Coarticulation as a Component in Articulatory Description

Katherine S. Harris

Coarticulation in Conventional Descriptions

In the recent past, the speech pathologist was often given a course in "articulatory phonetics." This study had as its goal teaching the student to make a series of alphabet-like symbols on a piece of paper, which, if the training was successful, would enable the student to perform such tricks as to read aloud in the "dialect" of the original speaker. Indeed, in the academic setting where this form was most highly developed--London University, the home base of Henry Sweet--these methods were used to change speech patterns not only of countless cockneys but also of the many non-native speakers of English who swarmed to London in the great days of the British empire. Of course, Sweet was the historical model for the hero of the play Pygmalion and the musical My Fair Lady (Borden & Harris, 1980).

In such training schemes, it was routinely assumed that there was no great difficulty about producing an adequate representation of the detailed act of speaking from alphabet-like marks on the page in which the only representation of time was the indication of visual succession (Lisker, 1974). Even now, it may be debated whether our knowledge of the principles of alphabetic writing is what underlies a belief in the adequacy of the symbol-by-symbol representation of speech, or whether, alternatively, the principles of alphabetic writing depend on some property of the perceptual system that makes such a representation seem adequate. Whichever formulation one prefers, there is a long history of a relationship between the study of phonetics and the desire of various authors, at various times, to commit oral narratives to writing. For example, as long ago as the 12th century, an Icelandic scholar wrote the "first grammatical treatise," an attempt to rework the orthography of Roman writing to suit the demands of representing the sounds of his native tongue (Fischer-Jörgenson, 1975).

The assumption that a series of symbols is an adequate representation of a child's articulation is one of the two basic assumptions of the typical course taken by the speech pathologist. The other is that, by listening, the transcriber can infer articulation or, at least, that aspect of articulation that is frequently all that the course provides—a schematic lateral view of the steady-state position of the articulators, to be associated with the left-to-right alphabetic labels of transcription.

†Also Graduate School, City University of New York.

Acknowledgment. This work was supported by NINCDS Grant NS-13617 and NIH Biomedical Research Grant RR-05596 to Haskins Laboratories.

While the skills of a well-trained phonetician to reproduce speech are often astonishing to the layman, it is not clear what information is being used in performing. Even very well-trained phoneticsians may not do a very good job of judging the articulator position associated with a given phone. For example, Ladefoged (1967) has shown that London-trained phoneticsians cannot accurately assign tongue positions to the "cardinal vowels" produced by their London-trained colleagues. (The cardinal vowels are a reference system of articulator positions that give a kind of grid for vowels.) Indeed, it is his contention that vowels are sorted into categories on the basis of acoustic rather than articulatory similarity. In part, the phonetician's difficulty in making articulator position inferences is the inevitable result of the asymmetry of the relationship between acoustics and articulator position. Theoretically, although the acoustic signal can be estimated from a sufficiently detailed knowledge of vocal tract shape, a given acoustic signal may be associated with any of an infinite number of vocal tract shapes. An amusing example is provided by Ladefoged (Ladefoged, Harshman, Goldstein, & Rice, 1978). He shows two lateral views of the vocal tract. In one view, the vocal tract has a physiologically sensible contour. In the other, the tongue appears to have been creased into pleats. The two shapes are acoustically indistinguishable.

Figure 1. Two vocal tract shapes which generate the same formant values. Reproduced from Ladefoged, P., et al., 1978, op. cit.
In many of the more recent versions of our hypothetical course in articulatory phonetics, it is suggested that students learn to transcribe in "feature" notation, although a number of alternative descriptions that fit this rubric have been proposed (e.g., Chomsky & Halle, 1968; Ladefoged, 1971; Singh, 1978). It is not my interest here to attack the value of feature descriptions in principle. Within speech pathology as a field, they are useful in describing such diverse phenomena as the confusion matrices generated by hearing-impaired listeners in speech perception studies (Bilger & Wang, 1976) and the transfer of training in articulation correction (Compton, 1976; Pollack & Rees, 1972). The classic feature description is temporally isomorphic with a phonological description. The feature description, in its most sophisticated form (Chomsky & Halle, 1968), was developed to capture certain kinds of generalization within linguistics, such as morphophonemic alternation rules; the fact that the features have a physiological referent is not, in principle, an issue within the generative phonology framework. From the point of view of temporal structure, the features are abstract and timeless in the same sense as the units they were designed to replace.

The picture of speech production that our hypothetical student might infer, then, would be that the act of speaking proceeds from steady state to steady state, with (since the articulators must move continuously) some uninteresting events between, and that the articulatory origins of the steady state events are fairly transparent.

For many members of the research community, the sheer conspicuousness of the dynamic, as contrasted to the static, characteristics of the speech signal was first revealed by the illustrations in the book Visible Speech (Potter, Kopp, & Green, 1947), when it appeared shortly after the Second World War. The book represented, in many ways, the culmination of efforts by the Bell Telephone Laboratories to execute a mission inherited from Alexander Graham Bell himself. Bell had an interest both in the visual representation of speech and in using this representation to aid the deaf in learning to talk (Borden & Harris, 1980; Bruce, 1973). The attitude taken by Potter, Kopp, and Green towards the temporal structure is an interesting one, given their pedagogical purpose; one must learn to recognize the "characteristic position" or "hub," and the coarticulatory influences on it. While mention was (necessarily) made of the time-varying nature of the pattern, they said almost nothing about the time course of events as characteristics of speech sound representation. In other words, they took a segmental approach, although the dynamic aspects of the pattern were quite conspicuous.

It is a mistake to suppose that phoneticians whose main work preceded the sound spectrograph were wholly unaware of temporal phenomena, although these phenomena fit uneasily into any transcriptional description. For example, diphthongs are conventionally transcribed with two symbols, although their dynamic character was recognized. Jones, Sweet's successor, said: "For the purpose of practical language teaching it is convenient to regard a diphthong as a succession of two vowels, in spite of the fact that, strictly speaking, it is 'a gliding sound" (Jones, 1956, p. 99).

Earlier phoneticians were also well aware of the consequences to articulator movement; articulator position for one sound might influence that for a temporally adjacent one. This is the phenomenon called assimilation by Jones. For a common example, in the pronunciation of these shoes in ordinary speech, the /s/ /f/ sequence is reduced to a single tongue movement to provide a suit-
able position for /ʃ/. However, phoneticians were unaware of the extent to which the phenomena described above were not special examples; that is, since articulator position changes continuously, context sensitivity is the rule, rather than a phenomenon to be explained in special cases. Much of the effort for the following decade in the study of both speech production and speech perception was to build theories to account for the mismatch, in perception and production, between transcriptional phonetics and the phenomena of speech production. Theories in this field may be divided into two broad classes—we might call them discrete and continuous.

Theories of Speech Production

In this section, we will discuss some fairly recent speech production and perception theories. To a certain degree, these theories were aimed at rationalizing transcriptional or perceptual simplicity in the face of acoustic or articulatory variability. As noted above, the theories are of two basic kinds: discrete and continuous.

As an example of a discrete model, one might choose Perkell's model of the speech production process (Perkell, 1980), which is, in turn, based on Stevens' quantal model (1973). Without going into the details of the model shown in Figure 2, it can be seen to have stages such that the input, at the top of the figure, is a series of segments (S₁, S₂, S₃, and S₄) with each segment specified by a feature matrix, which is transformed into an isomorphic sensory goal. In the output, due to various hypothesized mechanisms, the boundaries between segments are no longer perpendicular lines, so that the "motor goals" and the segments are no longer isomorphic. This model is very like the one proposed by Henke (1966) to explain coarticulation, which will be discussed below. Two points should be noted: the only representation of time in the input is a simple succession, as in transcription, and the effect of reorganization is to desynchronize the representations of the transcriptional units.

An alternative point of view, although in very primitive form, is represented by Liberman's motor theory (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The motor theory is designed to account for the finding that two acoustic synthetic speech patterns will both produce the perceptual impression of the same consonant /d/, coupled with two different vowels (see Figure 3). Apparently the percept depends in some rather direct way on the dynamics of the acoustic pattern. The motor theory assumed that the listener must perform some operation dependent on the articulatory dynamics of production. This is a continuous theory because the dynamics of the pattern are important in themselves. It should be pointed out, however, that while the motor theory can be described as a continuous theory, Liberman has produced a stage model of the production-perception process that is quite similar to Perkell's (Liberman, 1970), and that Perkell, while in this classification a discrete theorist, has produced an extremely elegant discussion of articulatory dynamics from a quantal theory perspective (Perkell & Nelson, 1982).

It should be noted here, as well, that there is an apparent dichotomy between theories with some kind of linguistic referent, as discrete, and theories with some kind of motor referent, as continuous. This dichotomy is certainly not a necessary one. Thus, Fowler has argued (1977) that although symbols for speech may be represented in a particular form on a page, this does not mean that their motor representations take the same form in the ner-
Figure 2. A figure showing Perkell's model of the translation of a feature matrix representation into articulatory units. Reproduced from Perkell, J., 1980, op. cit.

Figure 3. Two patterns perceived as /d/, followed by different vowels. Reproduced from Liberman, A., 1970, op. cit.
Harris: Coarticulation as a Component in Articulatory Description

vous system. It is possible to argue that a representation of a motor plan in which the speech act is conceived as present somewhere in the nervous system, stripped of its temporal properties, which are then added in the execution, is very like the conception of speech as a phonological string.

It is important when examining theories of coarticulation in detail, as we shall do below, to recognize that the study of coarticulation is merely a small part of the study of skilled movement. Speech is special, as a type of skilled movement, in some rather unfortunate ways. For one thing, as we discussed above, speech comes with a notation scheme developed for special purposes, which may lead us astray when we attempt a more physiological description (Moll, Zimmermann, & Smith, 1977). Speech comes, as well, with a very inaccessible set of independent variables, as most articulators are difficult or impossible to observe without special techniques. However, even if experimental data on the movement of the articulators were easily gathered, one could not develop a theory of coarticulation simply by turning to a formulation lying ready-made in, for example, robotics. Machines can be produced that will mimic particular acts, but machines cannot now be designed that will adapt to a wide variety of changed environmental conditions, as humans do (Kelso, 1981). Furthermore, while we know a great deal about the muscular and neurological structures that participate in movement, the increase in our knowledge of structure does not help us very much with respect to function. For example, although a recent review chapter (Matthews, 1981) testifies to the explosion of our knowledge of the microstructure of the muscle spindle, a specialized device that provides feedback information about movement, the basic behavioral questions we ask about movement today are not very different from those we asked in the early 1930s, when Bernstein began his studies of the coordination of gait (summarized in Bernstein, 1967) or, perhaps, even when Sherrington summarized his observations of the decerebrate cat (Sherrington, 1906). We still lack a comprehensive theory that explains why skilled movements can be scaled up and down in timing, what causes the resistance of movement patterns to disruption by environmental change, and, with reference to coarticulation, why the elements of skilled movement patterns can be so freely reassembled to form novel sequences. While we have theories of coarticulation, as we will see below, they can rather easily be shown to fail. In what follows, I will attempt to outline the proposals for a model of coarticulation and to show how existing data succeed or fail in supporting them.

Hypotheses about Organizational Units and Speech Planning

Coarticulation as conventionally described is but one of a number of phenomena indicating some kind of organizational cohesiveness in speech. A great deal of effort has been directed at defining the outer bound over which such organizational cohesiveness exists. Unfortunately, the larger the unit that has been investigated, the larger the unit over which organizational dependencies can be demonstrated. For example, Lehnert has shown evidence for paragraph cohesiveness over units that are larger than sentences. Speakers apparently signal first and last sentences in paragraphs by a number of means. The initial sentence in a paragraph is often signaled by high fundamental frequency, the last sentence by low fundamental frequency and laryngealization. There are, in addition, durational cues for the termination of paragraphs, although the way duration is used is language dependent (Lehnert, 1975, 1979, 1980a, 1980b). It may be that the question of the outer bound of such effects is not a meaningful one. However, even if the absolute outer bound of such effects is indeterminate, we can ask what these effects tell us.

24
Underlying an interest in many of these phenomena is what Monsell and Sternberg (1982) have called the utterance program hypothesis. "Certain basic assumptions [about theories of speech production] seem to be widely shared among psycholinguists, linguists, and other students of speech. One such is the claim, explicit or implicit, that the motor events of an utterance are controlled by the execution of a plan or program--an integrated and relatively detailed description of the utterance (or a large part of it) constructed as a whole before the utterance begins. We term this claim the utterance program hypothesis" (p. 1). The utterance program hypothesis has been used in explanation of coarticulation itself and in connection with discussions of related phenomena, such as declination and slips of the tongue. By considering the latter kinds of phenomena first, we can perhaps clarify our discussion of coarticulation theories, which follows.

First, however, we note that many discussions of speech motor plans are circular. An observation is made that speech is normal in one circumstance and abnormal in another. The difference is attributed to the correct or incorrect functioning of a motor plan. For example, our typical student of speech pathology has heard that the articulation difficulty of some populations is due to the failure of a motor plan. The important thing to note is that the invocation of the motor plan adds nothing to the behavioral observation that the population does not speak normally (Kelso & Saltzman, 1982).

A somewhat veiled version of this circular kind of argument is one in which the naked motor plan is given some kind of neuroanatomical or neurophysiological clothing. For an example outside of speech, the control of many kinds of rhythmic activity, such as walking, has been ascribed to the behavior of neural oscillators (Gallistel, 1980), which are not independently observed. While one might not wish to return to the kind of anathema on physiological theorizing dictated by Skinner (1938), it is important to recognize that one of the motivations for his prohibition still holds--there is no explanatory power in the restatement of an observation in different language, even when the language has an independent prestige. Thus, we find out nothing about an aphasic's speech by saying that it is due to the malfunctioning of a particular neural circuit unless we are experimentally prepared to launch a search for the circuit or unless what we know from other sources about a neural circuit of the proposed type allows us to make inferential predictions that we can test about the resultant behavior of aphasics.

A related problem with the metaphor of the motor plan has been raised by Turvey and his associates--it is that the existence of behavioral system activity does not require a single controlling mechanism that lies at a particular level in the nervous system and specifies in detail the properties to be controlled (Turvey, 1977), and, indeed, there are logical problems with the whole idea of a single control center. It may be that some of the characteristics of motor control, which have been attributed to the operation of a plan, are properties of the motor system itself, which emerge as it behaves. Thus, the fact that bees build hexagonal honeycombs does not mean that the bee has a hexagonal floor plan in his central nervous system--rather, the honeycomb may arise in its hexagonal form as a consequence of the interactional properties of the bee and his environment, as the honeycomb is constructed.

Given that we observe carefully these prohibitions on how much we attribute to motor plans, let us return to what we know about the temporal organization of speech. We will talk largely, but not entirely, about precursory ef-
ferts—that is, the effects that indicate anticipation of speech output before it occurs. While precursory effects tell us nothing about their causes, they tell us something about relevant temporal domains.

An important and much studied phenomenon is the slip of the tongue, in which exchanges occur between elements in speech. A famous example was produced by William Spooner, an English clergyman, who once said, "You've hissed my mystery lectures" in place of "You've missed my history lectures" (Fromkin, 1971). Slips of the tongue are important to a discussion of coarticulation for several reasons. The first, and most important, is that the primary sublexical exchange units seem to be elements very close to the single phonemic segment (Shattuck-Huffnagel, 1983). Apparently, these phonemic segments are correctly produced for their new positions. The existence of such shifts is probably the best evidence we have of the existence of a premotoric terminal stage in the speech production process (MacNeilage, Hutchinson, & Lasater, 1981). Apparently, the units that shift adapt to their new positions—that is, they are correctly coarticulated with their neighbors. Thus, even though we cannot precisely define the phonemic unit in such a way that we can isolate it in the speech stream, slips of the tongue provide some evidence that a phoneme has reality as an action unit. It is interesting to note that although phones participate as action units, single features do not, evidently, appear in exchange error units (Shattuck-Huffnagel & Klatt, 1979).

A final point may be made about slips of the tongue. The sphere over which they occur appears to be of the general length of a breath group. This is roughly the temporal domain of declination and, perhaps, of durational interaction, but it is substantially longer than the temporal extent over which conventional coarticulation spreads.

Another recently fashionable bit of evidence for speech motor planning is the so-called declination phenomenon—the tendency of utterances to decline in fundamental frequency from onset to termination. This is at the utterance level, an analog of the phenomenon studied by Lehtst, and cited earlier, that the onset of the sentence that comes first in a paragraph is higher in fundamental frequency (F0) than the onset of sentences in later positions. Figure 4 is a fairly typical example of sentence declination. Historically, this tendency has been characterized in two ways: as a terminal fall (Lieberman, 1967) and as declination (Maeda, 1975) through the utterance. Again speaking historically, it has been unclear whether the relevant phenomena should be conceived as localized at the end of the sentence, or as distributed throughout. Given that intonation is almost always studied in the context of syntactically complex and phonetically variable contexts, an experimentally clean decision between these alternatives has been difficult, but at least present thinking favors the declination description. That is, the downdrift in F0 appears to run through the sentence, rather than being localized at the end. A related question is whether the mechanism is passive or active. A passive mechanism would be one in which the generalized downdrift is a simple consequence of some physiological given. It might, for example, be a consequence of an uncorrected tendency for subglottal pressure and, hence, F0 to fall throughout the course of an utterance. An alternative would be that the shape of the fundamental frequency contour, regardless of its proximal physiological cause, is a consequence of active planning of the whole utterance. It has been suggested that the latter picture of events is correct because of the utterance length effect—the tendency of F0 to begin at a higher value in longer utterances. In a "speech planning" point of view, a speaker may begin the contour at a higher level in order to come out in the same place.
Figure 4. A figure showing declination in a complex "read" sentence. Reproduced from Cooper, W. S., and Sorenson, J., op. cit.

Figure 5. Declination in some sentences with one or two stress peaks. To appear in Gelfer, C., Collier, R., Harris, K. S., and Baer, T. Is declination actively controlled? In I. Titze (Ed.), Vocal fold physiology: Physiology and biophysics of voice. Iowa City: Iowa University Press, in press.
Harris: Coarticulation as a Component in Articulatory Description

Figure 5 shows F0 contours from a recent experiment by Gelfer, Collier, Harris, and Baer (1983). The contours were produced in reiterant speech—that is, the speaker mimicked himself producing a more complex utterance with the syllable *ma*. The utterances varied in stress placement and in length. Such structurally simple utterances produce simple fundamental frequency contours made up of one or more peaks. The initial peak varied in amplitude depending on the sentence length, but the effect was very small. From the point of view of speech planning, there were reliable precursive effects, which appear to reflect an overall rough schema for the utterance. However, notice that the whole utterance was not reorganized, depending on its length. Whatever utterance length effects are shown by the declination contour are small and localized. The domain of the effects, however, is the utterance—a domain of about the same size as that for slips of the tongue.

We can say, then, that although speakers may demarcate organizational units of greater length, the longest units over which there is evidence of planning, in the form of precursive effects, is a unit of the general length of a phrase. The examples given here involve slips of the tongue and declination. Similar material could be provided for unit duration. We turn now to conventional coarticulation, which operates over a far smaller temporal domain—on the order of a few speech segments.

Theories of Coarticulation

"Extrinsic Timing" Theories

Since classic theories of coarticulation spring from classic representations of phonological units, such theories almost by necessity attempt to represent coarticulatory phenomena themselves as essentially timeless. In the acoustic real world, no clear boundaries are seen between segments as conventionally defined. Furthermore, acoustic segments are context sensitive; therefore it is necessary to develop some theory that mediates between the acoustic representation and the (presumed) underlying units. Typical examples of such theories are Henke's look-ahead model of coarticulation (Henke, 1966) and Daniloff and Hammarberg's canonical forms model (Daniloff & Hammarberg, 1973). However, other examples of such models can be cited as well; the models as a class were discussed in more detail in a very thorough review several years ago (Kent & Minifie, 1977). Here, we will merely discuss a very well-known example, the Henke model, and refer readers to the review for more detail.

The Henke model assumes that all phonological units can be represented as bundles of features, which occur, in canonical form, as successive units along a time axis. Each phoneme has a specified value, zero, plus, or minus, for each feature. In forming articulatory sequences, the speaker performs an articulatory scanning operation on the phonemes arrayed in a buffer for output. If a feature is unspecified (that is, has a zero value) for several phones preceding the phone for which it is specified, then the feature will be anticipated during the intervening phones; that is, the intervening phones will assume the same feature value as the upcoming one.

Thus, in a sequence of a spread and a rounded vowel separated by nonlabial consonants, the consonants will assume the rounding feature. The test of this thesis has been to ask speakers to produce utterances like *once true* (Daniloff & Moll, 1968) and then to examine the sequence for the time of
the onset of rounding. Using tests of this sort, evidence has been produced that anticipatory coarticulation may spread over as many as four or five segments (Benguerel & Cowan, 1974; Daniloff & Moll, 1968). The model has also been used to explain the early onset of velar lowering in sequences of vowels concluding with nasal consonants. Presumably, in English, vowels are unspecified for nasality; hence, when they precede nasals, they become nasalized.

Success or failure of the model in explaining the data depends on its two assumptions—first, that coarticulatory spread is timeless, and, second, that whatever feature description is made of phones is adequate. For example, Kent and Minifie argue that while the Henke model is assumed to explain the findings of Benguerel and Cowan, it cannot explain the occasional spread of rounding to the end of the preceding spread vowel. One way of giving a "quick fix" to the theory is by relaxing the feature description requirements. Thus, one might assume that the end of a spread vowel is not specified for rounding. A second and more plausible approach is to give up the assumption of timeless- ness in speech production.

A Temporal Theory

In the light of a "coproduction" theory by Fowler (1980), discussed below, Bell-Berti and Harris (1981) proposed a temporal mode of coarticulation as a substitute for feature-based models. Because Fowler's theory has been somewhat enlarged since it was originally presented, we will discuss the Bell-Berti and Harris view first.

In brief, it was Fowler's thesis that current theories fail to make an appropriate recognition of the temporal dimension in speech production itself. Thus, a theory of anticipatory coarticulation that fails to acknowledge the time course of articulation will fail. She suggested, as an alternative to the view that static elements of vowel and consonant productions are exchanged, that vowel and consonant are coproduced.

A simple model of anticipatory coarticulation, then, makes three propositions (Bell-Berti & Harris, 1981): First, the articulatory period of a segment is longer than its acoustic period; second, for a given articulator, the period of anticipation is temporally independent of preceding phone string number, provided there is no articulatory conflict; and third, that articulatory period may begin at different times for different articulators.

These propositions were tested using electromyographic techniques for anticipatory coarticulation of lip rounding (Bell-Berti & Harris, 1979, 1981, 1982). The test is quite simple. If anticipatory coarticulation is segment based, then its onset will vary with the number of segments; if it is time based, then the duration of anticipatory coarticulation will be independent of the number of segments in a string, provided they do not themselves block coarticulation. Therefore, in order to provide a test, speakers were asked to produce utterances of the type [1C_n u], with a variable number of consonants in intervocalic position. Typical results are shown in Figure 6; the onset of lip-rounding, that is, the duration of the anticipatory period, is independent of the number of anticipatory segments, or of their durations, except for the single voiceless stop condition /itu/.
Figure 6. Onset of lip rounding in a series of minimal pairs. Part A shows electromyographic signals from the indicated phone string. Part B shows onset of separation between corresponding pairs. Presented in a paper entitled "Temporal organization of speech units over changes in stress and speaking rate" at the Tenth International Congress of Phonetic Sciences, Utrecht, 1983.
The Bell-Berti and Harris study was repeated, in part, by Sussman and Westbury (1981). They examined anticipatory coarticulation in the strings /kikstu/, /kakstu/, /tiku/, and /taku/. They found different onset times for all four utterances by electromyographic measures, although all differences were not statistically significant. A repeat of the experiment using strain gauge measures found no differences between /kikstu/ and /tiku/. They argued that the failure to find identical onsets for /kikstu/ and /tiku/ or for /kakstu/ and /taku/ argues strongly against time locking of coarticulation to the vowel. They also point out that anticipatory coarticulation is earlier for strings in which the first vowel is /i/ than those for which it is /a/. Their suggestion, in explanation of the latter finding, is that the rounding following /i/ begins earlier because of a necessity to counteract the biomechanical forces that spread the lips for /i/. They support neither the anticipatory scanning model nor the temporally locked model, although their loo-ahead scanner model is segment based.

Both their results and ours point strongly to one experimental issue, noted above; that is, that articulatory constraints are unpredictable from the feature specification of phones. The two vowels /i/ and /a/ are, in feature specifications for rounding, respectively minus and zero. Yet rounding onset time is affected in a manner that is contrary to the feature prediction.

The deviance of the data point for the /itu/ sequence is less conspicuous in the Harris and Bell-Berti data than the /iku/ sequence in the Sussman and Westbury paper, since the latter authors are plotting a two point continuum. On the assumption that the sequences are equivalent in the two experiments, a possible explanation of the deviance of the onset for rounding for single intervocalic stop sequences is provided by Engstrand (1983). He pointed out that relaxation of rounding has been shown to occur in the sequence /itu/ (Gay, 1978; Harris & Bell-Berti, 1983). He suggests that /t/-burst release may be incompatible with a fully rounded lip position. If this is so, then the lips must move rapidly from a fully rounded position, for /u/, to a partly rounded position for the preceding string. In sequences of the form /itu/, full rounding must be suppressed rapidly. For all other sequences (/is tu;/ /ist stu/, etc.), while full rounding must end relatively close to the consonant release, partial rounding can end anywhere in the preceding string. The general principle expressed is that production of dentals is incompatible with full rounding and compatible with partial or no rounding. We would, then, expect both the onset and the time course of rounding to be important in a full theory of coarticulation.

Two final experiments on anticipatory lip rounding may be cited—by Lubker (1981) and by Lubker and Gay (1982). The first, by Lubker, gives unequivocal support to the view that the onset of lip rounding varies with the length or duration of the preceding consonant string. The second shows individual differences in the form of the function relating the electromyographic onset of rounding to number of consonants in an intervocalic string. However, this study did not examine consonant string duration but merely consonant number.

In all of the above, we have concentrated on the anticipatory coarticulation of lip rounding. It should be pointed out that there is a similarly detailed, and almost as confused, literature on anticipatory nasalization. Indeed, Al-Bamerni and Bladon (1982) suppose that there may be two forms of anticipatory nasalization—one time locked and one variable. However, this seems a heuristically unsatisfactory solution.
In reading through this account of a series of experiments with their disparate results, the reader should be forgiven for some feeling of bewilderment. It may be worthwhile to consider what we do know. First, it is clear that conventional feature descriptions of phones are not strong enough to predict the details of their articulation, either spatially, that is, in terms of their detailed articulatory topology, or temporally, in terms of when one articulatory gesture begins with respect to another. At present, we are not sure why there are experimental differences among investigators. The only present solution seems to be a more thorough investigation, using simultaneous electromyographic, acoustic, and movement techniques.

Coarticulation and compensatory shortening. Fowler's comments on extrinsic timing theories of speech production have been cited above. However, the theory is far richer and more complex than we have indicated. It was developed, in part, as a means of explaining perceptual isochrony, the phenomenon that syllables perceived as being of more or less equal duration are systematically unequal. Some of its principles form a general theory of production.

Fowler assumes, following Ohman (1966) and Perkell (1969), that vowels and consonants are coproduced so that neighboring segments overlap; i.e., a consonant is produced while a vowel is being produced. The speaker can use such a strategy because vowels and consonants are different kinds of units. Succeeding vowels are produced as slow changes in the position of the tongue body in the mouth. Consonant production is more localized, may involve a partially non-overlapping set of muscles, and is superimposed on the continuous vowel-to-vowel movement of the tongue. Unstressed vowels are presumed not to interrupt the trajectory from one stressed vowel to the next. This model has both spectral and temporal consequences. Let us first consider the temporal consequences.

It has been shown, very often, that the measured duration of a vowel shortens as increasing numbers of consonants are added to it (for a review see Lehiste, 1970). There are backward shortening effects reported as well; that is, a vowel shortens as increasing numbers of consonants precede it (e.g., Lindblom & Rapp, 1973). However, backward shortening (that is, effects of consonants on succeeding vowels) is much the smaller effect. The effects of unstressed vowels on stressed vowels are analogous to the effects of consonants on vowels—for example, the stressed vowel in easy is shorter than in easily. In Fowler's model, the reason for the shortening is the articulatory overlap produced by coproduction.

The same mechanism produces spectral coarticulation. If an unstressed vowel is preceded or followed by a stressed vowel, it should coarticulate with it. Indeed, coarticulatory and shortening effects are but two measures of the same thing and should be highly correlated (Fowler, 1981). Fowler's test of the prediction shows usually significant correlation but some failures in the detailed prediction, apparently due to peculiarities of the particular experimental paradigm.

This theory does not make any predictions about lip rounding, because it is concentrated on the vocal tract manifestations of coproduction, which was the example used by Ohman. It is hard to believe, however, that some parts of the system operate on different principles than others. Furthermore, the model does not cover the well-known shortening effects of consonants on other consonants (Hawkins, 1973). Perhaps its most serious shortcoming, however, is
that it does not deal with competing articulation—the circumstance in which the articulators are constricted during consonant production so that free vowel-to-vowel coarticulation cannot take place. For example, Recasens (1983) has shown that in Catalan, vowel-to-vowel spectral coarticulation in vowel-consonant-vowel (VCV) disyllables varies systematically with the extent to which the intervening consonant engages the blade region of the tongue and, consequently, makes coarticulation physically impossible. If Fowler's theory were literally correct, one would expect that differences between VCV sequences in the extent to which they can be coproduced would be accompanied by corresponding differences in the amount of possible compensatory shortening.

Coarticulation and Context Sensitivity

The laboratory investigations discussed above are, perhaps, of interest to the speech pathologist in terms of what they can tell him or her about the practical problems of helping a client to improve a misarticulated sound. What, if anything, have we learned that is relevant?

It is the common observation that certain phonetic environments facilitate correct sound production; for example, Curtis and Hardy (1959), in a now classic paper, showed that some allophones of /r/ are more often correctly produced than others by misarticulating children. As Kent said, "An optimistic interpretation of this contextual facilitation is that some phonetic environments facilitate correct sound production and this facilitation can be exploited to clinical advantage" (Kent, 1982, p. 66). The limits on contextual generalization as a teaching strategy are entirely outside the province of this paper. However, what we can say something about, as a consequence of this brief review, is the task facing the child in learning to talk and the investigator in attempting to specify the contexts that may be relevant subjects of investigation. There are at least two factors that we will need to learn more about:

1. Relative production variability. The first section discussed the insecurity that an observer should feel in making inferences about the articulatory details of production from perceptual judgment. The observer is right, by definition, in judging a child's production to be correct. What he or she cannot do is to infer the articulation from the acoustics, the effects of perceptual factors on his criterion, or the nature of the articulatory error when the speaker is judged to be wrong. Even with respect to the variability of the acoustic signal for a given phonemic percept in a given environment, it is obvious that there is more acoustic production variability in some environments than in others. Some contextual effects may be contextual effects on listener criterion. For example, the formant values for correct stressed vowels are less variable than for unstressed vowels (Summers & Soli, 1982). A more often studied case is /s/, a phone that is notoriously difficult for children and also notoriously subject to contextual inconsistency (Mazza, Schuckers, & Daniloff, 1979). It may be that part of the contextual variability is associated with criterion variability, rather than articulatory variability.

2. Context specification. A lesson to be learned from the literature on coarticulation is that a decision to consider sounds as dividing into allophonic classes leads to balkanization. However, it is questionable whether House's (1981) suggestion that improved transcription may lead to better accounts of context sensitivity will help. A sound can be shown to be differ-
ent in endless ways, depending on factors both within and without the transcriptional record. In truth, we do not know what contexts form natural classes.

It has now been shown repeatedly that children learn phones in words, without uniform generalization across all environments (e.g., Macken, 1980). These types of context sensitivities must have some significance for practical decisions about contexts important in defining a class of phones. On the other hand, certain kinds of context sensitivity are apparently not part of the learning process in children nor are they stored as separately learned patterns in adults. The demonstration that two productions are acoustically or gesturally different does not tell us whether or not the two members form a natural class. It is only careful study of the natural variability of children's articulation, coupled with better assessment of what constitutes motor equivalence and cohesiveness in the adult, that will allow us to make progress in this difficult field.

References


