of a gestural account may be established, namely that the analyses of place assimilation could be different for the stop and fricative cases. Structures identified as assimilated in the case of stops could arise from the general gestural processes that we have hypothesized, for example from simple overlap of the velic lowering gesture with the following oral constriction gesture. However, as Clements observes, simple overlap will not produce the results observed in the fricative cases in these languages, because there is evidence for an oral closure gesture during the nasal. In an articulatory phonology, these cases of fricative assimilation would probably be described as involving two oral constriction gestures, sharing the same articulator set (more or less) but not the same constriction degree. The phonology would include, then, a specific statement 'licensing' structures with just these properties.

Such gestural analyses embody the pos-



sibility that place assimilation in nasals might involve different phenomena for stops and fricatives, with the case of stops but not fricatives resulting from independently motivated gestural processes. (Kohler also expects place assimilation in nasals to involve different phenomena for stops and fricatives, but for a different reason.) More generally, the articulatory phonology analysis (as well as Kohler's view) predicts that (nasal) place assimilation to fricatives should be a marked phenomenon, and should occur more rarely than assimilation to stops in languages of the world. In fact, a difference between place assimilation for nasals in the case of stops and in the case of fricatives is strongly supported by Padgett [1991]. He shows that nasal assimilation to fricatives is much rarer in languages than nasal assimilation to stops, and that languages adopt a variety of strategies to avoid assimilating nasals to fricatives. To handle this, and a number of other (unrelated) observations, he proposes an alternative feature geometry, in which stricture features are dependents of the articulator node, creating what he calls the 'articulator group'. This articulator group roughly corresponds to the gesture (minus its dynamic aspect).

Even given that the phenomena described as assimilation behave differently in the case of stops and fricatives, it is necessary to determine whether such a difference obtains within a given language, or just across languages as a whole. Two competing hypotheses seem most likely. On the one hand, place assimilation for nasals before stops might be the same phenomenon in all languages, and place assimilation before fricatives a different phenomenon. On the other hand, if a language has place assimilation in nasals before both stops and fricatives, the differences might have been levelled, presumably by adding an oral closure gesture (associated with the velic lowering gesture) in all cases: whereas a language with assimilation only before stops might use a dif-

ferent, gesturally motivated, type of assimilation. Here we will not attempt to support one or the other hypothesis, but merely point out that if either holds true, it is consistent with the gestural approach, and not predicted by a purely featural characterization in which place and stricture are independent of one another.

Finally, Clements also presents an argument for the potential independence of constriction degree (CD) and constriction location (CL) in vocalic assimilatory processes. Vowels assimilate in height (our CD) in Kimatumbi, but not in backness (related to our CL). The example is certainly suggestive of a problem in how vowels are described within the articulatory phonology model at present. However, the entire issue of the proper representation of vowels is lively but unstable at the moment, both among phoneticians and phonologists [cf. Fischer-Jørgensen, 1985; Ladefoged and Maddieson, 1990; Wood, 1982; Clements, in press]. A revision of vowel descriptions within articulatory phonology may be required, but currently seems premature.

### 2.5. Assimilation of Final Coronal Closures

Kohler uses the facts of German place assimilation, in which coronal stops can be assimilated in word- or syllable-final position, to argue that speech production processes must be under active auditory control. (We will use the term 'coronal' to refer to the class of tongue tip/blade constrictions that Kohler calls 'apical', so as to avoid the ambiguity of 'apical' in this sense vs. its meaning when contrasted with 'laminal'.) In discussing the cases in which assimilation fails to occur, Kohler suggests that the common denominator among coronal fricatives, released coronal stops, and (word- or syllable-) initial coronal nasals and stops, which all resist assimilation, is that they are each 'perceptually salient' in some way. We suggest the common denominator is much simpler: these are all 'not-X'.

where X = final coronal closing gesture. That is, the phenomenon of concern should be that word- or syllable-final coronal closing gestures are often assimilated. Final coronal closing gestures are distinguished articulatorily by their tendency to be reduced; as discussed in the position paper, assimilations often involve reduced gestures, so that the predisposition of final coronal closing gestures to be reduced probably contributes to their tendency to assimilate.

Three articulatory considerations (positional, anatomical, and functional) pick out final coronal closing gestures as candidates for substantial reduction in speech. First, there seems to be a general articulatory predisposition to reduce the magnitude of movements that occur later in a unit. Second, tongue tip movements seem particularly susceptible to reduction. Third, closing gestures have a different dynamical character than the more constrained gestures that produce the 'critical' constriction degree found in fricatives. We will examine each of these considerations in turn.

*Positional reduction.* In addition to the evidence discussed in the position paper on reductions of final gestures, there is a growing body of evidence that articulation is generally reduced over the course of a unit, possibly because of a general reduction in speaking effort or perhaps, as proposed by Gelfer [1987], the regulation of expiration. Evidence for different kinds of declination (sometimes when measuring vowels and sometimes consonants, and sometimes within syllables and sometimes within phrases) has been found for the intonation contour [e.g., Pierrehumbert, 1979; Cooper and Sorenson, 1981], the velum [Bell-Berti and Krakow, 1991], and the jaw [Machi, 1988; Vayra and Fowler, 1992; Gracco et al., in preparation]. This in no way denies the perceptual salience of initial portions of a word or syllable. However, while there surely

is a correlation between articulatory precision and perceptual salience, as Kohler suggests, the primacy probably lies in the articulation, given the pervasiveness of articulatory reduction across different articulatory systems, and across different sized domains. That is, under this hypothesis, initial gestures would have higher signalling value to listeners than final ones because they are reduced less (not vice versa).

*Tongue Tip Reduction.* We have recently observed a greater susceptibility to positional reduction for tongue tip movements in some data from an American English speaker. The peak vertical positions of x-ray microbeam pellets on the lower lip, tongue blade, and tongue dorsum were measured in the words 'pop', 'tot', and 'caulk' in the sentence: 'MY\_\_ huddles'. The peak was always higher for the initial consonant than the final one in the test word, but the magnitude of the difference was greater in the case of the coronal gesture (about 3.5 mm) than for the labial or dorsal gestures (about 1.4 mm). Note that this greater reduction for the tongue tip/blade gesture is found in an environment where there is no possibility of assimilation or, as traditionally described, flapping. Rather it seems to indicate a general tendency for more extreme reduction of final tongue tip gestures. In turn, it is possible to speculate that such reduction might result from the peripheral location of the tip on the surface of the tongue and the fact that it tends to rest on the lower teeth or floor of the mouth during vowels [see Browman, in press]. This positioning means that the tip has to travel further to form a closure than is the case for the tongue dorsum or lower lip, a pattern that is observed in the displacement data of Kuehn and Moll [1976]. The greater required distance could result in greater deviation from target.

*Closure vs. Critical Constrictions.* Constrictions that give rise to turbulence have a

relatively narrow range for constriction degree. Possibly as a result of this precision requirement, constriction gestures for fricatives have different dynamic properties from those that produce closures. For example, the data of Kuehn and Moll [1976] show that fricatives are produced with slower velocities than their homorganic stop counterparts, even when controlling for the size of the required displacements. While much more needs to be done to pinpoint the nature of the dynamical differences between closed and critical gestures, it seems reasonable that the increased precision associated with the critical gestures might render them less susceptible to reduction. This, in turn, could account for the failure of final fricative gestures to assimilate.

2. 5. 1. *'Beamte' vs. 'Beamten'*. We will argue here that Kohler's hypothesis that assimilation in German is effectively blocked by a perceptually salient type of sound is untenable. Specifically, Kohler argues for the role of acoustic stop releases in blocking assimilation by presenting pairs like German *'Beamte'* and *'Beamten'*, which differ in their susceptibility to assimilation. In *'Beamte'*, the [t] is never assimilated to the preceding [m], whereas in *'Beamten'* the [t] (and following [n]) can be assimilated ([bə'ʔampm]). Kohler attributes the differential behavior to the presence vs. the absence of an audible release of the tongue tip constriction, proposing that an audible release is perceptually salient, and that the only articulatory modifications possible are those that have small perceptual effects. However, there are cases in which the posited perceptual salience of audibly released stops fails to prevent their modification or even elimination.

As Kohler himself points out, both here and elsewhere [e. g., Kohler, 1990], stops that have an audible release (into a schwa vowel) in careful speech can lose this release in casual speech and then can also be assimilated. Kohler [1990] cites examples such as

*'Wagen'* ([va:gən] → [va:gn] → [va:gŋ]) and *'gegeben'* ([gəge:bən] → [gəge:bn]) → [gəge:bm]). If stops with clear acoustic releases have a high degree of perceptual salience, then the fact that the stops become unreleased argues against at least one of Kohler's claims: either it is not true that acoustic release has the relevant kind of perceptual salience, or it is not true that the only articulatory modifications allowable are those that do not affect perceptual salience very much. Thus, the lack of assimilation in *'Beamte'* cannot be attributed to the perceptual salience of the release.

We propose that the difference in behavior of *'Beamte'* and *'Beamten'* very likely originates in the predisposition for final reduction discussed above, and also in gestural overlap and its consequences. In *'Beamte'*, the coronal stop is neither word- nor syllable-final, so it would not be expected to undergo reduction. Now consider a non-assimilated version of *'Beamten'*, but one with increased overlap, such that the two coronal closure gestures (that of [t] and that of [n]) are now largely overlapping ([bə'ʔamtŋ]). The resulting tongue tip movement will likely be smaller in amplitude, due to the reduction of the final coronal gesture. Assimilation in *'Beamten'* may be reinforced by a kind of perceptual 'hiding' of the final coronal gesture akin to that which we have described in casual speech. That is, even in a non-assimilated (but increased overlap) version of *'Beamten'* ([bə'ʔamtŋ]), there is likely very little acoustic evidence of the coronal closure gesture. The movement of the tongue tip into closure presumably takes place while the lips are closed, and thus produces no acoustic consequences. The tongue tip movement out of closure occurs only after the end of the nasal murmur, which in utterance-final position is after phonation has ceased. In such cases, the nasal murmur contains the only acoustic evidence for the tongue tip constrict-

tion. However, listeners cannot identify place information for coronal nasal murmurs very accurately [for American English, close to chance level, Repp and Svastikula, 1988; see also Malécot, 1956]. Therefore, the presence of a tongue tip closure gesture at the end of 'Beamten' would not be very perceptibly different from its absence.

Without knowing more about the articulations involved, it is not possible to know whether there is a reduced coronal gesture present in the assimilated version of 'Beamten'. That is, even palatograph data showing lack of gestural closure cannot reveal either the presence or absence of a reduced gesture. And, as seen above, perceptual evidence for the place of a coronal nasal is not trustworthy. That is, a percept of an assimilated 'Beamten' ([bəʔampm]) might result even in the presence of a tongue tip closure gesture.

Nevertheless, let us assume, for the moment, that there is indeed no coronal gesture in at least one assimilated version of 'Beamten', and moreover, that the labial movement is longer than in the unassimilated version. In such a case, the question would then become whether this is the end result of a continuum, or whether this is simply another variant. If the posited assimilated version of 'Beamten' were the end result of a continuum, one would expect to find intermediate forms. Such forms should show varying degrees of overlap, akin

to 'gegeben' and 'Wagen', varying degrees of coronal reduction (with deletion the extreme case), and possibly also varying degrees of lengthening of the labial gesture (although the mechanism that would cause such gradient lengthening is not clear). If, however, the posited assimilated version of 'Beamten' were simply another variant, one would not expect to find intermediate forms. In such a case, the assimilated variant would presumably come from a listener-based sound change [Ohala, 1981] in which the listener misinterprets the gestural structure of 'Beamten' as having either two labial gestures, or a lengthened labial gesture, instead of a labial gesture and a coronal gesture, presumably influenced by the mechanisms of reduction and overlap as discussed above.

Notice that regardless of the nature of the assimilated form of 'Beamten', it is likely that overlap contributes to the account of the assimilations. Although at least in some instances overlap alone cannot account for the assimilations being considered, in general it would clearly be a mistake to eliminate overlap as an explanatory phenomenon in German assimilation, given that many (and perhaps all) of the assimilated cases show signs of increased overlap. That is, as overlap increases between two successive consonantal constrictions, first the schwa disappears and then place assimilation occurs.

## References

- Baer, T.; Alfonso, P. J.; Honda, K.: Electromyography of the tongue muscles during vowels in /apVp/ environments. *Ann. Bull. RILP* 22: 7-19 (1988).
- Barry, M. C.: Temporal modelling of gestures in articulatory assimilation. *Proc. 12th Int. Congr. Phonet. Sci.* 4: 14-17 (1991).
- Beek, P. J.; Bingham, G. P.: Task-specific dynamics and the study of perception and action: a reaction to von Hofsten (1989). *Ecol. Psychol.* 3: 35-54 (1991).
- Bell-Berti, F.; Krakow, R. A.: Velar height and sentence length: declination? 121th Meet. Acoust. Soc. Am., Baltimore 1991. *J. acoust. Soc. Am.* 89: 1916 (1991).
- Benguerel, A. P.; Hirose, H.; Sawashima, M.; Ushijima, T.: Laryngeal control in French stop production: A fiberoptic, acoustic and electromyographic study. *Folia phoniat.* 30: 175-198 (1978).
- Bhat, D. N. S.: Retroflexion: an areal feature. *Work. Pap. Lang. Universals* 13: 27-67 (1973).

- Browman, C. P.: Lip aperture and consonant releases: in Keating, Papers in Laboratory Phonology III (Cambridge University Press, Cambridge, in press).
- Browman, C. P.: Goldstein, L.: Tiers in articulatory phonology with some implications for casual speech. *Haskins Lab. Stat. Rep. Speech* 92: 1-30 (1987).
- Browman, C. P.: Goldstein, L.: Some notes on syllable structure in articulatory phonology. *Phonetica* 45: 140-155 (1988).
- Browman, C. P.: Goldstein, L.: Articulatory gestures as phonological units. *Phonology* 6: 201-251 (1989).
- Browman, C. P.: Goldstein, L.: Gestural structures: distinctiveness, phonological processes, and historical change: in Mattingly, Studdert-Kennedy, Modularity and the motor theory of speech perception, pp. 313-338 (Erlbaum, Hillsdale 1991).
- Chomsky, N.; Halle, M.: The sound pattern of English (Harper & Row, New York 1968).
- Clements G. N.: Place of articulation in consonants and vowels: a unified theory: in Laks, Rialland, L'architecture et la géométrie des représentations phonologiques (Editions du CNRS, Paris, in press).
- Cohn, A. C.: Phonetic and phonological rules of nasalization. *UCLA. Work. Pap. Phonet.* 76 (1990).
- Cooper, W. E.; Sorenson, J. M.: Fundamental frequency in sentence production (Springer, New York 1981).
- Fischer-Jorgensen, E.: Some basic vowel features, their articulatory correlates, and their explanatory power in phonology: in Fromkin, *Phonetic linguistics: essays in honor of Peter Ladefoged*, pp. 79-99 (Academic Press, New York 1985).
- Fujimura, O.: Elementary gestures and temporal organization - What does an articulatory constraint mean?: in Myers, Laver, Anderson, *The cognitive representation of speech*, pp. 101-110. (North-Holland, Amsterdam 1981).
- Gelfer, C. E.: A simultaneous acoustic and physiological study of fundamental frequency declination: unpubl. diss., City University of New York (1987).
- Goldstein, L.: On articulatory binding: comments on Kingston's paper: in Kingston, Beckman, *Papers in laboratory phonology I: between the grammar and physics of speech*, pp. 445-450 (Cambridge University Press, Cambridge 1990).
- Gracco, V. L.; Fowler, C. A.; V.-Bateson, E.; Romero, J.: Global correlates of stress accent in spoken sentences (in preparation).
- Hardcastle, W. J.: *Physiology of speech production: an introduction for speech scientists* (Academic Press, London 1976).
- Keating, P.; Linker W.; Huffman, M.: Patterns in allophone distribution for voiced and voiceless stops. *J. Phonet.* 11: 277-290 (1983).
- Kelso, J. A. S.; Schoner, G.; Scholz, J. P.; Haken, H.: Phase-locked modes, phase transitions and component oscillators in biological motion. *Physica Scripta* 35: 79-87 (1987).
- Kingston, J.: Articulatory binding: in Kingston, Beckman, *Papers in laboratory phonology I: between the grammar and physics of speech*, pp. 406-434 (Cambridge University Press, Cambridge, 1990).
- Kohler, K. J.: Segmental reduction in connected speech in German: Phonological facts and phonetic explanations: in Hardcastle, Marchal, *Speech production and speech modelling*, pp. 69-92 (Kluwer, Dordrecht 1990).
- Krakow, R. A.: The articulatory organization of syllables: a kinematic analysis of labial and velar gestures: unpubl. PhD diss., Yale University, New Haven (1989).
- Kuehn, D. P.; Moll, K.: A cineradiographic study of VC and CV articulatory velocities. *J. Phonet.* 4: 303-320 (1976).
- Ladefoged, P.; Maddieson, I.: Vowels of the world's languages. *J. Phonet.* 18: 93-122 (1990).
- Lieberman, A. M.: The grammars of speech and language. *Cogn. Psychol.* 1: 301-323 (1970).
- Macchi, M.: Labial articulation patterns associated with segmental features and syllable structure in English. *Phonetica* 45: 109-121 (1988).
- Maddieson, I.: *Patterns of sounds* (Cambridge University Press, New York 1984).
- Malecot, A.: Acoustic cues for nasal consonants. *Language* 32: 274-284 (1956).
- Ohala, J. J.: The listener as a source of sound change: in Masek, Hendrick, Miller, *Papers from the parasession on language and behavior*, pp. 178-203 (Chicago Linguistic Society, Chicago 1981).
- Ohala, J. J.: The phonetics and phonology of aspects of assimilation: in Kingston, Beckman, *Papers in laboratory phonology I: between the grammar and physics of speech*, pp. 258-275 (Cambridge University Press, Cambridge 1990).
- Padgett, J.: *Stricture in feature geometry: unpubl. PhD diss., University of Massachusetts, Amherst (1991).*
- Pierrehumbert, J.: The perception of fundamental frequency declination. *J. acoust. Soc. Am.* 66: 363-369 (1979).
- Repp, B. H.; Svastikula, K.: Perception of the [m]-[n] distinction in VC syllables. *J. acoust. Soc. Am.* 83: 237-247 (1988).
- Steriade, D.: Gestures and autosegments: comments on Browman and Goldstein's paper: in Kingston, Beckman, *Papers in laboratory phonology I*, pp. 381-397 (Cambridge University Press, Cambridge 1990).
- Stevens, K.: Phonetic evidence for hierarchies of features: in Keating, *Papers in laboratory phonology III* (Cambridge University Press, Cambridge, in press).
- Turvey, M. T.: Coordination. *Am. Psychol.* 45: 938-953 (1990).
- Vayra, M.; Fowler, C. A.: Declination of supralaryngeal gestures in spoken Italian. *Phonetica* 49: 48-60 (1992).
- Wood, S.: X-ray and model studies of vowel articulation. *Work. Pap. Lund Univ.* 23 (1982).
- Yamanishi, J.; Kawato, M.; Suzuki, R.: Two coupled oscillators as a model for the coordinated finger tapping by both hands. *Biol. Cybernet.* 37: 219-225 (1980).