424

Journal of Experimental Psychology: General
1983, Vol. 112, No. 3, 386-412

Converging Sources of Evidence on Spoken and Perceived Rhythms of Speech: Cyclic Production of Vowels in Monosyllabic Stress Feet

Carol A. Fowler

Journal of Experimental Psychology: General
1983, Vol. 112, No. 3, 386-412

Converging Sources of Evidence on Spoken and Perceived Rhythms of Speech: Cyclic Production of Vowels in Monosyllabic Stress Feet

Carol A. Fowler
Dartmouth College and Haskins Laboratories, New Haven, Connecticut

SUMMARY

The article reviews the literature from psychology, phonetics, and phonology bearing on production and perception of syllable timing in speech. A review of the psychological and phonetics literature suggests that production of vowels and consonants are interleaved in syllable sequences in such a way that vowel production is continuous or nearly so. Based on that literature, a hypothesis is developed concerning the perception of syllable timing assuming that vowel production is continuous.

The hypothesis is that perceived syllable timing corresponds to the timed sequencing of the vowels as produced and not to the timing either of vowel onsets as conventionally measured or of syllable-initial consonants. Three experiments support the hypothesis. One shows that information present during the portion of an acoustic signal in which a syllable-initial consonant predominates is used by listeners to identify the vowel. Compatibly, this information for the vowel contributes to the vowel's perceived duration. Finally, a measure of the perceived timing of a syllable correlates significantly with the time required to identify syllable-medial vowels but not with time to identify the syllable-initial consonants.

Further support for the proposed mode of vowel–consonant production and perception is derived from the literature on phonology. Language-specific phonological conventions can be identified that may reflect exaggerations and conventionalizations of the articulatory tendency for vowels to be produced continuously in speech.

To their speaker/hearers, both naive (Donovan & Darwin, 1979; Lehiste, 1972) and expert (Abercrombie, 1964; Classe, 1939; Pike, 1945), languages sound rhythmical. The term *rhythm* as applied to speech refers generally to an ordered recurrence of strong and weak elements. In this general sense, languages clearly are rhythmical: Consonants and vowels approximately alternate, and, in stress languages such as English, so do stressed and unstressed syllables. However, attempts to validate the intuition that speech is rhythmical have focused on recurrence defined temporally—in particular, on the question of whether the regular recurrence of certain spoken units is isochronous.

Three classes of rhythm have been proposed for languages; stress timing (English, Swedish), syllable timing (Spanish, Italian, French), and mora timing (Japanese). In rhythmical utterances a unit of speech—the stress foot, the syllable, or the mora—is said to be regulated temporally, so that onset–onset intervals between units are approximately isochronous.[1] In a stress-timed language, for example, intervals between onsets of stressed syllables are said to approach isochrony, even though some intervals may be monosyllabic and others di- or trisyllabic (e.g., Abercrombie, 1964; Catford, 1977; Classe, 1939; Pike, 1945).

The bases for linguists' and other listeners' impressions of isochronous rhythms in speech are unknown. However, with the possible ex-

[1] A foot is a unit of metrical structure in speech consisting of a strong syllable and one or more weak syllables. In English, the weak syllables of a foot always follow the strong syllable. A mora is a "light" syllable (i.e., a short vowel optionally preceded by a consonant), or it is part of a "heavy" syllable. A heavy syllable consists of a syllable-initial consonant, if any, a long vowel or a short vowel, and a postvocalic consontant; it is two morae in length.

ception of mora timing in Japanese (e.g., Han, 1962; Dalby & Port, Note 1), it is known that the basis is not acoustic isochrony, or, in stress-timed languages, even near-isochrony, of the intervals that have been proposed as relevant. English is probably the most studied language in this regard, and many researchers have reported large departures from measured acoustic isochrony of stress feet in spontaneous (Lea, Note 2; Shen & Peterson, Note 3) and more constrained (Classe, 1939; Lehiste, 1972) utterances.

It is unlikely, then, that any units of naturally produced speech are *realized* isochronously. In view of that, the interesting questions to ask now are where the impression of rhythmicity comes from, whether recurrence of any of the units of speech that do recur is perceptually significant, whether it is linguistically significant, and whether it is articulatorily significant. Evidence bearing on these questions is derived from research reported in the psychological literature and the linguistics literature on phonetics and phonology. This article and one following (Fowler, Note 4) are intended to bring together these research lines and thereby to assess the state of our understanding of spoken and perceived rhythms of speech.

The two articles in the series differ in scope. This article considers only monosyllabic utterances in which all syllables are stressed (e.g., from Bolinger, 1965: "Pa made John tell who fired those guns"). The reason for this narrow focus is that fairly extensive but disparate lines of research—in psychology relating to perception, in phonetics concerning articulation, and in phonology concerning structure in sound sequences—converge to suggest a coherent perspective on rhythmic speech production and on perception of rhythmic speech in an idealized stress-timed language where feet are monosyllabic. Less extensive lines of research provide a less coherent picture of production and perception of speech where unaccented syllables are produced. This latter literature is the subject of the second article.

In this article, discussion is limited also in a second way. Initially, I consider ways in which talkers comply with *instructions* to produce stress(syllable)-timed speech and the ways in which listeners assess those productions. Before it is possible to draw realistic conclusions concerning rhythms that may or may not underlie production of spoken languages, and before we can ascertain whether the impression of rhythm is realistic or illusory, it is imperative that we learn how to recognize rhythm in speech when it occurs.

I first review the literature concerning production and perception of sequences of monosyllabic stress feet. The literature under review suggests two conclusions, one concerning the production of vowels in fluent speech and one concerning their perception. These proposals are tested in a series of three experiments.

In the second part of the article, I introduce evidence from the linguistics literature on phonology that may converge with the experimental evidence reviewed or presented in the first part. In the second part, I attempt to introduce and defend three basic ideas. One is the general idea that direct investigation of linguistic structure can provide a useful source of converging evidence with that provided by experimental investigations of language use. The second is the more specific idea that some phonological rules can be identified as exaggerations and conventionalizations of articulatory dispositions, and as such, can provide converging evidence for the identity of dispositions. Third, I attempt to identify several instances of phonological rules that are "natural" (i.e., reflect articulatory dispositions) if the manner of vowel production proposed in the first part of the article is in fact an articulatory disposition.

In the final part of the article, conclusions are drawn from the array of findings reviewed and presented in the first two parts.

This research was supported by National Science Foundation Grant BNS 8111470 and by National Institute of Child Health and Human Development Grant HD 16591-01 to Haskins Laboratories. I thank Alan Bell and Gary Dell for their comments on drafts of this article.

Experiment 1 was carried out in collaboration with Louis Tassinary and has been summarized in Fowler and Tassinary (1981).

Requests for reprints should be sent to Carol A. Fowler, Department of Psychology, Dartmouth College, Hanover, New Hampshire 03755.

Monosyllabic Stress Feet

The Perceptual Evidence and Some Articulatory Correlates

Several years ago, Morton, Marcus, and Frankish (1976; Marcus, 1981) reported a systematic discrepancy between the measured timing of a sequence of digits and its perceived timing. In particular, they found that sequences of digits with acoustically isochronous onset–onset intervals sound unevenly timed to listeners. Given an opportunity to adjust the intervals between digits until the timing sounds isochronous, listeners introduce systematic departures from measured acoustic isochrony. This finding is almost complementary to one reported by Lehiste (1972) and others (Donovan & Darwin, 1979) on listeners' perceptions of sentential rhythms. This literature (reviewed by Fowler, Note 4) reports that listeners may *fail* to detect departures from measured isochrony in spoken sentences. Although this latter collection of studies is interpreted as revealing listener insensitivity to foot durations, the findings by Morton et al. (1976) cannot have that interpretation. Indeed, taken together, the two sets of findings suggest that listeners' impressions of speech timing are not based on the same intervals measured by investigators. This was the interpretation offered by Morton et al. of their own findings.

An investigation of talkers' productions of isochronous sequences suggests one important difference between measured and perceived rhythmic intervals. In particular, the latter but not the former sometimes can be identified with rhythmic articulatory intervals (Fowler, 1979; Fowler & Tassinary, 1981). Asked to produce isochronous sequences of monosyllables, talkers produce sequences with just the measured departures from isochrony that listeners require in order to hear the sequences as evenly timed (Fowler, 1979).

This research indicates that talkers' and listeners' notions of rhythmicity in speech agree but differ from those of experimenters. Such a pattern of agreement and disagreement invites two interpretations. One is that talkers and listeners are subject to an illusion that experimenters, working on visible rather than audible displays of speech, evade. Another is that talkers produce rhythmic speech on request in these studies and listeners recognize it as such. For their part, experimenters fail to detect the rhythmicity because their experimental measurements somehow fail to reflect the natural structure of the spoken sequences. The latter is the more conservative of the two views because it ascribes no special processes or behaviors to listeners and talkers. The talker is assumed simply to follow instructions, and the listener, to detect the natural structure of the acoustic signal provided by the talker. In addition, this interpretation appears a realistic one in view of the well-known difficulties involved in the measurement of speech because it is coarticulated.

From the perspective of this second interpretation, assessments of the rhythmic structure of naturally produced speech sequences will be inaccurate until experimenters discover what counts as rhythmicity for talkers and listeners. This best can be determined, to begin with perhaps, by studies in which talkers are asked to produce sequences with specified timing and in which their performances are examined.

In the study by Fowler (1979), talkers produced sequences consisting of a pair of rhyming consonant–vowel–consonant (CVC) syllables in alternation (e.g., /bad sad bad . . ./). In these sequences, talkers produced long intervals between measured acoustic onsets of syllables when the first syllable in the interval began with a long-duration prevocalic segment. Indeed the departures from measured isochrony of successive intervals could be predicted very closely from differences in the measured durations of the syllable-initial consonants. Figure 1 displays the relationship found in Fowler's (1979) article. The onset–onset time differences in these productions ranged from a minimum of about 35 msec for sequences such as /mad nad . . ./, in which initial consonants were similar in manner class, to a maximum of about 200 msec when consonants differed in manner and in other features (e.g., /bad sad . . ./).

Although measured vowel onsets tend to be aligned more evenly than onsets of acoustic energy for the initial consonants of the syllables, intervals between vowel onsets are

not isochronous either; instead they show departures from isochrony complementary to those of syllable onsets.

Articulation may be isochronous in these productions, however. When monosyllables in a sequence are rhyming CVCs, measures of intervals between onsets of muscle activity involved in segment production have revealed isochrony both of initial consonant and of vowel-related muscle activity. This is found even in sequences showing substantial departures from measured acoustic isochrony (Tuller & Fowler, Note 5). For example, in a sequence /bak fak bak . . ./, electromyographic (EMG) activity of the orbicularis oris muscle involved in lip closure was found to be isochronous; this implies that lip closures for /b/ and /f/ also were isochronous in these utterances. Necessarily, however, acoustic intervals from stop release for /b/ to onset of frication for /f/ were shorter than the opposite intervals from frication to release. This departure from isochrony of acoustic-energy onsets follows from the timing relation between the consonant articulations and their acoustic correlates. Consonants are produced in three broad phases: a closing phase, a closure interval, and a release phase. During the closure interval for the stop consonant /b/, the lips are shut, and in stressed, syllable-initial position, the interval is silent. The stop burst occurs on release of the closure in the final phase of consonant production. In contrast to the stop consonant /b/, the fricative /f/ has a noisy closure interval. During closure, the lower lip approximates the upper teeth but does not seal off the oral cavity to the passage of air. Air passing through the narrow constriction produces frication. Consequently, a talker who aligns closure phases of syllable-initial stops and fricatives will produce syllables with systematically anisochronous onsets of acoustic energy.

These studies suggest, then, that talkers comply with instructions to produce isochronous monosyllables by producing isochronous *articulations*. Intervals between onsets of acoustic energy for successive monosyllables, then, are anisochronous because different manner classes of consonants have nonsilent acoustic consequences at different times after articulatory onset. Talkers do not attempt to compensate for this anisochrony of acoustic-energy onsets. For their part, in these experiments, listeners only hear isochrony when articulation is isochronous. They hear uneven timing when acoustic-energy onsets of different manner classes of consonants are aligned. I conclude, therefore, that in these experiments

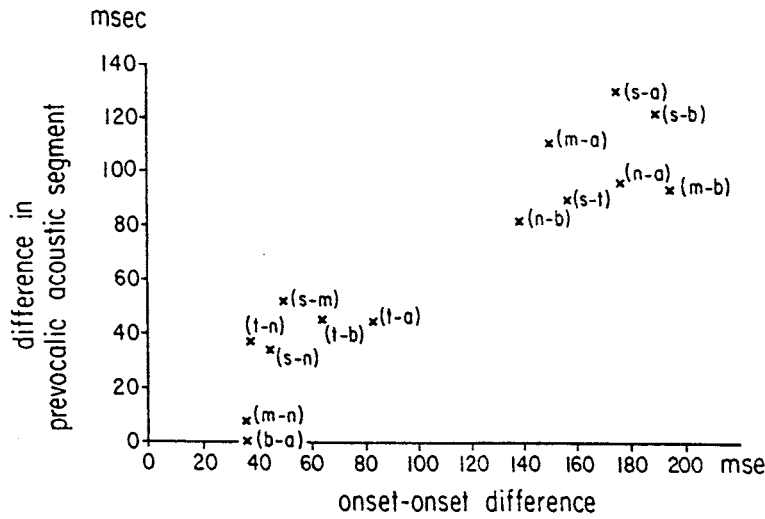

Figure 1. Differences in duration of prevocalic acoustic energy in syllables produced in alternation (Fowler, 1979) plotted as a function of syllable onset–onset asynchrony. (Data is from a single talker instructed to produce the syllables evenly stressed and timed. Paired letters on the figure refer to syllable-initial segments. For example, [s–a] refers to utterance /sad ad . . ./.)

listeners' perceptions of the rhythmic structure of speech are based on their extraction of acoustic information specifying articulatory timing (cf. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). This conclusion is compatible with that based on other evidence (e.g., Fitch, Halwes, Erickson, & Liberman, 1980; Lehiste, 1970). For example, listeners' judgments of the relative loudness of two vowels correspond more closely to the articulatory effort required to produce them than to their relative intensities (Lehiste, 1970).

The conclusion that perceived timing is produced timing does not tell the whole story, however. The experiment by Tuller and Fowler (1980) found isochrony of both consonant- and vowel-related muscle activity. A later experiment (Fowler & Tassinary, 1981) showed that initial consonants are not always articulated at isochronous intervals in sequences that talkers intend to be isochronous. Figure 2 displays measurements of a set of syllables produced in time to a metronome by three talkers (see Rapp, Note 6, for similar data on Swedish talkers, and Allen, 1972a, 1972b, for analogous data on English obtained using a different procedure). The location of the metronome pulse in the CVCs is indicated by the vertical line at zero in the figure. Points generally to the left of the metronome pulse indicate the onset of acoustic energy of the syllable. Points generally just to the right of the pulse indicate the measured vowel onset, and points farther to the right indicate measured vowel offset. By showing the alignment of rhyming syllables with the metronome pulse, the figure also reveals how syllables are aligned in relation to one another. The figure shows the effect reported by Morton et al. (1976) and studied further by Fowler (1979) and by Tuller and Fowler (1980). Acoustic-energy onsets for fricatives are early relative to those for voiced stops. Of interest here is another finding, however. Acoustic-energy onsets of intervals beginning with consonant clusters are early relative to others. A talker producing the sequence /sad strad sad . . ./ in time with the metronome does not produce isochronous acoustic onset–onset times—as he or she would if /s/ production were initiated at temporally equidistant intervals. Consequently, whatever the

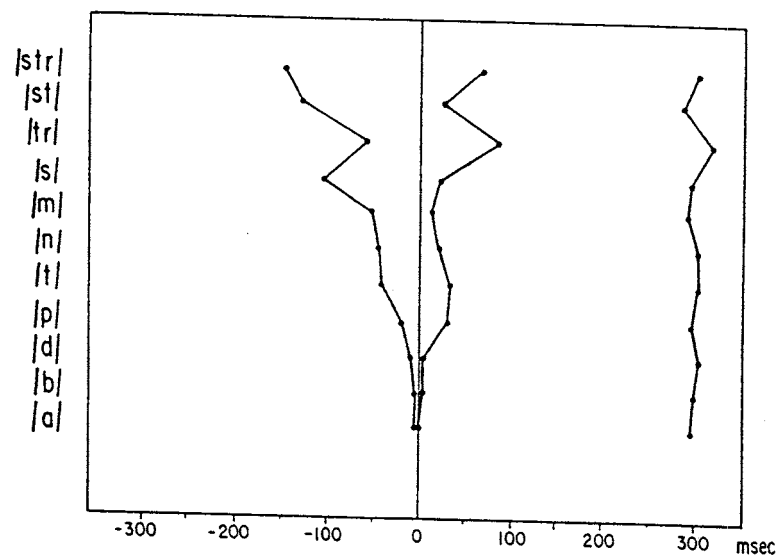

Figure 2. Measures of syllables produced by talkers in time with a metronome. (The vertical line at zero represents the metronome pulse. Different syllables are plotted top to bottom in the figure. The points generally to the left of the line represent the onset of acoustic energy for each syllable relative to the metronome pulse. Points generally just to the right of the pulse represent the measured vowel onset, that is, the onset of voiced oral formants for the vowel. Points to the far right represent measured vowel offset—the beginning of closure for final /d/. From "Natural Measurement Criteria for Speech: The Anisochrony Illusion" by C. A. Fowler and L. Tassinary, in J. Long and A. Baddeley (Eds.), *Attention and Performance, IX.* Hillsdale, N.J.: Erlbaum, 1981. Copyright 1981 by Erlbaum. Reprinted by permission.)

talker may have been producing rhythmically in these utterances, it was not initial-consonant production.

The alignments are not related to the amplitude contours of the syllables (Morton et al., 1976; Tuller & Fowler, Note 5) or, apparently, to their fundamental frequency contours (Rapp, Note 6).

In this study, the only acoustic measure temporally equidistant from the metronome pulse and consequently isochronous in these productions was the measured vowel offset. This finding perhaps can be rationalized by examining two separate research lines that investigate the temporal and articulatory microstructure of syllables: studies of phonetic shortening and of coarticulation.

The Temporal and Articulatory Microstructure of Syllables

Figure 2 reveals a pattern of vowel shortening in the context of various syllable-initial consonants. This pattern of shortening has been reported by other investigators for other languages (e.g., Lindblom, Lyberg, & Holmgren, 1981). In Figure 2, the measured duration of the vowel shortens as that of the prevocalic consonant or consonants increases in duration. Figure 3a replots the shortening effects in Figure 2 beside others (3b) reflecting effects of syllable *final* consonants on vowel duration.[2] These data resemble those reported by Lindblom et al. on speakers of Swedish and show that a vowel's measured duration also shortens as syllable-final consonants are added to the syllable.

Two interpretations of the shortening effects suggest themselves. According to one, talkers attempt to maintain a constant syllable duration in production (e.g., Shaffer, 1982). This might be a manifestation of a syllable- or stress-timing tendency. If for whatever reason talkers *are* trying to maintain a constant syllable duration, however, they are unsuccessful, as Figure 2 reveals. An examination of the articulatory evidence suggests a different interpretation.

In syllables, the production of consonants and vowels is context sensitive, usually in an assimilative way. The context sensitivity, called *coarticulation*, occurs very generally in syllables (e.g., MacNeilage & DeClerk, 1969).

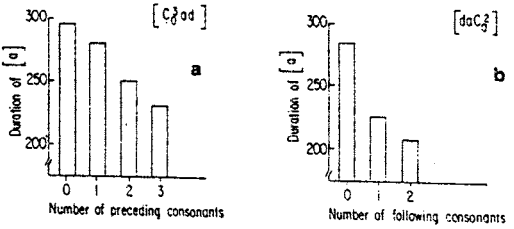

Figure 3. Measured vowel shortening in the context of preceding (a) and following (b) consonants in English.

For example, closure for a /b/ followed or preceded by the close vowel /i/ is achieved with a more closed jaw than that for /b/ followed or preceded by the open vowel /a/ (Sussman, MacNeilage, & Hanson, 1973). Similarly, the place of articulation of /k/ is fronted in the context of a front vowel as compared with a back vowel (e.g., Perkell, 1969).

Coarticulation has various explanations in the literature. One explanation, first proposed by Öhman (1966), appears to account for the vowel-shortening effects just described as well as for the context sensitivity of segment production. Öhman proposed that syllable-initial and -final consonants are superimposed on a vowel's leading and trailing edges. Moreover, in a VCV disyllable, vowel-to-vowel gestures of the tongue body are produced somewhat separately from articulatory gestures for the consonant. Öhman's evidence for his rather counterintuitive view of disyllable production was meager, but it has been substantiated by several subsequent studies. His evidence was derived from acoustic measures of implosive and explosive formant transitions in VCV disyllables produced by a Swedish talker. In Öhman's data, implosive transitions, representing the closing phases of voiced stop production, were affected by both vowels in the disyllable. So were the explosive transitions following consonant release. This seemed to indicate diphthongal production of the two vowels in the disyllables *during* production of the consonant.

[2] The data in Figure 3b were collected from a single talker (myself) who produced CVC syllables in a carrier phrase.

Compatible articulatory data have been provided by several investigators. Carney and Moll (1971) provided cinefluorographic tracings of tongue movements during production of $C_1V_1C_2V_2$ disyllables in which the second consonant is a fricative. They find movement of the tongue body from V_1 to V_2 during production of C_2. Similarly, Kent and Moll (1972) found indistinguishable trajectories and velocities of the tongue moving from /i/ to /a/ in "he monitored" and "he honored," even though in one, but not the other, utterance, the two vowels are separated by a bilabial consonant. Compatible findings are reported by several other investigators (Barry & Kuenzel, 1975; Butcher & Weiher, 1976; Perkell, 1969). This set of findings establishes the vowel as the articulatory foundation of a syllable, in the sense that it is produced throughout the syllable's articulatory extent, and suggests that in VCVs, (stressed) vocalic gestures are realized in relation to production of other (stressed) vowels, even if a consonant intervenes. In addition, this view of vowel and consonant production may explain the measured shortening effects that consonants exert on vowels.

Figure 4 illustrates the relationship between coarticulation and shortening implied by these studies. The figure's horizontal dimension represents time, and its vertical dimension, an abstract attribute: prominence. Prominence refers at once to the extent to which vocal tract activity is given over to the production of a particular segment and the extent to which the character of the acoustic signal reflects articulatory gestures associated with the segment. During the closure phase of a consonant, for example, the character of the acoustic signal is largely determined by the consonant's manner and place of closure; the signal is noisy if the segment is a fricative, silent if it is a stop, and so on. Even though a coproduced vowel can influence the signal during consonant closure, giving rise to the context sensitivity of the signal for the consonant, the voiced formant structure most characteristic of vowels is absent during consonant closure. This is indicated in the figure by giving the vowel a lesser degree of prominence than the consonant during consonantal closure.

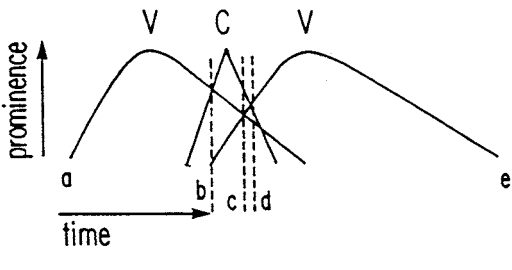

Figure 4. Schematic representation of vowel and consonant production. (The horizontal axis represents time, and the vertical axis, an abstract dimension: prominence.)

Measuring conventions locate segment boundaries approximately where ordinal changes take place in the prominence of two segments. Thus, boundaries delimit acoustic intervals during which an individual phonetic segment is the most prominent one in the signal. (Moreover, ambiguities arise concerning where a boundary should be located—e.g., between a voiceless stop and a vowel, see Lisker, 1972—when it is not obvious over a certain extent of the signal which of two segments is predominant.) In the VCV depicted in Figure 4, vowels would be given boundaries at *a* and *b* and at *d* and *e*, while the consonant would extend from *b* to *d*. If the consonant were deleted and a VV were produced, the first vowel's measured extent would be from *a* to *c*, and the second vowel's from *c* to *e*. Because of these conventions, even if the vowels in the VCV and the VV had identical articulatory extents, both would be measured to shorten in the VCV as compared with the VV. A first-approximation hypothesis, however, in view of the bidirectional coarticulation and shortening effects, would be that vowels do not change their produced durations in consonantal contexts. Rather, the consonants overlap them more or less. Although this most conservative hypothesis almost certainly will have to undergo revision, it is the simplest one to explain both coarticulation and shortening in syllables.

Now let us consider syllables produced in sequence. Öhman proposed that in VCVs, transconsonantal vowels are produced as continuous diphthongal gestures, to a first approximation, unperturbed by a medial consonant (see also Kent & Moll, 1972). Extrapolation of this view to longer speech sequences (at least to longer sequences of

stressed syllables) suggests that vowels are produced cyclically—that is, continuously, one after the other—and constitute a somewhat separate articulatory stream from gestures involved in consonant production.[3]

This hypothesis gives rise to the question of how consonants might be timed relative to the vowel stream. Some research by Tuller, Kelso, and Harris (1982) suggests part of an answer. Across utterances of the form pV_1CV_2p, produced at various rates with different stress patterns and two different medial consonants, Tuller et al. found an invariant linear relationship between duration of a vocalic cycle (i.e., the interval between the onset of muscle activity for V_1 and that for V_2) and the time lag between onsets of activity for V_1 and C. That is, timing of consonant production relative to vowel production was invariant over substantial changes in the duration of a vocalic cycle. The evidence suggests a strategy of initiating production of a consonant at an invariant phase in the production of a vowel's cycle. (Evidence of vowel shortening as consonants are added to a cluster implies, however, that the critical phase in production of a vowel at which consonant production is initiated would be different for the single consonants studied by Tuller et al. than for clusters.) As Tuller et al. pointed out, preservation of relative timing of muscle activity or gestures over changes in rate and amplitude of movement is commonly observed across a variety of activities, for example, handwriting (Viviani & Terzuolo, 1980; Wing, 1978; Hollerbach, Note 7), locomotion (Grillner, 1975), and respiration (Grillner, 1977).

Spoken and Perceived Syllabic Isochrony Reconsidered

The temporal structure of the syllable as just outlined may help to rationalize the behaviors of talkers and listeners in the experiments by Morton et al. (1976), Fowler (1979), and Fowler and Tassinary (1981) summarized earlier. By interpretation, the measured shortening of a vowel estimates how much it has been overlaid by surrounding consonants.[4] Estimates of the effective overlapping of a vowel by a consonant can be obtained by examination of Figure 2. In the figure, the metronome pulse is temporally equidistant from the measured vowel offset across the syllables. Moreover, in /ad/, with no initial consonant, the metronome pulse nearly coincides with the measured vowel onset. In other syllables, then, vowel shortening is the same as the interval from the metronome pulse to measured vowel onset. This interval estimates the interval of effective CV overlap in these syllables. By hypothesis, based on the EMG evidence provided by Tuller and Fowler (1980), talkers initiate vowels at temporally equidistant intervals under instructions to produce isochronous sequences of syllables. For their part, listeners appear to hear vowel timing; moreover, their judgments evidently are based on the articulatory timing of vowels, not on the timing of their periods of prominence in the acoustic signal as reflected by usual ways of identifying their onsets.

For listeners to hear produced rather than measured vowel timing, they must segment the speech stream in an unexpected way. They must do so in such a way that the summed duration of the segmented consonants and vowels exceeds the duration of the spoken syllable from which they have been segmented. The duration of the vowel must be its measured duration plus the extent of its effective overlap by the consonant.

Experiments 1 and 2 are designed to ask

---

[3] Further evidence in support of the view that vowel and consonant production is separate is available in the literature on speech errors. Anticipation errors, perseverations, exchanges, and substitutions never involve interaction *between* consonants and vowels. Instead, vowels intrude on other vowels, and consonants on consonants.

[4] This may be an oversimplification in two senses. First, vowels shorten for some reasons that have nothing to do with coarticulation, for example, when speech rate increases. Therefore, whereas coarticulation implies shortening, the reverse need not be true. Second, stressed vowels coarticulate with consonants *and* with unstressed vowels that precede or follow them (e.g., Fowler, 1981a, 1981b; and see Experiment 2). To coarticulate with an unstressed vowel, stressed-vowel production necessarily extends throughout (and beyond) production of a medial consonant at least in utterances where an unstressed vowel precedes the stressed vowel. But the vowel's measured shortening is less than the full extent of its overlap by other segments (again, at least in utterances including unstressed vowels). Possibly, the *effective* duration of a stressed vowel for a listener does not include the entire period of time during which it influences the acoustic signal.

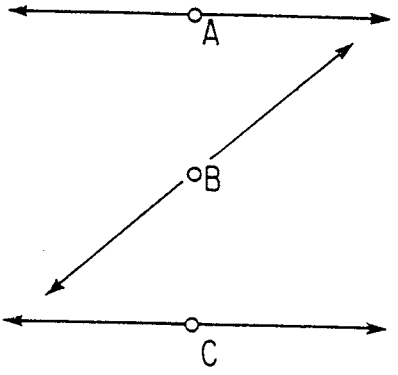

Figure 5. A display used by Johansson (1950) to study perceptual vector analysis. (Lights A and C move horizontally back and forth in phase; light B moves diagonally.)

whether such a segmentation occurs in perception. First, however, we ask, in an abstract way, *how* such a segmentation might occur.

In the literature on perception, investigators are familiar with an analogous segmentation in which separate contributions to complex events are perceptually distinguished. Figure 5 displays an example from Johansson's (1950) research (see also Johansson, 1974). The figure represents a visual display in which three moving lights are shown to subjects. The top and bottom lights, A and C, move horizontally in phase, while a third light, B, moves in a diagonal trajectory. Viewers do not report seeing two lights moving horizontally and one diagonally. Instead they report horizontal movement of an apparent rod extending from A to C, with B moving vertically along the rod.

Based on this and similar evidence, Johansson concluded that viewers perform a "perceptual vector" analysis in which movements common to a set of points serve as a perceptual frame relative to which residual motions are perceived. In the figure, all points include vectors of horizontal motion. Horizontal motion extracted from points A and C exhausts the description of their movements, but extracted from B leaves a residual, vertical motion vector.

Perceptual vector analysis is a realistic perceptual behavior. Ordinarily when components of a visual scene move together, they belong to the same event; consequently, the common movements are appropriately ascribed to coherent movement of a common frame. Imagine, for example, watching a child on a merry-go-round. If the child is seated on a horse that moves up and down relative to the surface on which it is mounted, then the child on the horse in fact moves in a complex, cycloid motion. The complex motion combines the rotation of the merry-go-round with the up and down movement of the horse relative to the floor of the merry-go-round. Observers do not see the complex movement, however. Instead, and appropriately, they see rotational movement of the merry-go-round as a whole, and an up-and-down motion of the child and the horse relative to the rotational movement. That is, they extract rotational movement, which is common to the merry-go-round and its components. This exhausts the movement of the merry-go-round's fixed structure, but, extracted from the motion of the horses, it leaves a vertical motion vector.

When we ask whether a listener can detect a vowel's produced extent despite coarticulatory overlap of part of it by a consonant, we are asking whether listeners can do the speech-perception equivalent of a perceptual vector analysis. We have seen that the vowel serves as the articulatory foundation of the syllable; for clarity in making the analogy to the visual examples, we call the vowel the *frame.* It is produced during syllable-initial and -final consonants as well as during its own interval of prominence in the signal. Therefore, acoustic reflections of the vowel's component tongue body and jaw movements provide the analogue to the vectors of common movement. These reflections exhaust the contributions to the acoustic signal during the time that the vowel is the most prominent segment in the syllable, but not during consonant production. During consonant production, two kinds of articulatory gesture contribute to the acoustic signal: the relatively slow gestures of the tongue body and jaw associated with the vocalic frame, and the relatively fast gestures of the articulators (possibly including the tongue body and jaw) associated with the consonant. If a perceptual vector analysis is possible, the gestures common to the vocalic frame may be "factored" from those specific to the consonantal portion, leaving, on the one hand, perception of

the whole vocalic frame and, on the other hand, as residual, a relatively context-free version of the consonant.

This proposed analysis, like its visual counterpart, would be a realistic one for perceivers, because it recovers the natural structure of speech events.

Experiments 1 and 2 were designed to test two predictions derived from the hypothesis that listeners perform a perceptual vector analysis on syllables and, hence, may attend to articulatory timing of vowels in the experiments outlined at the beginning of the first part of this article. One prediction is that the effective duration of a vowel for a listener is its measured duration plus its effective overlap by a syllable-initial consonant. The second prediction is that information for vowel identity is available to listeners during the production of an overlaid segment. Experiment 1 tests the first prediction, and Experiment 2, the second. Experiment 3 is designed to assess the relation between vowel perception and the perceived timing of syllables in experiments such as that by Morton et al. (1976).

Experiment 1

To ask whether listeners are sensitive to the temporal microstructure of syllables and in particular to the relationship of overlap between syllable-initial consonants and post-consonantal vowels, we used a technique developed by Raphael (1972). Raphael has shown that a syllable-final stop or fricative can be synthesized that is identified as voiced after a long-duration vowel and as voiceless after a short-duration vowel. This is compatible with the fact that, particularly in English, voiced syllable-final consonants are preceded by longer vowels than are voiceless consonants. By generating a set of stimuli with a range of vowel durations before the final consonant and by asking subjects to label the final consonant as voiced or voiceless, Raphael was able to identify a voicing boundary within the continuum of vowel durations. The boundary is defined as the vowel duration at which subjects label the syllable equally often as /d/ or /t/, that is, the 50% crossover point. In later studies, Raphael, Dorman, and Liberman (Note 8) and Raphael and Dorman (1980) showed that the crossover point is shifted toward the /t/ (short vowel) end of the continuum by a syllable-initial consonant. That is, the final consonant is heard more frequently as /d/ when a consonant precedes the vowel than when the vowel is syllable-initial. This may indicate that the vowel is heard as being longer when preceded by a consonant than when it is syllable-initial. For syllable-initial /d/, all or most of the transitions, which in these stimuli were necessary in order to specify the initial /d/, were also heard to belong to the vowel. This interpretation is consistent with the facts of production; the direction and extent of second formant (F2) transitions appropriate for /d/ are conditioned by the following vowel because the two segments are coarticulated during the release of the consonant.

In the study by Raphael et al. (Note 8), an initial /r/ also shifted the /t/–/d/ boundary substantially, whereas steady-state frication characteristic of /s/ shifted it only slightly. This latter outcome was replicated by Raphael and Dorman (1980) with natural speech. These experiments made it clear that the perceived voicing of a final stop can be affected by vowel length. In the following experiment, I attempt to extend these findings to some of the syllables depicted in Figure 2. If adding initial consonants to a vowel increases the vowel's effective duration, then, following Raphael et al. (Note 8), we should observe a change in the voicing boundary of syllables beginning with /a/, /b/, /m/, and /s/. Furthermore, we predict a greater effective lengthening of the vowel by consonants that according to Figure 2 shorten the vowel substantially (e.g., /s/) than by those that shorten it very little (e.g., /b/).[5]

[5] This prediction may appear contradictory to the findings of Raphael et al. who found limited effects of /s/ on apparent vowel duration and substantial effects of /d/. The difference in prediction and outcome is derived from a difference in measurement criteria for the vowel. In experiments by Raphal et al., voiced formant transitions following release of /d/ were identified as belonging to the consonant and not to the vowel; hence, when the addition of transitions affected the voicing judgments, the influence was identified as one of the consonant on the effective duration of the vowel. In our measurements, however, voiced formant transitions are included in the measurement of vowel duration. Therefore, the predicted *additional* effect of a voiced stop such as /d/ or /b/ on voicing judgments is small.

Method

Subjects. Subjects were 63 introductory psychology students at Dartmouth College.

Stimuli and materials. We selected the syllables /ad/, /bad/, /mad/, and /sad/ spoken by two of the talkers who provided the data for the experiment reported by Fowler and Tassinary (1981; and who were two of the three talkers who provided that data shown in Figure 2).[6] These syllables had shown a range of vowel shortening that spanned 20 msec collapsed over the two talkers. The order of measured vowel durations decreased in the series: /ad/, /bad/, /mad/, and /sad/.

For each talker, a single token of each of the four syllables was selected from the nonmetronome condition of the experiment reported by Fowler and Tassinary. These syllables were digitized and edited using the pulse-code modulation system at Haskins Laboratories.

The final portion of the syllable /ad/ was spliced from the rest (50 msec for talker 1 and 85 for talker 2). The portion excluded any voicing during the closure for the /d/ and any release of the /d/ to facilitate a shift in identification from /d/ to /t/. This final section of the syllable /ad/ replaced the final portion of the other three syllables to ensure that the final consonant of the four syllables was equivalently /d/- or /t/-like. Finally, the vowels in each syllable were made equal in duration (within a pitch pulse) by deleting pitch pulses from the steady-state portions of syllables with longer vowel durations. The original vowel durations of the four syllables averaged 225 msec for talker 1 and 236 for talker 2. From each of these syllables, a 10-step continuum was constructed by successively deleting one pitch pulse taken for talker 1 and two for talker 2 (a female), insofar as possible from the relatively steady-state portion of the vowel. This gave continua with a range of approximately 75 msec for talker 1 and 90 msec for talker 2.

For each talker, four test orders were constructed, one for each continuum (syllable). Each test order began with 20 trials in which the two end points of the continuum were repeated 10 times each in alternation. These served to familiarize the listeners with the most /d/- and /t/-like sounds they would hear. The introductory series of 20 trials was followed by 100 trials in which the 10 stimuli were presented 10 times each in random order. This pattern, 20 trials in which the end point stimuli were repeated in alternation and 100 randomized trials, was repeated twice more for a total of 60 introductory trials and 300 test trials. The first third of the test served as practice; the data to be reported are from the last set of 200 test trials. There were 2 sec between trials with a longer delay of 4 sec following every 10th trial.

Design. Subjects were nested within the four levels of the independent variable, syllable (/ad/, /bad/, /mad/, and /sad/), and the two levels of the variable talker. With a single exception, eight subjects were assigned to each cell in the design. Only seven subjects were run for the syllable /bad/ produced by the first talker. We expected a shift in the /d/–/t/ boundary toward the short-vowel (/t/) end of the continuum progressively in the sequence /ad/, /bad/, /mad/, and /sad/.

Procedure. Subjects listened to the test orders over earphones in groups of one to four in a sound-treated room. They were instructed to listen to the initial 20 sounds of alternating /d/- and /t/-final syllables on each third of the test, writing *d* or *t* as appropriate on their answer sheet as they followed along. On the next 100 trials in each third of the test, they were instructed to write *d* or *t* depending on which final consonant they heard, choosing only between the responses *d* and *t*.

Results and Discussion

The prediction that the voicing boundary would shift toward /t/ progressively in the series /ad/, /bad/, /mad/, and /sad/ was assessed by comparing the four syllables on the measure of number of *d* responses to each stimulus in the continuum. Figure 6 displays the results of this procedure collapsed over talkers 1 and 2. The ogival curves for the four syllables cross over the 50% point in just the predicted order. Interpolating from the figure, the boundaries for /ad/, /bad/, /mad/, and /sad/ are 5.36, 5.70, 5.90, and 6.39.

In an analysis, the average number of *d* responses given to the four syllables was compared for stimuli near the voicing boundaries, that is, stimuli 5, 6, and 7. Collapsed over talker and stimulus number (5–7), because neither variable interacted with syllable, the average number of *d* responses out of 20 to the four syllables was 7.4, 8.7, 9.1, and 11.4. This increase reflects the increasing resistance to labeling the final consonant as *t* throughout the series. The increase was significant according to a trend test in which the mean for each syllable was weighted according to its measured vowel shortening in the syllables displayed in Figure 2. In the analysis, both subject and talker were treated as random factors, $F(1, 3) = 18.86$, $p = .02$.

In this analysis, listeners' judgments of syllables produced by talker 1 showed just the predicted increase, whereas their judgments of talker 2 showed a reversal of /bad/ and /mad/. This reversal in fact occurred on just one of the three crossover stimuli.

The outcome of this analysis, though certainly not striking, is compatible with the hypothesis that the duration of the vowel as perceived by listeners increases with increases in the vowel's measured overlap by the consonant (its measured shortening). Nonethe-

[6] We attempted to create continua using syllables of the third talker in the metronome study. However, we were not successful in creating continua of syllables that listeners could label consistently.

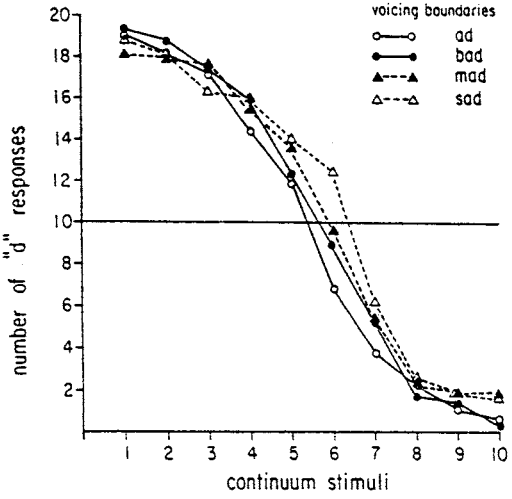

Figure 6. Number of /d/ responses to four different vowel-duration continua.

less, whereas the range of shortening was about 20 msec in the experiment by Fowler and Tassinary (1981), the difference in perceived vowel duration as assessed by the present experiment was only about 10 msec.

Experiment 2

Experiment 1 has an alternative interpretation to the one that we have proposed. Possibly, listeners are familiar with different durations of vowels following /b/, /m/, and /s/; consequently, they expect relatively shorter vowels following /s/ than /m/ and following /m/ than /b/. If so, the results of Experiment 1 document those expectations but do not reveal a tendency to hear a vowel during that part of the acoustic signal in which vowels and consonants coarticulate but consonants predominate in the signal.

Experiment 2 was designed to provide evidence converging with Experiment 1 that perceivers extract vowel information during production of segments that coarticulate with it. If they do, then time to identify a vowel, timed from the vowel's *measured* acoustic onset, should be shorter, the more extensive its effective overlap with preceding segments. After estimating overlap by vowel shortening, then, time to identify /a/ should be shorter in /sa/than in /ma/ and shorter in /ma/ than in /ba/. Experiment 2 was designed to test that prediction.

Method

Subjects. Subjects were 14 undergraduates at Dartmouth College.

Stimuli. Stimuli were naturally produced VCV disyllables in which the first vowel was unstressed schwa, the consonant was /b/, /m/, /s/, or /p/, and the second vowel was /a/ or /i/. A disyllable with /p/ replaced the syllable /ad/ in Experiment 1. As Figure 2 shows, vowel shortening after /p/ is greater than that following /s/. Therefore, predicted time to identify a vowel is expected to decrease in the series əbV, əmV, əsV, əpV.[7]

Three tokens of each disyllable were produced, giving 24 different stimuli in all. The stimuli were randomized into five 48-trial blocks with the constraint that in each block each token occurred twice. Stimuli were recorded on audiotape with 2 sec between trials and 10 sec between blocks.

Table 1 provides durational measures of the stimuli. Measures of schwa duration were taken from the onset of periodicity in the signal to closure for the consonant. For the consonants, the onset of the closure interval to the onset of voicing for the vowel was measured. Stressed vowels were measured from the earliest evidence of voicing following release of the consonant to signal offset. As others have found (see also Figure 3), the durations of consonants and stressed vowels were negatively correlated ($r = -.76$).

Table 2 provides measures of F2 during the initial schwa of each disyllable. (Measures were obtained by the method of linear prediction.) Measures were taken during the four 20-msec time frames preceding closure for the consonant. The table shows that F2 for schwa is lower when the forthcoming stressed vowel is /a/ than when it is /i/. This is compatible with the substantially higher F2 for the high vowel /i/ than for the low vowel /a/ and indicates that anticipatory coarticulation of the stressed

[7] This prediction requires clarification. The observation that vowel shortening in /pV/ is greater than in other syllables is true if vowel onset is defined as the onset of voicing following release of a syllable-final consonant. If the onset were located instead at the onset of the formant transitions following release of the /p/—an equally defensible location because the transitions provide vowel information as well as being sufficient to specify the /p/ to a listener—the rank ordering would change. However, it is not necessary for the aims of the present experiments to defend either of these measuring points as superior. Indeed, according to the present arguments, any measuring point is indefensible that purports to divide an acoustic signal into nonoverlapping phonetic segments. The aims of the experiments can be met if a reference point is selected and used consistently in assessing syllable-timed productions (Figure 2), judgments of vowel duration (Experiment 1), vowel and consonant classifications (Experiments 2 and 3), and syllable-timing judgments (Experiment 3). If syllables are aligned similarly around the selected reference point for syllable-timed productions and judgments as for assessments of vowel durations and for vowel classifications, but not for consonant classifications, then the conclusion is warranted that syllable timing is related to vowel sequencing more than to consonant sequencing.

Table 1
Durational Measures (in msec) of the Disyllables Used in Experiments 2 and 3, Averaged Over the Three Tokens of Each Type

Disyllable	/ə/	C	V
əba	61	128	434
əbi	61	123	465
əma	43	144	402
əmi	56	123	390
əsa	45	189	387
əsi	51	195	337
əpa	42	206	370
əpi	40	218	371

vowel precedes closure for the consonant (see also Fowler, 1981a, 1981b).

Figure 7 displays this more clearly by plotting the difference between F2 for /ə/ preceding /i/ and /a/ separately for each disyllable pair during the last four 20-msec intervals preceding consonant closure. This evidence of coarticulation is compatible with Öhman's findings and other evidence cited earlier.

Until the final frame, disyllables including /b/ and /m/ appear to be more differentiated than those containing /s/ and /p/. If listeners use average frequency of the second formant of schwa over these time frames as a source of information about the forthcoming vowel, they will not show the rank ordering of response times we have predicted. However, the predicted ordering is reflected in the rate of change in the plotted difference score over the last three frames where the change is monotonic; /b/ shows the lowest rate of change, and /p/ the highest. If this measure reflects information about ongoing adjustments in vocal tract shape for the forthcoming vowel to which listeners are sensitive, then Figure 7 may offer acoustic support for the predicted ordering of response times.

Design. The major independent variable was consonant identity; a second was vowel identity. All subjects participated at all levels of the independent variables.

Table 2
Measures of F2 of Schwa During the Four 20-msec Frames Preceding Consonant Closure Averaged Over the Three Tokens of Each Type

	Frame number before closure			
Disyllable	4	3	2	1
əba	1,464	1,406	1,334	1,304
əbi	1,676	1,641	1,628	1,619
əma	1,469	1,470	1,403	1,314
əmi	1,755	1,687	1,640	1,670
əsa	1,689	1,698	1,693	1,702
əsi	1,794	1,791	1,853	1,921
əpa	1,451	1,415	1,373	1,328
əpi	1,517	1,426	1,517	1,683

Note. F2 = Second formant.

The dependent variable was time to classify the vowel, timed from the vowel's measured onset. Based on the findings of Fowler and Tassinary (1981) displayed in Figure 2, I expected reaction time to classify a vowel as /i/ or /a/, measured from the acoustic onset of the vowel's period of prominence, to decrease in the series əbV, əmV, əsV, əpV because the measured vowel durations decrease in the series. I had confidence that this rank ordering of vowel durations is stable because the same rank ordering was reported by House and Fairbanks (1953) for vowels in symmetrical bVb, mVm, sVs, and pVp contexts. Having previously examined only stimuli in which the stressed vowel was /a/, I had no reason to expect a difference in reaction time to /i/ or /a/ nor any interaction between the variables, consonant and vowel identity.

Procedure. Subjects were tested individually. They listened to the test sequence over earphones, classifying the stressed vowel on each trial as /i/ or /a/ by making a button-press response. For half the subjects, /i/ corresponded to the left-hand button and for the other half, /a/ corresponded to the left-hand button. Responses and reaction times were collected by microcomputer. Times were measured from the acoustically defined vowel onset by placing a click on the second channel of the audiotape, 100 msec prior to measured vowel onset on the first channel. In the experiment, these clicks caused a millisecond clock to be read; the clock was read again on receipt of the subject's button-press response, and the difference in the times minus 100 msec was the subject's reaction time.

Subjects were instructed to make their responses as quickly as possible but to minimize errors.

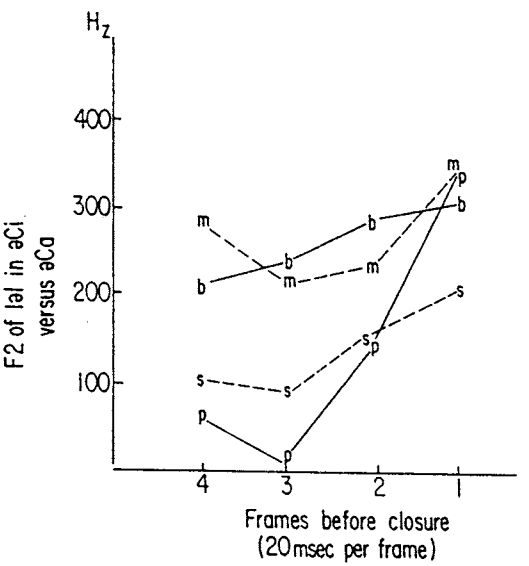

Figure 7. Anticipatory coarticulation of stressed /i/ and /a/ in the disyllables of Experiments 2 and 3. (Second formant [F2] of initial schwa in əCa subtracted from F2 of schwa in əCi is plotted for each of the four disyllable pairs and for four 20-msec frames preceding closure of the consonant.)

Results

Results are reported for the final four blocks of the experiment, the first block serving as practice. Subjects were quite accurate, averaging 95% correct overall.

Average reaction times to the disyllables əbV, əmV, əsV, and əpV were 483, 468, 463, and 424 msec, respectively. The effect of consonant identity is significant, $F(3, 39) = 33.7$, $p < .001$. More important, however, the decrease in reaction time in the series occurred as predicted. Based on the measured shortening in Figure 2 (averaged over three talkers—those whose productions provided stimuli for Experiment 1 and one other), the predicted differences in reaction time in the series is 14 msec for əbV versus əmV, 8 msec for əmV versus əsV, and 8 msec for əsV versus əpV. The first two predicted differences fit the observed differences fairly well; however, the obtained difference between əsV and əpV is 39 msec rather than the predicted 8 msec. A planned comparison weighting reaction times according to the predicted differences is highly significant, $F(1, 39) = 81.10, p < .0001$.

The main effect of vowel identity is nonsignificant in the analysis, $F(1, 13) = 1.65$, $p = .22$, but the interaction between consonant and vowel identity is significant, $F(3, 39) = 9.55$, $p < .001$. One reason for the interaction is that the ordinal relation of əmV and əsV is as predicted when the vowel is /i/ (465 vs. 441 msec) but is reversed when the vowel is /a/ (472 vs. 484 msec). In addition, when the vowel is /i/, reaction times to əsV and əpV are the same (441 msec) but differ when the vowel is /a/ (406 vs. 484). I had no reason to predict a difference in rank ordering of reaction times based on vowel identity, because in earlier studies the vowel was invariably /a/. Whereas articulatory support for this interaction or other reasons for it will have to be investigated, the reasons for the interaction will not be pursued here. However, a similar interaction will be sought in listeners' assessments of the timing of the syllable sequences in the next experiment.

Discussion

This experiment provides evidence that vowels are detected during intervals when the vowels coarticulate with prevocalic segments (including initial consonant and the preceding schwa). Experiment 2 shows that the time to identify a vowel, timed relative to measured vowel onset, is correlated with the vowel's measured shortening. Based on the coarticulation evidence cited earlier (and represented schematically in Figure 4), we interpret the relative shortening as an index of relative overlap by the prevocalic consonant (and, perhaps, by the unstressed schwa; see also Fowler, 1981a, 1981b). Therefore we interpret the decrease in vowel classification time with shortening as evidence that listeners use information for the vowel in the prevocalic segments *as* information for vowel identity.

These results converge with those of Experiment 1. That experiment found that the measured duration of a vowel at which judgments of voicing of a syllable-final consonant shift from voiced to voiceless decreases progressively in the series /ad/, /bad/, /mad/, and /sad/. One interpretation of this outcome is that listeners are sensitive to the shortening effects of consonants and vowels displayed in Figure 3a, but another interpretation is promoted by the results of Experiment 2—that the effective duration of a vowel for a listener is the vowel's measured duration plus the overlap of part of its perceived extent by a syllable-initial consonant.

Previous experiments in this series (Fowler, 1979; Fowler & Tassinary, 1981) have used the vowel /a/ exclusively. Experiment 2 introduced the vowel /i/ and obtained an interaction between initial consonant and vowel in vowel-classification times. In Experiment 3, assessments are made of the relative rhythmic alignment of the syllables used in Experiment 2. If perception of vocalic timing underlies the perception of speech rhythms as I propose, then the interaction found in Experiment 2 should be reflected also in listeners' rhythmic alignments of these disyllables. Experiment 3 tests this prediction.

Experiment 3

In this experiment, I relate listeners' vowel-classification times, obtained in Experiment 2, to listeners' perceptions of rhythmicity, which I propose have their bases in percep-

tion of cyclic vowel production. In addition I also assess the relation of listeners' consonant classifications to their perception of rhythm. According to the view of perception being developed here, consonant classifications are not related to the perceived timing of syllables.

Method

Subjects. Subjects were 30 Dartmouth undergraduates. Fifteen participated in the tapping task, and 15 in the consonant-classification task.

Stimulus materials. The experiment used the audiotape devised for the vowel-classification task of Experiment 2.

Procedure. In Experiment 2, subjects were asked to classify the stressed vowel on each trial as /i/ or /a/. In the present experiment, one group of subjects was asked to tap a key in time with the successive disyllables, tapping once for each disyllable at a point corresponding to the syllable's "beat." This technique, like the metronome technique used by Rapp (Note 6) and by Fowler and Tassinary (1981), enables discovery of the perceived temporal alignment of different syllables (see Figure 2).

A second group of subjects was asked to classify the consonants on each trial as /b/, /m/, /p/, or /s/, making a button-press response as quickly as possible. Assignment of phoneme labels to buttons was varied over subjects.

Design. As in Experiment 2, independent variables are consonant identity (/b/, /m/, /p/, /s/) and stressed-vowel identity (/i/, /a/). The dependent measure is response time, measured initially as relative to measured vowel onset and next as relative to measured stressed-syllable onset. I expected vowel-classification times obtained in Experiment 2 to correlate with tap times in the present experiment. This would suggest a close relation between information necessary to identify a vowel and perceived relative timing of the disyllables. No such relation was predicted between consonant-classification times and tapping times.

Results

When tapping times are measured relative to vowel onset, the effect of consonant is highly significant, $F(3, 42) = 297.78$, $p < .0001$. Tap times follow vowel onset by 207, 187, 137, and 125 msec for the disyllables əbV, əmV, əsV, and əpV, respectively. This is exactly the rank ordering of disyllables obtained in Experiment 2, although responses to əsV are closer in reaction time to əpV in the present experiment and to əmV in Experiment 2.

As in Experiment 2, the effect of vowel identity is nonsignificant, $F(1, 14) = 2.16$, $p = .16$, but the interaction is significant, $F(3, 42) = 20.63$, $p < .001$. In Experiment 2, there were two reasons for the interaction. First, the rank ordering of times to əmV and əsV was as predicted (based on measured shortening in Figure 2) when the vowel was /i/ but reversed when the vowel was /a/. Next, there was no difference in reaction time to əsi and əpi, but a large difference between əsa and əpa. In the present experiment, the predicted rank ordering of əmV and əsV was obtained for both vowels. However, as in Experiment 2, there was essentially no difference in tapping times to əsi and əpi (123 vs. 121 msec), but the predicted direction of difference appeared between əsa and əpa.

Table 3 provides mean response times in the tapping and consonant-classification tasks, respectively, with response times now measured relative to onset of acoustic energy for the consonant (i.e., release for /b/ and /p/). Table 3 provides comparable times for the vowel classifications of Experiment 2. As predicted, vowel and tap times pattern similarly. The correlation between them, computed over the eight disyllables, is .95. Consonant times also pattern similarly to tap times ($r = .79$). Moreover, the patterns of vowel and consonant times are correlated ($r = .73$). All of these correlations are significant. However, the significant relationship between tap times and consonant response times is due to shared variance between vowel and consonant times. When that variance is partialed out, the correlation between tap times and consonant times falls to .46, a nonsignificant value. In contrast, when variance shared by consonant- and vowel-identification times is partialed from the tap–vowel correlation, the partial correlation remains significant ($r = .90$). In a multiple regression analysis, only the vowel times contribute significantly to predictions of tap response times. This suggests that perceived timing of stressed syllables is a function only (or primarily) of perceived information pertaining to vowel identity as predicted and is not significantly a function of perceived consonant identity.

Discussion of Experiments 1–3

I have attempted to establish a relationship, on the one hand, between the temporal and articulatory structures of spoken sylla-

Table 3
Measures of Response Time (in msec) in Experiments 2 and 3

Disyllable	Tap			Consonant			Vowel		
	RT(AE)	RT(C)	*SD*	RT(AE)	RT(C)	*SD*	RT(AE)	RT(C)	*SD*
əba	328	205	45	728	605	83	618	495	74
əbi	338	218	46	757	637	82	600	480	79
əma	—	328	55	—	670	83	—	616	72
əmi	—	313	54	—	668	83	—	588	77
əsa	—	339	59	—	683	59	—	673	73
əsi	—	320	54	—	673	73	—	638	94
əpa	335	233	58	762	660	127	612	510	83
əpi	339	246	49	703	610	99	660	567	76

Note. RT(AE) = response timed from onset of acoustic energy for the consonant; RT(C) = response timed from onset of closure for /b/ and /p/.

bles and, on the other hand, between both of these systematic properties of produced speech and the perceived timing of syllables in productions that talkers intend to be rhythmical. I have proposed that measured vowel shortening in the context of surrounding consonants is an index of coarticulatory overlap of the vowel by consonants. This proposal is supported by the coarticulation literature, which shows that vowels are coproduced with consonants (Barry & Kuenzel, 1975; Butcher & Weiher, 1976; Carney & Moll, 1971; Öhman, 1966) and provides evidence for vowel-to-vowel gestures of the tongue body occurring concurrently with medial consonant production. Based on an elaboration of Öham's proposal suggesting that vowels are produced continuously in sequences of stressed vowels, I hypothesized that the perceived timing of syllables is based on the perceived timing of vowels.

The research presented here supports this view, showing that both the perceived duration of a vowel (Experiment 1) and the time necessary to identify a vowel (Experiment 2) are affected by the identity of the syllable-initial consonant. In particular, Experiment 1 showed that the more extensive the shortening effect of a consonant on a vowel (and hence, by hypothesis, the more the consonant overlaps the vowel), the more the consonant helps resist shifts in perceived voicing of the syllable-final consonant, which occur as the vowel's measured duration decreases. Experiment 2 found that the more extensive the shortening effect of a consonant on a vowel, the shorter the subjects' response time to classify the vowel as /i/ or /a/, timed from the vowel's measured onset.

Experiment 3 established a relation between perception of the stressed vowel in a sequence of disyllables and the perceived timing of the sequence. Vowel-classification times and tap times were highly correlated.

Some problems with the present view of vowel production as continuous have been raised in a recent article by Shaffer (1982). Shaffer pointed out that with changes in rate of production, vowels change in duration more than consonants. But if vowels and consonants were produced coordinately but separately as proposed here, either of two different outcomes would be expected. Just one segment type might be affected by rate change without any effect on the other; alternatively, being coordinate, consonants and vowels might change proportionately. Neither outcome corresponds to what is observed.

There is a way in which separate, but coordinate, segment types could change disproportionately, however. There is nonlinearity in the articulatory system in the form of an upper limit on segment shortening due to rate changes. If at slow rates of talking, consonants are closer to this limit than are vowels, then they would shorten less with an increase with rate than do vowels. Consonant gestures *are* faster than vocalic gestures at slow or conversational rates of talking. In a recent study, Tuller, Harris, and Kelso (in press) reported a shorter duration of muscle activity supporting consonant than of vowel production

at a slow rate of talking. At a fast rate, duration of activity for the consonant and vowel is more similar, that for the consonant having decreased by 13% and that for the vowel by 23%.

Shaffer (1982) also argued that the present proposal

> fails to account for the coarticulation of consonants and for coarticulation across syllable boundaries: it does not consider the timing of postvocalic consonants or show why syllable duration is affected by the size of the consonant clusters. (p. 121)

The present view *does* fail to account for the coarticulation of consonants, but only because it does not yet address consonant production except in relation to vowel production. Consonants are considered primarily as they may affect perceived rhythm or, more often, as they mask evidence of vowel production used by listeners to guide rhythm judgments. However, I do not detect anything in principle that will prevent incorporation of information about relative timing of consonants into a theory of vowel production as separate from consonant production. The timing of postvocalic consonants *relative* to the vowel and the coarticulation of consonants with vowels across syllable boundaries are addressed.

As for increases in syllable duration with increases in consonant cluster size, the theory can offer two possible hypotheses. Segments have compression limits (e.g., Klatt, 1976). In particular, the constraint that consonants be initiated at a particular phase in the production of vowels (Tuller et al., 1982) may prevent excessive overlap of the vowel by consonants in a cluster. If so, then production of a large cluster may force a discontinuity in vowel production with the consequence that initial consonants in a prevocalic cluster may not coarticulate with the following vowel but may with a preceding vowel; similarly, final consonants in a postvocalic cluster may not coarticulate with the preceding vowel but may with the subsequent one. However, in view of the findings that stressed vowels coarticulate over long extents when unstressed vowels follow (Bell-Berti & Harris, 1979; Fowler, 1981a, 1981b), a different outcome is also possible. Consonant clusters may force an *increase* in the duration of a vowel cycle to preserve continuity of the vowel stream. Further research will have to distinguish these possibilities and to distinguish them from others that might be proposed.

Contributions From Phonetics and Phonology

In this part of the article, I develop the three ideas outlined in the introduction. First is the general idea that investigation of language structure, which proceeds largely independently from studies of language use, can provide a useful source of evidence converging (or failing to converge) with results of experimental studies. The second, more specific, idea is that some phonological rules are "natural" in the specific sense that they reflect exaggerations and conventionalizations of articulatory dispositions. Insofar as they can be identified as such, they offer a source of evidence concerning the nature and identity of some dispositions. Third, I provide examples that I suggest are exaggerations and conventionalizations of the articulatory tendency to produce vowels in a continuous, cyclic fashion.

Phonological descriptions of languages characterize systematic properties in the phonological forms of lexical items. That is, the descriptions factor systematic (general) phonological properties common to lexical items, expressed as general rules, from properties idiosyncratic to individual items. This factoring reveals a number of characteristics of the lexicons of languages that are relevant to psychological interests. Spoken-language systems exist only as they are used by speaker/hearers; moreover, they are evolutionary acquisitions of speaker/hearers. In view of these facts, systematic phonological properties provide clues to the nature of the speaker/hearers themselves (see also Chomsky, 1980, who, however, focuses on their revealed cognitive nature rather than on their perceptual and articulatory natures as I will emphasize here).

Some of these clues appear to be more fundamental or significant than others. They are systematic properties that are popular across languages. For example, many languages devoice final obstruents. In German, the noun *Bund* is pronounced /bunt/ in the nominative, but /bund/ in the genitive *Bundes*. In Polish, *snow* is /s'n'ek/ in the nominative but

/s'n'ega/ in the genitive. In Russian, the nominative of *leg* is /noga/, but the genitive plural is /nok/. (The German example is from Comrie, 1980, and the Polish and Russian examples, from Kenstowicz & Kisseberth, 1979.) That this phonological rule is somehow natural to language users is suggested by the fact that children learning language also have a tendency to devoice final consonants. This occurs even in English where it is inappropriate (Oller, Wieman, Doyle, & Ross, 1976).

Systematic phonological properties that are popular across languages may be popular for a reason. Indeed, there may be many reasons why a particular kind of systematic property is favored by languages, but of interest here is the possibility that many properties are natural in resembling articulatory dispositions. Word-final devoicing may be an example.

If some phonological regularities do resemble articulatory dispositions, then phonological investigation can serve a useful function for psychological investigation of speech production. Articulation is difficult to study with respect to issues of psychological (as opposed, say, to physiological) interest, not simply because the articulators are difficult to access, but also because direct study of articulation tends to provide more detail than current psychological perspectives on speech motor-control can organize and explain. Identification of popular systematic properties of the phonologies of languages can contribute to direct study of articulation in two ways. First, it can suggest the kinds of articulatory regularities that have served as resources for the evolution of phonologies. These suggestions can help to focus the search for regularities or organizing principles in articulation. Next it can serve as converging evidence for hypothetical organizing principles, such as that of cyclic vowel production, that may have emerged, perhaps, dimly, from articulatory or perceptual investigations of speech. That is the use to which phonological evidence will be put here.

Systematic and Idiosyncratic Properties of Language

Not all systematic properties of lexical items are factored out in phonological rule systems. Two kinds of systematic properties of lexical items can be identified that I will call *conventional* and *necessary*. Conventional systematic properties are expressed by general rule, whereas necessary ones are not. Conventional systematic properties are specific to individual languages; they are conventions, which are used to convey linguistic information. An example is the formation of the plural in English. The plural is formed by adding (morphological) *s* to a word. The pronunciation of the *s* is conditioned in a ruleful way by properties of the phonological segment adjacent to which the *s* is appended. If the segment is unvoiced and is neither a fricative nor an affricate, the plural is realized as /s/. If the segment is voiced and neither a fricative nor an affricate, the plural is /z/. Otherwise, the realization is /ɨz/. This conditioning is systematic—it can be expressed as a rule—but it is a convention. An alveolar fricative after a voiced segment need not be voiced (witness *dance*, phonemically /dæns/). And other languages have other plural formation rules.

Other systematic properties of language are "necessary"; that is, they are essentially universal and (to a first and close approximation) could not be other than they are. An example is the f_0 contour on a vowel following a voiced or voiceless stop. Following release of a voiced stop, the fundamental frequency of the voice is low and gradually rises over a period of more than 100 msec (e.g., Hombert, Ohala, & Ewan, 1979; Ohala, 1978). After a voiceless stop, f_0 is high and gradually falls. The reasons for this patterning are not fully understood, but it is generally agreed that the f_0 contour is a necessary consequence of the aerodynamic and articulatory adjustments made to maintain or resist voicing during stop closure (Ohala, 1978). The f_0 contour following a stop is a systematic property of a word but is not a convention and is not expressed as a phonological rule in the phonologies of languages.

In the subsequent sections, I focus on both necessary and conventional phonological properties. Necessary systematic properties are direct sources of evidence about articulatory constraints on production. For this reason, they are very useful to study. However, I focus primarily on a second aspect of

necessary properties—they may serve as a source of new linguistic conventions as languages change. Thus it is important to look at the evolution of conventions to gain insight into necessary systematic properties.

Leakages From Articulation Into the Phonologies of Languages

Ohala (1974, 1981) has argued that exaggerated versions of necessary systematic properties of languages occasionally enter the language as conventions due, in his view, to systematic misperceptions by listeners. For example, Ohala suggested (1974, 1981) that tone languages such as Punjabi may have evolved from atonal languages with voicing distinctions among stop consonants.

This evidence is derived from comparisons of related languages, one of which is a tone language and the others of which are not. Punjabi, for example, is a tone language related to Hindi and other languages that are not. In Punjabi, the distinction between aspirated voiced consonants and unaspirated unvoiced consonants, present in Hindi, is absent. Words starting with an aspirated voiced consonant in Hindi have a low tone on the vowel in Punjabi. In the history of Punjabi, apparently, the distinction between voiced aspirated and unvoiced unaspirated consonants was lost, leaving behind a tonal distinction between words formerly differing in voicing of the initial consonant.

Ohala ascribed this sound change to consistent misperceptions by listeners. Hearing the f_0 contours produced by voiced and voiceless consonants on following vowels, language learners may have interpreted the contours mistakenly as systematic conventions. Consequently, when these listeners produced voiced or voiceless stop-initial syllables, they intentionally produced a tone on the following vowel. Being exaggerated, the contours were more salient than the unintentionally produced contours that necessarily accompany stop voicing or voicelessness. As numbers of language learners made the error (uncorrected for unexplained reasons[8]), syllables differing in voicing of the initial consonant were marked in two ways—one by the voicing distinction itself and the other by the tonal pattern on the vowel. In some languages, the tonal contours replaced the voicing difference as the critical difference between certain syllables. These languages became tone languages. Ohala (1981) offered many other examples in which conventions apparently entered languages as exaggerations of necessary systematic properties of speech (see also Wright's, Note 10, analysis along similar lines of the [continuing] vowel shift in English).

If the examples are real, they imply that some systematic conventions that are popular among unrelated languages may reflect exaggerations of necessary regularities in speech production and, hence, in fact may provide clues to the identity of some of these regularities. Review of the phonological literature reveals several systematic properties suggestive of the mode of vowel production proposed here to underlie (in part) the impression of rhythmicity of speech. As we have characterized (stressed) vowel production, it has two central aspects. Vowels' leading and trailing edges are overlaid by consonants, and vowels are produced as a cyclic stream somewhat separate from the production of consonants. Reflections of both of these aspects can be found in the phonologies of languages. I know of no conventions that contradict the proposed mode of vowel production.

Language Conventions Suggestive of Continuous Vowel Production

Vowel Shortening and Lengthening

A number of languages have adopted conventions whereby consonant and vowel length serve a distinctive function in the language (i.e., a long vowel, V:, or long consonant, C:,

[8] Louis Goldstein (Note 9) has suggested a reason for this. Locke's (e.g., 1979) research on the so-called "fis" phenomenon in children reveals that immediately after producing a word, children are more aware of what they meant to say than of what they in fact uttered. Locke's research focuses on children whose speech does not seem to distinguish pairs of sounds (e.g., /w/–/l/ or /r/–/w/) that are distinct in adult language. After having produced something like /weyk/ meaning "rake," they will deny having said "wake." But if their production is recorded and replayed to them 1 day later, they are no better than other listeners in distinguishing their *wake*s from their *rake*s.

is considered a different vowel or consonant from its short counterpart). In some of these languages, rules ensure that consonant and vowel length are complementary. These rules may constitute exaggerations and conventionalizations of the shortening effects of consonants on vowels depicted in Figure 3.

For example, Swedish distinguishes long and short versions of vowels and consonants phonologically. In Swedish, constraints on syllable structure prevent long postvocalic consonants and long vowels from co-occurring in a syllable, and they prevent short vowels and (only) short postvocalic consonants from co-occurring in stressed syllables (Elert, 1964, cited in Lindblom & Rapp, Note 11). Allowed stressed syllable structures are (C)V:(C) and (C)VC:(C). (Parentheses indicate that segments are optional.) This reciprocal relationship between vowel and consonant length at the phonological level of description of the language is *not* the same as the (phonetic) shortening depicted in Figure 3. Lindblom et al. (1981) showed that Swedish *long* vowels are shortened by intra- or transsyllabic consonants, just as English vowels are. But the phonetic shortening of the long vowels does not transform them into phonologically short vowels. (Thus, although V: in V:C is shorter than V: in isolation, both are phonologically long vowels.) In Swedish, then, a reciprocal relation exists between consonants and vowels at two levels—at a phonetic level where it also occurs generally across languages, and at a phonological level where it is a convention special to Swedish.

Yawelmani, a native American language once spoken in California, like Swedish, distinguishes phonologically long and short vowels. Also like Swedish, Yawelmani maintains a reciprocal relation between vowel length and, in this case, the *number* of following consonants. In Yawelmani, a phonologically long vowel in a stem is made short if a suffix is added to the stem causing the stem vowel to be followed by more than one consonant.

According to Kenstowicz and Kisseberth (1979):

> Examination of a variety of other languages reveals that alternations in [phonological] vowel length typically revolve around differences in the consonant–vowel structure of words, with long vowels preferred in "open syllables" (__CV) and short vowels preferred in "closed syllables" (__CC). (p. 83)

This is just what we would expect if languages tend to conventionalize by exaggeration, properties of production that already are necessarily systematic in language. By virtue of the coproduction of vowels and consonants in syllables, vowels are overlaid by consonants, leading to their measured shortening. In many languages, vowel length is made phonologically distinctive and, in some of these languages (Swedish, Yawelmani, and others), rules conventionalize the reciprocal relation between vowel duration and consonant duration.

Historical Sound Change

Some historical sound changes reflect a similar reciprocal relation between vowel length and the vowel's consonantal context. These changes are called *compensatory lengthening* (e.g., Ingria, 1980) and occur when a consonant is lost in a word or set of words and a vowel in the vicinity of the consonant—formerly phonologically short—becomes long. This occurred both in Latin and Greek. Both languages lost /s/ in certain contexts. In Latin, /sisdo:/ became /si:do:/, for example, and in Greek, /ekrinsa/ became /ekri:na/ (Ingria, 1980). Phonetically, loss of a consonant should "uncover" part of a vowel's produced extent, giving it a longer measured duration. The historical change appears analogous except that the lengthening of the vowel is phonological. (However, see deChene & Anderson, 1979, for a skeptical look at the historical phenomenon of compensatory lengthening.)

Vowel Infixing[9] *and Vowel Harmony*

Languages reveal two other conventional structures suggestive of the basic organization of consonants and vowels that I have suggested. In contrast to the conventions just described, which reflect (so we suppose) the overlap of consonants and vowels in production, the following conventions may reflect the separateness of the vowel "stream" from the production of consonants. In particular,

[9] I am grateful to Judy Kegl for pointing out the relevance of McCarthy's analysis to my proposal that vowel production is continuous.

they are conventions in which phonetically nonadjacent vowels are treated in some respects as if they were adjacent (and hence a separable stream from the consonants).

In Arabic (McCarthy, 1981), derivationally related words may share a triconsonantal root. For example, words in which *ktb* occurs all have to do with the concept *to write.* Examples of words are /katab/, /ktaabab/, /kutib/, and /uktab/. McCarthy did an analysis of these word systems in which separate vocalic and consonantal tiers are proposed to underlie word generation.

To generate a particular verb form in Arabic, three choices are made. The choice of the triconsonantal root determines the word family. The choice of a "prosodic template" selects the derivational form of the verb. Finally, selection of a vocalic infix determines the voice and aspect of the verb.

The prosodic template is a word schema that specifies the numbers and orderings of the consonants and vowels in the word (e.g., CVVCVC). Some templates have more vowel slots than vowels in the infix and more consonant slots than consonants in the root. In general, consonants in the root are assigned left-to-right to the C slots, and vowels in the infix, left-to-right to the V slots of the template. If there are unfilled C or V slots, the right-most consonant or vowel is "spread" to the unfilled slots of the appropriate type. So, for example, /ktb/ and the infix /a/ (perfective, active), inserted into the template CVCVC, give /katab/ (*write*); inserted into CCVCVC, give /ktabab/.

McCarthy has captured this system's structure using a so-called "autosegmental" analysis (Goldsmith, 1976). An autosegmental approach differs from the usual segmental/suprasegmental approach in allowing several segmental tiers to underlie the expression of an utterance. Traditionally, one or two are allowed: one for phonological segments and, perhaps, another for tonal contours and other aspects of prosody. However, according to Goldsmith, utterances cannot be sliced vertically (perpendicular to the time axis) in such a way that the utterance is partitioned into coherent units. Instead, different features of the utterances start and stop at their own individually appropriate intervals and to a degree independently of the startings and stoppings of other features. In an autosegmental formulation, properties regulated separately are assigned to different tiers of a structure representing the utterance. The different tiers are related by simple rules of association.

In McCarthy's analysis, vowels and consonants are assigned to separate tiers. So, for example, /katab/ is represented by the structure in Figure 8a, and /ktabab/ by that in Figure 8b. In this kind of formulation, the "spreading" to unfilled consonant or vowel slots now can literally be a spreading. For /a/, there are no relevant segments (see discussion below of the Relevancy Condition) intervening between two V slots.

This autosegmental structure, proposed by McCarthy, obviously is compatible with the articulatory dynamics proposed to underlie syllable production. It differs from the structure, however, in being a convention of Semitic languages, not a necessary property of syllable production. Nonetheless, its existence suggests that of an underlying necessary property of production not unlike the one proposed in the first part of this article.

Another, more frequent, language convention possibly reflecting the same articulatory structure is *vowel harmony,* that is, a tendency for certain vowels to assimilate to other vowels in their neighborhood. Vowel harmony occurs in many languages including Turkish, Hungarian, Yawelmani, and Igbo. In Turkish, for example, properties of a suffix vowel are assimilated in backness and rounding to the stem vowel to which it is attached. Rules of vowel harmony operate over any number of intervening consonants. Thus, vowel harmony, like vowel infixing, is captured naturally in an autosegmental analysis in which vowels and consonants occupy separate tiers.

Vowel harmony may be an instance of a class of rules tending to conform with a constraint on phonological rules known as the *Relevancy Condition* (Jensen, 1974; Jensen & Stong-Jensen, 1979).[10] The constraint specifies the conditions under which phonological rules can refer to influences of segments on nonadjacent segments ("action at a distance").

---

[10] I thank Alan Bell for directing me to the work of Jensen and Stong-Jensen.

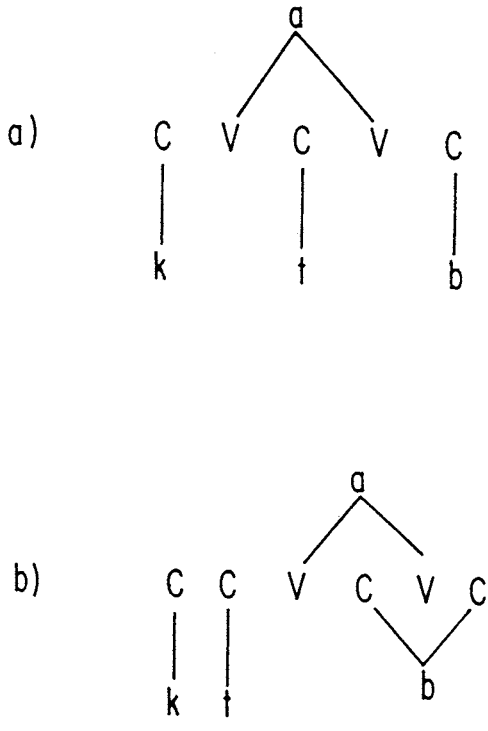

Figure 8. Vowel infixing in Arabic from McCarthy (1981).

Phonological rules may be characterized as having the following abstract form:

focus → structural change/determinant,

irrelevant segments, ___.

For example, a rule of vowel harmony in Yawelmani can be written as follows:

$$\begin{bmatrix} +\text{ syll} \\ \alpha\text{ high} \end{bmatrix} \rightarrow \begin{bmatrix} +\text{ round} \\ +\text{ back} \\ -\text{ low} \end{bmatrix} / \begin{bmatrix} +\text{ syll} \\ +\text{ round} \\ \alpha\text{ high} \end{bmatrix} C_0 \_\_\_.$$

In words, a vowel (focus) is realized as rounded, back, and nonlow (structural change) following a rounded vowel matching it in height (determinant) and by any number of intervening consonants (irrelevant segments). According to the Relevancy Condition, any features shared by the focus and the determinant (here, any vowel) define a class of *relevant segments*. The complement of that class, the irrelevant segments, serves as the "distance" over which a phonological segment can exert its effect. The influence cannot skip over relevant segments. Hence in the Yawelmani harmony rule, the irrelevant segments skipped over are all and exclusively consonants.

Conceivably, the relevancy conditions of a language may be useful in defining its autosegmental tiers. The relevant segments defined by a rule may define segments that share a tier, and irrelevant segments define a different tier or tiers. If so, it is interesting that in the examples of rules conforming to the constraint provided by Jensen and Stong-Jensen (1979), relevant segments are either consonants only or vowels only, never both.

Conclusions

Talkers

When talkers produce sequences of stressed vowels and consonants, production of the two segment types overlaps. This is shown by coarticulatory evidence, by evidence of measured shortening of vowels in consonantal contexts, and, by inference, by the existence of phonological rules in some languages that ensure a complementary relation between consonant and vowel length.

In addition, evidence suggests a degree of separateness of vowel from consonant production, which in fact allows the overlap just described. Evidence for the separation of vowel from consonant production is threefold. Coarticulation suggests it, the patterning of speech errors suggests it, and so, inferentially, does the existence of phonological rules, in which an autosegmental analysis distinguishes a vocalic from a consonantal tier.

When talkers intend to produce a rhythmic sequence of stressed monosyllables, evidence suggests they produce evenly timed vowels. Timing of syllable-initial consonants depends on the ways in which consonants or clusters are produced relative to vowels. A relaxed cyclicity in production of stressed vowels in natural speech may explain in part the impression of temporal rhythm in stress- and syllable-timed languages.

As to *why* talkers might produce speech in this way, only tentative answers may be given.

Liberman and Studdert-Kennedy (1978) suggested that speech is coarticulated ("encoded") for the listener's sake. Speech has to be produced at a rapid rate to enable retention of sufficient speech for syntactic analysis. But at the required rate, were speech a sequence of discrete sounds, listeners would be unable to recover the segments or their order (see, e.g., Warren, 1976). Coarticulation allows a large number of relatively long sounds to occupy the same interval as a much smaller number of shorter, but temporally discrete, segments. We have shown here that listeners make use of information for a vowel during the portion of the signal dominated by consonant information. This is entailed by the proposal of Liberman and Studdert-Kennedy that coarticulation *facilitates* the perceptibility of serially ordered speech sequences (see also, Shankweiler, Strange, & Verbrugge, 1977).

A second reason for separate vowel and consonant production may have to do with production rather than perception. Elsewhere (Fowler, 1977; Fowler, Rubin, Remez, & Turvey, 1980) I have proposed that talkers may exploit the fact that vowels constitute a natural articulatory class. All vowels, in contrast to consonants, are produced as relatively slow changes in the global shape of the vocal tract effected largely by movements of the tongue body and jaw.

Each particular vowel itself is a *class* of tongue body and jaw positions that yield approximately the same global vocal tract shape. This is shown by perturbation studies where, for example, talkers produce vowels clenching a bite block between the teeth so that the jaw is fixed. In these studies, the acoustic properties of the vowels are near normal (e.g., Fowler & Turvey, 1980; Lindblom, Lubker, & Gay, 1979), suggesting that tongue movement has compensated for the inability of the jaw to move. It is shown, too, by studies of coarticulation where positioning of the jaw in CV and VC syllables is affected jointly by the identity of the consonant and vowel (Sussman et al., 1973). These observations are displayed schematically in Figure 9. In the figure, each vowel is represented as a curve in a jaw–tongue coordinate space. This is meant to show the capacity that a speaker has to achieve any given vowel by a class of

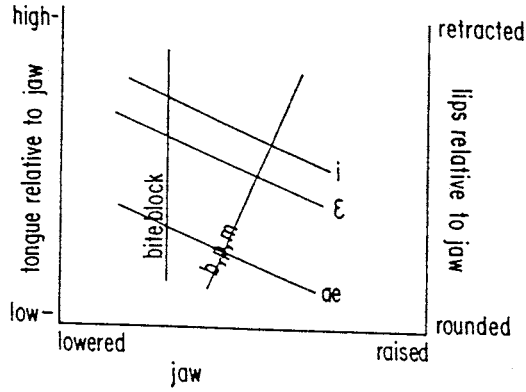

Figure 9. Schematic representation of constraints on the jaw and tongue during production of vowels /i/, /ɛ/ and /æ/ and on the jaw and lips during bilabial consonant production. (A vowel is produced by a range of negatively correlated jaw and tongue positionings that yield the same tongue–palate approximation. Similarly, a bilabial stop is realized by a variety of negatively correlated jaw and lip positionings that achieve bilabial closure; e.g., Folkins & Abbs, 1975.)

jaw positionings and tongue positionings relative to the jaw. Due to this capacity, when a bite block prevents jaw movement, or when a consonant perturbs it, all is not lost; an acceptable version of the vowel is achieved by adjusting the tongue to the special constraints on jaw position.[11]

Vowels differ one from the other largely (but not entirely) in terms of the tongue body's positioning (front/back, high/low) relative to the palate. The idea that vowels constitute a natural articulatory class is indicated in Figure 9 by showing /i/, /ɛ/, and /æ/ as if the functions for each vowel relating jaw position to the position of the tongue relative to the jaw were parallel. By hypothesis, producing a vowel, any vowel, involves organizing the musculature of the jaw and tongue body so that the two structures work in a compensatory fashion. Producing a *particular* vowel may be modeled as choosing a parameter value for the jaw–tongue relationship that ensures an "equilibrium position" for the jaw–tongue system appropriate to the selected vowel.

[11] In Figure 9, I have drawn the curves for each vowel as if they were straight lines and the lines for different vowels as if they were parallel. There is no reason to suppose that either constraint is accurate. The lines are meant to serve as schematic representations.

This proposal is analogous to Bizzi's (1978) hypothesis that pointing to positions by monkeys is achieved when the monkey establishes appropriate levels of activation of agonist and antagonist muscles in the arm. Appropriate activation levels create an equilibrium position of the arm (i.e., the position of the arm when the opposing muscle forces balance) at the target position.

What would such a system buy a talker? First, establishing a compensatory relationship between jaw and tongue may constitute an example of a general way in which movement systems responsible for reproducing positions (as opposed to movements) tend to be organized. The organizations have the advantage of "equifinality," that is, of enabling achievement of the goal position in a variety of ways without requiring reorganization (see, e.g., Keele, 1981; Kelso & Holt, 1980). This makes vowel production context sensitive.

Second, the aspects of vowel organization that hypothetically are shared among vowels may buy the talker an increment in efficiency in facilitating cyclic vowel production. Cyclic activities such as locomotion and respiration (see Grillner, 1977) are efficient in terms of the motor organizations they require. In locomotion, muscle systems are organized to generate a step. Once so organized, the same muscle systems will produce an indefinite number of subsequent steps without requiring any change in organization. Cyclic vowel production may provide another example of this kind of motor organization. If it is possible for a talker to coordinate his or her tongue and jaw in a compensatory fashion but also in a way that is *general* to the class of vowels, then once established, the organization can serve the production of vowels throughout an utterance, individual vowels being produced by cyclic reparameterizations of the tongue–jaw system.

Of course, this proposal currently begs a number of critical questions: Most important, how might the muscles of the jaw and tongue be coordinated in a compensatory fashion? Second, is the notion of a difference in values of *parameters* of an invariant organization of muscles a realistic way to describe the different jaw–tongue relations characteristic of different vowels?

However, if vowel production *were* cyclic, it would help to rationalize the linguist's and naive listener's judgments of rhythm in speech. Indeed, this is my tentative proposal, based on studies of monosyllabic stress feet and subject to revision when I turn to more natural productions in a subsequent article (Fowler, Note 4).

Listeners

The most important conclusion to be drawn about listeners' perceptions of rhythmic speech is that they mirror the natural structure of the spoken utterance. Listeners hear speech sequences largely as talkers produce them and essentially as talkers intend them to be heard.

Doing so involves hearing through coarticulatory overlap of segments, and I have shown at least one circumstance in which listeners appear to do just that (Experiment 2). I have proposed that their hearing through coarticulation is analogous to their perceptual segmentation of visually complex events and involves something like a perceptual vector analysis of the acoustic speech stream.

By interpretation, listeners hear isochronous speech when talkers produce it by attending to acoustic information specifying timing of (stressed) vowel production. In the isochronous sequences of stressed monosyllables, talkers produce vowels cyclically, and listeners attend to the timing of vowels.

Measurement

As I have argued elsewhere (Fowler & Tassinary, 1981), conventional measurements of phonological segments and measures of acoustic segments do not always reflect the psychological structure of the spoken or perceived utterance. This is not because (or only because) listeners "interpret" the acoustic message, whereas measurements are "objective" assessments. Rather, there are other possible objective segmentations of a signal than conventional ones, and the listener's perspective on the signal may constitute an alternative objective segmentation. In particular, conventions for measurement in which phonological segments are demarcated as if they were temporally discrete do not reflect

the possibly equally objective perspectives that respect coarticulatory overlap. The judgments of listeners may in the future guide decisions concerning natural measurement criteria for speech.

Sources of Evidence

Products of linguistic analysis offer a reservoir of evidence, largely untapped by psychologists, that can converge with evidence obtained from experimental investigation. Although the procedures of phonological analysis are nonexperimental, the products of the analysis, systematic phonological properties of languages, *are* behavioral regularities because they reflect language use. As such, they are relevant to psychological theories of language use including theories of speech production and perception.

Here we have used evidence from phonological analysis of language to buttress proposals that the talker's overlap of vowels and consonants is perceptually real and that separate, perhaps cyclic, vowel production is sufficiently real for language users that it gives rise to analogous phonological phenomena.

Reference Notes

1. Dalby, J., & Port, R. Temporal structure of Japanese: Segment, mora and word. In *Research in phonetics* (Report 2). Bloomington: Indiana University, Department of Linguistics, 1981.
2. Lea, W. *Prosodic aids to speech recognition: IV. A general strategy for prosodically-guided speech understanding* (Report No. PX10791). Arlington, Va.: Advanced Research Projects Agency, 1974.
3. Shen, Y., & Peterson, G. Isochronism in English. In *Studies in linguistics, Occasional Papers* (Vol. 9). Buffalo, N.Y.: University of Buffalo, 1962.
4. Fowler, C. A. *Converging sources of evidence on spoken and perceived rhythms of speech: II. Polysllabic sequences.* Manuscript in preparation, 1983.
5. Tuller, B., & Fowler, C. A. *The contribution of amplitude to the perception of isochrony* (SR-65). New Haven, Conn.: Haskins Laboratories, 1981.
6. Rapp, K. A study of syllable timing. In *Speech transmission laboratory: Quarterly progress report* (Vol. 1). Stockholm, Sweden: University of Stockholm, 1971.
7. Hollerbach, J. M. *An oscillation theory of handwriting.* Cambridge, Mass.: Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1980.
8. Raphael, L., Dorman, M., & Liberman, A. *The perception of vowel duration in VC and CVC syllables* (SR-42/43). New Haven, Conn.: Haskins Laboratories, 1975.
9. Goldstein, L. Personal communication, 1982.
10. Wright, J. The behavior of nasalized vowels in the perceptual vowel space. In *Report of the phonology laboratory, Berkeley* (Vol. 5). Berkeley: University of California, 1980.
11. Lindblom, B., & Rapp, K. Some temporal regularities of spoken Swedish. In *Papers in linguistics from the University of Stockholm* (Vol. 21). Stockholm, Sweden: University of Stockholm, 1973.

References

Abercrombie. D. Syllable quantity and enclitics in English. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, & J. L. M. Trim (Eds.), *In honour of Daniel Jones.* London: Longman, 1964.

Allen, G. The location of rhythmic stress beats in English: An experimental study. Part I. *Language and Speech,* 1972, *15,* 72–100. (a)

Allen, G. The location of rhythmic stress beats in English: An experimental study. Part II. *Language and Speech,* 1972, *15,* 170–195. (b)

Barry, W., & Kuenzel, H. Coarticulatory airflow characteristics of intervocalic voiceless plosives. *Journal of Phonetics,* 1975, *3,* 263–282.

Bell-Berti, F., & Harris, K. Anticipatory coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America,* 1979, *65,* 1268–1270.

Bizzi, E. Processes controlling arm movements in monkeys. *Science,* 1978, *201,* 1235–1237.

Bolinger, D. Pitch accent and sentence rhythm. In I. Abe & T. Kanekiyo (Eds.), *Forms of English: Accent, morpheme, order.* Cambridge, Mass.: Harvard University Press, 1965.

Butcher, A., & Weiher, E. An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics,* 1976, *4,* 59–74.

Carney, P., & Moll, K. A cineflourographic investigation of fricative consonant–vowel coarticulation. *Phonetica,* 1971, *23,* 193–202.

Catford, J. *Fundamental problems in phonetics.* Bloomington: Indiana University Press, 1977.

Chomsky, N. On cognitive structures and their development: A reply to Piaget. In M. Piattelli-Palmarini (Ed.), *Language and learning.* Cambridge, Mass.: Harvard University Press. 1980.

Classe, A. *The rhythm of English prose.* Oxford: Blackwell, 1939.

Comrie, B. Phonology: A critical review. In B. Butterworth (Ed.), *Language production, I.* London: Academic Press. 1980.

deChene B., & Anderson, S. Compensatory lengthening. *Language,* 1979, *55,* 505–535.

Donovan. A., & Darwin, C. The perceived rhythm of speech. *Proceedings of the Ninth International Congress of Phonetic Sciences,* 1979, *2,* 268–274.

Fitch, H., Halwes, T., Erickson, D., & Liberman, A. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics,* 1980, *27,* 343–350.

Folkins, J., & Abbs, J. Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Jour-*

nal of Speech and Hearing Research, 1975, *18*, 207–220.

Fowler, C. Timing control in speech production. Bloomington: Indiana University Linguistics Club, 1977.

Fowler, C. A. "Perceptual centers" in speech production and perception. *Perception & Psychophysics*, 1979, *25*, 375–388.

Fowler, C. A. Perception and production of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research*, 1981, *46*, 127–139. (a)

Fowler, C. A. A relationship between coarticulation and compensatory shortening. *Phonetica*, 1981, *38*, 35–50. (b)

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

Fowler, C. A., & Tassinary, L. Natural measurement criteria for speech: The anisochrony illusion. In J. Long & A. Baddeley (Eds.), *Attention and performance, IX*. Hillsdale, N.J.: Erlbaum, 1981.

Fowler, C. A., & Turvey, M. T. Immediate compensation in bite-block speech. *Phonetica*, 1980, *37*, 306–326.

Goldsmith, J. *Autosegmental phonology*. Bloomington: Indiana University Linguistics Club, 1976.

Grillner, S. Locomotion in vertebrates. *Physiological Reviews*, 1975, *55*, 247–304.

Grillner, S. On the neural control of movement—A comparison of different basic rhythmic behaviors. In G. S. Stent (Ed.), *Function and formation of neural systems*. Berlin: Dahlem, 1977.

Han, M. The feature of duration in Japanese. *Study of Sounds*, 1962, *10*, 65–75.

Hombert, J.-M., Ohala, & Ewan, W. Phonetic explanation for the development of tones. *Language*, 1979, *55*, 37–58.

House, A., & Fairbanks, G. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, 1953, *25*, 105–113.

Ingria, R. Compensatory lengthening as a metrical phenomenon. *Linguistic Inquiry*, 1980, *11*, 465–495.

Jensen, J. A constraint on variables in phonology. *Language*, 1974, *50*, 675–686.

Jensen, J., & Stong-Jensen, M. The Relevancy Condition and variables in phonology. *Linguistic Analysis*, 1979, *5*, 125–160.

Johansson, G. *Configurations in event perception*. Uppsala, Sweden: Almqvist and Wiksell, 1950.

Johansson, G. Projective transformations as determining visual space perception. In R. MacLeod & H. Pick (Eds.), *Perception: Essays in honor of James J. Gibson*. Ithaca, N.Y.: Cornell University Press, 1974.

Keele, S. Behavioral analysis of movement. In V. Brooks (Ed.), *Handbook of physiology: Motor control*. Washington: American Physiological Society, 1981.

Kelso, J. A. S., & Holt, K. Evidence for a mass-spring model of human neuromuscular control. In C. Nadeau, W. Halliwell, K. Newell, & G. Roberts (Eds.), *Psychology of motor behavior and sport*. Champaign, Ill.: Human Kinetics, 1980.

Kenstowicz, M. J., & Kisseberth, C. W. *Generative phonology: Description and theory*. New York: Academic Press, 1979.

Kent, R., & Moll, K. Tongue body articulation during vowel and diphthong gestures. *Folia phoniatrica*, 1972, *24*, 278–300.

Klatt, D. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 1976, *59*, 1208–1221.

Lehiste, I. *Suprasegmentals*. Cambridge, Mass.: MIT Press, 1970.

Lehiste, I. Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America*, 1972, *54*, 1228–1234.

Liberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, *74*, 431–461.

Liberman, A., & Studdert-Kennedy, M. Phonetic perception. In H. Leibowitz & H.-L. Teuber (Eds.), *Handbook of sensory physiology, Vol. VIII: Perception*. Berlin, West Germany: Springer-Verlag, 1978.

Lindblom, B., Lubker, J., & Gay, T. Formant frequencies of some fixed-mandible vowels and a model of speech-motor programming by predictive simulation. *Journal of Phonetics*, 1979, *7*, 147–161.

Lindblom, B., Lyberg, B., & Holmgren, K. *Durational patterns of Swedish phonology: Do they reflect short-term memory processes?* Bloomington: Indiana University Linguistics Club, 1981.

Lisker, L. On time and timing in speech. In T. Sebeok (Ed.), *Current trends in linguistics, 12*. The Hague: Mouton, 1972.

Locke, J. The child's processing of phonology. In W. A. Collins (Ed.), *Minnesota Symposium on Child Psychology* (Vol. 17). Hillsdale, N.J.: Erlbaum, 1979.

MacNeilage, P., & DeClerk, J. On the motor control of coarticulation in CVC monosyllables. *Journal of the Acoustical Society of America*, 1969, *45*, 1217–1233.

Marcus, S. Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics*, 1981, *30*, 247–256.

McCarthy, J. A. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry*, 1981, *12*, 373–418.

Morton, J., Marcus, S., & Frankish, C. Perceptual centers (P-centers). *Psychological Review*, 1976, *83*, 405–408.

Ohala, J. Experimental historical phonology. In J. Anderson & C. Jones (Eds.), *Historical linguistics, II: Theory and description in phonology*. Amsterdam: North-Holland, 1974.

Ohala, J. The production of tone. In V. Fromkin (Ed.), *Tone: A linguistic survey*. New York: Academic Press, 1978.

Ohala, J. The listener as a source of sound change. In M. F. Miller (Ed.), *Papers from the parasession on language behavior*. Chicago: Chicago Linguistic Association, 1981.

Öhman, S. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 1966, *39*, 151–168.

Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. Infant babbling and speech. *Journal of Child Language*, 1976, *3*, 1–11.

Perkell, J. *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge, Mass.: MIT Press, 1969.

Pike, K. *Intonation of American English*. Ann Arbor: University of Michigan Press, 1945.

Raphael, L. Preceding vowel duration as a cue to the

perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America*, 1972, *51*, 1296–1303.

Raphael, L., & Dorman, M. The contribution of CV transition duration to the perception of final-consonant voicing in natural speech. *Journal of the Acoustical Society of America*, 1980, *67*, S51.

Shaffer, L. H. Rhythm and timing in skill. *Psychological Review*, 1982, *89*, 109–122.

Shankweiler, D., Strange, W., & Verbrugge, R. Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving acting and knowing: Toward an ecological psychology*. Hillsdale, N.J.: Erlbaum, 1977.

Sussman, H., MacNeilage, P., & Hanson, R. Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. *Journal of Speech and Hearing Research*, 1973, *16*, 397–420.

Tuller, B., & Fowler, C. A. Some articulatory correlates of perceptual isochrony. *Perception & Psychophysics*, 1980, *27*, 277–283.

Tuller, B., Harris, K., & Kelso, J. A. S. Stress and rate: Differential transformations of articulation. *Journal of the Acoustical Society of America*, in press.

Tuller, B., Kelso, J. A. S., & Harris, K. Interarticular phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology: Human Performance and Perception*, 1982, *8*, 460–472.

Viviani, P., & Terzuolo, B. Space-time invariance in learned motor skills. In G. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland, 1980.

Warren, R. Auditory sequence and classification. In N. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press, 1976.

Wing, A. Response timing in handwriting. In G. Stelmach (Ed.), *Information processing in motor control and learning*. New York: Academic Press, 1978.

Received June 25, 1982
Revision received October 19, 1982 ■