Perception & Psychophysics
1979, Vol. 25 (5), 375-388

"Perceptual centers" in speech production and perception

CAROL A. FOWLER
Dartmouth College, Hanover, New Hampshire 03755
and Haskins Laboratories, New Haven, Connecticut 06510

Morton, Marcus, and Frankish (1976) report that listeners hear acoustically isochronous digit sequences as anisochronous. Moreover, given a chance to adjust intervals in the sequences until they are perceptually isochronous, the listeners introduce systematic deviations from isochrony. The present series of studies investigates these phenomena further. They indicate that when asked to *produce* isochronous sequences, talkers generate precisely the acoustic anisochronies that listeners require in order to *hear* a sequence as isochronous. The acoustic anisochronies that talkers produce are expected if talkers initiate the articulation of successive items in the sequence at temporally equidistant intervals. Items whose initial consonants differ in respect to manner class will have acoustic consequences (other than silence) at different lags with respect to their articulatory onsets thereby generating the observed acoustic anisochronies. The findings suggest that listeners judge isochrony based on acoustic information about *articulatory* timing rather than on some articulation-free acoustic basis.

Sequences of spoken digits are judged to be unevenly timed when the digits are presented at acoustically regular intervals. Moreover, given an opportunity to adjust the intervals to make them sound isochronous, subjects introduce systematic deviations from acoustic regularity (Morton, Marcus, & Frankish, 1976). The deviations are such that long intervals are interposed between a digit starting with a consonant of long acoustic duration and one starting with a short-duration consonant or a vowel (e.g., six-eight); correspondingly short intervals are inserted between the same digit types presented in reverse order (e.g., eight-six).

These findings disconfirm a hypothesis about listeners' judgments of rhythmicity in speech that, a priori, seems simple and plausible—namely that listeners base rhythmicity judgments on the intervals between the onsets of acoustic energy of successive syllables (or, more likely, perhaps, of successive *stressed* syllables) in the sequence.

Morton et al. coin the term "perceptual center" or "P-center" to reference the locus in a word that must be equidistant, temporally, from corresponding loci in surrounding words in order for the sequence to sound isochronous to a perceiver. Thus, the perceptual center is the "psychological moment of occurrence" of a word.

This research was supported by NIH Grant HD 01994 and NSF Grant NS 13617 to Haskins Laboratories. I thank Betty Tuller and George Wolford for reviewing earlier versions of the manuscript.

The interval-adjustment technique used by Morton et al. did not enable them to locate the P-centers of the different digits, but only to discover the relative temporal alignments of the acoustic onsets of successive digits in a sequence that would make the sequence sound isochronous. As noted, the critical variable affecting the temporal alignment of a digit with respect to surrounding digits was the duration of its acoustic energy prior to the acoustic onset of its vowel. Morton et al. were unable to discover any obvious acoustic markers of the P-centers. They excluded as markers the onset of the word, the onset of the stressed vowel, and the word's or vowel's peak intensity.

Two other experimental investigations, both designed to discover the locus of the stress beat in a word (Allen, 1972; Rapp, Note 1) apparently do pinpoint the P-center, although neither demonstrates how it is marked acoustically. In Allen's study, subjects listened to sentences, each presented repeatedly on a tape loop. During a block of 50 repetitions of the same sentence, subjects tapped their fingers "on the beat" of a designated syllable in the sentence. Over blocks of trials, the subjects tapped to different syllables in the sentence, and over experimental sessions, they listened to different sentences. Allen found that subjects' taps tended to be located near the onset of acoustic energy for the vowel in a stressed syllable, but to precede the vowel's onset by a duration that correlated positively ($r = .6$) with the duration of acoustic energy of the prevocalic consonant or consonant cluster. For example, the tap preceded the vowel onset by 19 msec on the average

0031-5117/79/050375-14$01.65/0

when the vowel followed a (short-duration) voiced stop, but by 96 msec when it followed a (long-duration) consonant cluster. Rapp (Note 1) found a much higher correlation of precisely the same kind when she asked talkers to repeat various nonsense words in time with a regularly occurring pulse. In this instance, talkers located a nonsense disyllable on the pulse so that the pulse preceded the (second) stressed vowel's acoustic onset by a variable duration. The duration varied directly with the duration of the prevocalic consonant or consonant cluster.

It is evident in Rapp's data that despite the progressive shift *backward* of the stress beat with increases in prevocalic consonant duration, the stress beat tends also to *follow* the acoustic syllable onset by increasingly longer intervals the greater the prevocalic consonant duration. Assuming that stress beats are P-centers and that P-centers are aligned temporally in perceptually rhythmic utterances, this latter outcome of Rapp's predicts the acoustic anisochronies reported by Morton et al. That is, syllables with acoustically long prevocalic consonants should be located temporally closer to their predecessor than syllables with acoustically short-duration consonants. Thus, evidently, the data of Rapp and of Allen are compatible with those of Morton et al., but add information about P-center location in a word or syllable.

These three rather different experimental procedures agree that a word's stress beat or psychological moment of occurrence does not correspond to the word's acoustic onset, to the acoustic onset of its stressed vowel, or to any other obvious acoustic marker. Two of the studies locate the beat within the prevocalic consonant or consonant cluster, but at a variable distance both from the vowel's acoustic onset and from the onset of the word.

Among other areas of investigation, these findings require consideration in examinations of the production and perception of stress-timed rhythms in English. In respect to production, several linguists have claimed that English is a stress-timed language—a language, that is, in which the intervals between major stresses in a naturally produced utterance are fairly uniform in duration (see, for example, Abercrombie, 1964; Classe, 1939, cited in Lehiste, 1973; Pike, 1945). But these claims are based primarily on intuitive judgments, and they may have been disconfirmed by several experimental tests. Tests of the stress-timing claim have shown clearly that the acoustically defined intervals between stressed-syllable onsets vary enormously in duration within an utterance (Duckworth, 1965; Lehiste, 1973; Lea, Note 2; Shen & Peterson, Note 3).[1]

The work of Morton et al., however, suggests a need to reexamine these disconfirmations. If it is the case that talkers regulate the same intervals that listeners judge—that is, those between P-centers and not those between stressed syllable onsets—these disconfirmations may be spurious. The variability among interstress intervals might be substantially reduced were the intervals measured that talkers in fact regulate.[2]

In the perceptual domain, some investigators (for example, Coleman, 1974; Lehiste, 1973) have suggested that stress timing is largely an imposition by a listener, not by a talker. Indeed, Lehiste finds that naive listeners (no less than the linguists cited above who first posed the stress-timing claim) perceive speech sequences, but not nonspeech analogues, to be more stress timed than acoustic measurements substantiate (see also Coleman, 1974).

However, a proposal that listeners *impose* a stress-timing rhythm on an arhythmic utterance is difficult to rationalize, and it is rendered implausible by other perceptual data. Some evidence, obtained primarily in phoneme-targeting experiments, (e.g., Cutler, 1975; Shields, McHugh, & Martin, 1974; see also Martin, 1970) suggests that subjects know with some precision *when* a stressed syllable (but not an unstressed syllable) is due to occur in a sequence they are monitoring. This evidence implies strongly that spoken utterances have some stress-based rhythm that listeners are able to track. The simplest stress-based rhythm that the perceptual data promote as plausible is stress timing. The evidence of Morton et al., Allen, and Rapp suggests that the rhythmic intervals may be bounded by P-centers. Thus, it is possible that the acoustic measurements of stress timing in the Lehiste study might better have mirrored listeners' judgments had the interval edges been established at the P-centers.

The suprasegmental rhythms of speech, including stress timing perhaps, most probably are symptoms of those aspects of articulatory control having to do with coordinating the various structures of the vocal tract, the larynx and the respiratory system. In particular, they may manifest the workings of articulatory controls whose regulated events are temporally substantially coarser-grained than the durations of individual phonetic segments. A proposal that speech is stress-timed, then, is a claim about a talker's style of coarse-grained articulatory control, which suggests that one controlled, coherent, event in speaking is the interval between stresses. A proposal like this is fairly easy to understand if the interstress interval begins and ends at the onsets of stressed syllables. It is far less comprehensible, on the surface anyway, if the interval is bounded by P-centers, since P-centers do not correspond in any obvious way to the edges of linguistic or acoustic units.

The present experiments were designed to investigate further the perceptual phenomenon found by Morton et al. In particular, the experiments had two major aims. The first was to rationalize the finding that listeners require perceptually stress-timed utterances to be acoustically anisochronous in particular ways. Morton et al. sought an acoustic explanation for the phenomenon. Here, an articulatory account is developed. The second experimental aim was parasitic on the first; it was to obtain information enabling an interpretation of the stress-timing hypothesis, preparatory to reevaluating it.

The first two experiments ask, respectively, whether talkers produce anisochronous rhythms when they are told to talk rhythmically and, if so, whether their deviations from isochrony are just those that listeners require in order to perceive a sequence as stress timed. Experiments 1 and 2 provide affirmative answers to both of these questions. The third experiment generalizes the production findings to a slightly more natural type of utterance. The final two experiments further examine an articulatory account of the acoustic deviations from isochrony reported by Morton et al.

The results of these experiments suggest that talkers may well produce stress-timed *gestures* of the vocal tract on request. However, these gestures of the vocal tract produce an anisochronous acoustic signal if the initial consonants of successive syllables are different (primarily, if they differ in manner class). For their part, listeners behave as if the acoustic speech signal provides information about the timing of the talkers' *articulatory* gestures, and as if they base their rhythmicity judgments on that acoustic information about gestural timing. In this respect, the evidence of the present experiments concerning the perception of articulatory rhythms is highly compatible with that of Liberman and his colleagues (see Liberman & Pisoni, 1977, for a review of these studies) relating to the perception of phonetic segments. In particular, both sets of studies show that silence in an utterance provides critical information to a perceiver of speech, putatively about those articulatory gestures that have silence as an acoustic correlate.

The present experiments provide a simple articulatory account of the acoustic deviations from rhythmicity reported by Morton et al. However, they do not immediately explain the variable locus of the P-center as found by Allen (1972) and Rapp (Note 1). Thus, they do not suggest what the boundaries of a stress-timed interval might be. A plausible hypothesis about P-center identity can be formulated in articulatory terms however; this hypothesis will be given in the General Discussion.

EXPERIMENT 1

The first study asks whether talkers produce acoustically stress-timed utterances when instructed to or, instead, if their utterances deviate from stress timing in systematic ways.

Method

One subject, naive to the purposes of the experiment, was asked to produce a series of nonsense sentences each composed of six monosyllables. The monosyllables all rhymed with /ad/, but differed in initial consonant. A nonsense sentence was either homogeneous in composition—that is, its component words were the same (e.g., "mad mad mad mad mad mad")—or it was alternating. An alternating utterance was composed of two different nonsense words produced in alternation (e.g., "mad sad mad sad mad sad"). The syllables, ad, bad, mad, nad, tad, sad, composed the vocabulary from which the homogeneous and alternating utterances (each syllable paired with every other) were produced in all.

The subject, a male adult, was asked to speak at a slow rhythmic rate, stressing every syllable. His utterances were recorded on tape, and sound spectrograms were made of each one. Compatibly with the measures of Morton et al. and with those of the above cited investigators of stress timing in speech, measurements were made of the intervals between acoustic onsets of the syllables in each utterance. Measurements were made in millimeters (1 mm = 7.5 msec). To avoid contamination of the rhythmicity effects by utterance-initial and -final lengthening (Klatt, 1976; Oller, 1973; Lindblom & Rapp, Note 4), only the intervals between Syllables 2 and 3, 3 and 4, and 4 and 5 were measured. These interstress intervals will be called, respectively, ISI_2, ISI_3, and ISI_4.

Results

The relevant measurements are presented in Table 1. They are absolute mean differences in ISI duration for the homogeneous and alternating utterances. Thus, the first column of data in Table 1 presents the mean difference in duration between ISI_2 and ISI_3 for the homogeneous and alternating utterances. The second column presents the mean difference between ISI_3 and ISI_4. And the third column presents the difference between ISI_2 and ISI_4. Absolute values of the individual difference scores were used in the computations of the mean difference values.

Table 1 verifies that the homogeneous utterances were produced in a near stress-timed rhythm. The mean deviations from isochrony ranged between 19.7 and 27.5 msec. These deviations are rather small given that the average interval duration was 474 msec for the homogeneous utterances. An analysis of variance (with initial-segment identity as the random factor) shows that the three deviation values do not differ significantly [$F(2,10) < 1$].

Table 1
Absolute Values of Durational Differences (in Milliseconds) Among the Interstress Intervals of Homogeneous and Alternating Utterances in Experiment 1

	$\lvert ISI_2\text{-}ISI_3 \rvert$		$\lvert ISI_3\text{-}ISI_4 \rvert$		$\lvert ISI_2\text{-}ISI_4 \rvert$	
	Mean	SD	Mean	SD	Mean	SD
Homogeneous	20.5	11.9	27.5	18.8	19.7	17.8
Alternating	116.0	80.2	120.0	70.1	19.4	18.4

In contrast to this are the corresponding values for the alternating utterances. Here Table 1 shows that ISI_2 and ISI_3 have durations that differ by 116 msec on the average. Likewise, ISI_3 and ISI_4 differ by 120 msec, whereas ISI_2 and ISI_4 differ only by 19.4 msec. Recall that in an alternating utterance (e.g., "mad sad mad sad mad sad"), ISI_2 (the interval between the first "sad" and the second "mad") and ISI_3 (the interval between the second "mad" and the second "sad") are different in that their initial and final consonants are reversed with respect to each other. Likewise for ISI_3 and ISI_4. But ISI_2 and ISI_4 are alike—they start with the same consonant (/s/ in the present example) and end with the same consonant (here /m/).

An analysis of variance of the three mean difference scores for the alternating utterances verifies that the values differ significantly [$F(2,28) = 19.6$, $p < .0001$]. Scheffé's tests attribute the significant F value to the difference between the value 19.4 and the other two values, 116 and 120. The latter two values do not differ significantly.

Parallel to the perceptual findings of Morton et al., among the alternating utterances of the present study, the *degree* of deviation from isochrony of an utterance is closely related to the relative (acoustic) durations of the component syllable's prevocalic acoustic portions. For example, the average deviations from isochrony of syllable onsets (i.e., the average of Columns 1-3 in Table 1) was 191 for the utterance, "bad sad bad sad . . . ," but was only 44 msec for "bad ad bad ad . . . ." The prevocalic acoustic signal in "bad" is less than 5 msec in duration. This contrasts with 120 msec in "sad" and 0 in "ad." The Pearson product-moment correlation between the average deviation from isochrony and the average difference in duration between the component prevocalic consonants of an alternating utterance is .92, a highly significant value ($p < .001$).

Discussion

Although the experiment provides a rather limited amount of data, the patterning is clear. When the talker was asked to produce stress-timed utterances, his productions deviated from *acoustic* isochrony in systematic ways. The deviations are like those created by the listener-subjects of Morton et al. out of acoustically isochronous digit sequences. For both the talker of the present experiment and the listeners described by Morton et al., intervals beginning with acoustically long-duration consonants and ending with short-duration consonants were long relative to intervals that were just the reverse of this. This compatibility across the two experiments strongly suggests that in order to hear an utterance as stress-timed, listeners require precisely the deviations from acoustic isochrony that talkers create when they are asked to produce a stress-timed sequence. Experiment 2 establishes the agreement between talkers and listeners using the production data from Experiment 1.

EXPERIMENT 2

Method

Stimuli. Twelve utterances were selected from those of the preceding experiment. They were chosen from among the 15 alternating utterances and were the 12 whose deviations from isochrony were largest. Two versions of each utterance were constructed using the PCM system at Haskins Laboratories. One version of each utterance consisted of the middle four syllables of the original six-syllable sentence with the ISIs unaltered. The second version was constructed from the first so that the three ISIs were equal or nearly equal in duration. Isochrony was achieved by electronically splicing silence onto an interval, or less often by deleting silence from intervals of the first versions of each utterance. No portion of an utterance was deleted in which acoustic energy was present (except that the first and sixth syllables were deleted as noted above); only silence was added or removed.

The mean deviations from isochrony among the ISIs for the natural and altered versions of the 12 utterances used in this study are, respectively, 125 and 17 msec. These values verify that the altered versions were substantially more homogeneous in interval duration than were the naturally produced versions.

The mean ISI duration for the altered versions was 473 msec, while that for the natural versions was 416 msec. This difference was due to the necessity, for the most part, of *adding* silence to natural intervals to achieve isochrony, rather than deleting silence. Often, deleting a requisite duration from a long interval was impossible because it would have entailed deleting part of the acoustic signal. However, this difference in mean interval duration is not large and, in any case, is comparable to the (unexplained) difference in the mean duration of homogeneous-utterance ISIs (474 msec) and alternating-utterance ISIs (418 msec) found in the first experiment. Thus, the ISIs of the altered utterances in the present experiment are within the range of ISIs that are naturally produced in utterances with monosyllabic ISIs. But they may better correspond to the ISIs of naturally produced homogeneous and (like the altered utterances) acoustically isochronous utterances than to alternating, anisochronous utterances.

In the case of one utterance ("bad tad bad tad"), silence was added to all three intervals (in approximately equal amounts), but *not* so as to alter significantly the degree of anisochrony of the utterance. The difference in duration between the longest and shortest ISI in the natural version was 60 msec; in the constructed version it was 53 msec. However, the mean ISI duration in the original version was 405 msec, while in the altered version it was 465 msec. This utterance pair was used as a "catch trial" in a way that is explained below.

On each of 12 trials, the natural version of a sentence was paired with its altered, isochronous, counterpart. The natural version occurred in first position of the pair on six trials, and the isochronous version, on the remaining six trials. The ordering of trials with respect to this manipulation was random. Between the first and second member of each utterance pair, 1,500 msec intervened; each pair was repeated once. Four and a half seconds intervened between trials.

Procedure. The subjects' task on hearing each pair was to indicate whether the first or the second version was the more "rhythmic." "Rhythmic" was defined for the purposes of the experiment as equality of intervals between syllable onsets. If a subject judges rhythmicity by comparing the intervals between acoustic syllable onsets, then he should choose the isochronous version on each trial. On the other hand, if he listens for the same intervals that talkers regulate (perhaps the intervals between

P-centers), then he should choose the natural version on each trial.

The catch trial provides a check on the basis for the subjects' choices. If, rather than choosing the perceptually more rhythmic versions of a sentence pair as instructed, the subjects choose on the basis of something like perceived naturalness (e.g., because the durations of the intersyllable pauses in the altered versions may be unnaturally long), then they should choose the natural version on the catch trial as often as they choose it on the other trials. But if they are choosing on the basis of perceived rhythmicity, they should choose the natural version of the catch trial about half the time.

Subjects. Ten students enrolled in the Introductory Psychology course at the University of Connecticut participated in the experiment in exchange for course credit.

Results

Subjects chose the natural anisochronous version of each sentence pair with far greater than chance frequency. On the 11 trials excluding the catch trial, the natural version was chosen 9.8 times on the average. This value differs significantly from the chance value of 5.5 according to a paired t test [$t(9) = 9.22$, $p < .0001$]. Half of the subjects chose the natural version on all 11 trials.

On the average, 9 of the 10 subjects chose the natural version on any given trial. One of the natural utterances was chosen by only 7 subjects; the remaining ones were selected by 8, 9, or all 10 subjects. In contrast, 6 of the 10 subjects chose the natural version of the catch trial. This difference was not due to the natural version of the catch trial utterance having been minimally deviant from acoustic isochrony in the first place, since 10 of the 10 subjects chose the natural version of another utterance ("mad sad mad sad") that was as little deviant from stress-timing as the catch trial utterance "bad tad bad tad." If there was, indeed, a preference on the subjects' part to choose the natural version of each pair regardless of rhythmicity considerations, it was not of sufficient magnitude to account for the overwhelming tendency in addition to choose the acoustically anisochronous version of each pair.

Discussion

The results of the first experiment suggest that when a talker is asked to produce a rhythmic utterance, he regulates something other than the acoustic onsets of its component syllables. The experiment does not establish what interval is regulated. However, whatever the interval may be, Experiment 2 shows that it is the same interval to which listeners attend when asked to judge the rhythmicity of an utterance.

Since talkers are responsible for the acoustic anisochronies of perceptually stress-timed utterances, it is reasonable to look to the dynamics of articulation for their explanation. However, an articulatory account is only of interest if these anisochronies occur in natural utterances as well as in lists of the sort investigated in the studies of Morton et al. and in Experiment 1. Therefore, consideration of an articulation-based rationale for the P-center phenomenon is deferred until the generality of the articulatory findings of Experiment 1 is tested.

EXPERIMENT 3

Optimally, the generality and naturalness of the findings are tested by examining a large corpus of unconstrained conversation. Under these conditions, if words beginning with acoustically long-duration consonants tend to be "located" temporally closer to preceding words than words starting with short-duration consonants, we could conclude that these timing effects are general to the production of speech and are not peculiar to rhythmic nonsense strings. (These results would not permit any conclusions to be drawn about stress-timing in natural speech, since, as the Discussion will suggest, they may occur for reasons unrelated to the production of speech prosody.)

However, at present, a procedure involving unconstrained conversation is difficult to implement properly. First, the conditions under which the P-center phenomenon may be clearly observed have not been mapped out in any detail. There is some evidence (Marcus, Note 5) that the findings of systematic deviations from acoustic isochrony are somewhat obscured when disyllables are substituted for some of the monosyllables in a string. The reason for this is unknown, but it may have to do with special timing regulations that surround the production of unstressed syllables (see Fowler, Note 6). These timing effects may *interact* with P-center-related effects and may obscure them. Until they are studied systematically, and until the implications with respect to the P-center concept are understood, it is better to control for them than to let them vary freely as they do in ordinary conversation. Therefore, in this preliminary investigation, the utterances produced are restricted to sentences that consist only of monosyllabic stressed syllables.

A second consideration that obviates the use of informal conversation as a data base has to do with the phonetic composition of an ISI. Of particular concern, the acoustic interval between a syllable-final consonant and a following phonetic segment may vary with the degree of articulatory incompatibility between the two segments. It is important that this effect on ISI duration not contribute in a biased way to assessments of temporal alignments of stressed syllables as it is likely to in ordinary conversation. Thus, a preferable procedure is to vary the identity of critical syllable-final consonants and that of syllable-initial segments orthogonally rather than at random. In the present experiment, a more feasible, but somewhat less adequate, procedure than this was adopted. Here the critical syllable-final consonant was held *constant* over variation in the following syllable initial segment. The velar stop /k/ was selected as the critical syllable-final consonant, while the neighboring syllable-initial segments had alveolar or labial places of articulation. The production of these segments should be minimally and equivalently incompatible with the production of the preceding /k/.

Method

The sentence frame used in the experiment was "Jack likes black – – – – –" (borrowed from Lehiste, 1973) and the words "acts, bats, mats, gnats, tacks," and "sacks" were spoken in the frame. The sentences were printed on 3 × 5 in. file cards and were presented to subjects in random order. Following several practice runs to familiarize subjects with the utterances, the test trials began. On each trial, the subject was given a file card and was asked to read the sentence on the card at a slow natural rate. Each sentence was produced three times; the same sentence was never produced twice consecutively. Subjects were tested in a quiet room, and their productions were recorded on audio tape.

Subjects. The subjects were two students in the Introductory Psychology course at the University of Connecticut. They participated in the experiment in exchange for course credit.

Measurements. Spectrograms were made from the recorded utterances and the following intervals were measured: (1) The offset of voicing in "black" to the onset of the test word. (This measurement was made rather than that between /k/ release and test-word onset because it was easier to detect reliably. There is no reason to suppose that this choice of interval should affect the results in any way.) (2) The acoustic onset of the test word to the acoustic onset of its vowel (the acoustic onset of a voiced formant pattern).

Results

For each subject, the measurements made of a given utterance type were averaged across its three repetitions. Then correlations were computed between the pair of measurements described in the Methods section (the interword interval and the test word's prevocalic acoustic duration). For one subject, the Pearson product-moment correlation is $-.96$ ($p < .001$); for the other subject, it is $-.75$ ($p = .02$). Thus, long intervals are interposed between "black" and short-duration syllable-initial consonant, while short intervals are interposed between "black" and a long-duration syllable-initial consonant. This is precisely the result reported by Morton et al. and replicated in Experiments 1 and 2.

Discussion

The results of this experiment and of Experiment 1 may be given a simple articulation-based explanation. The independent variable in each experiment is the identity of certain syllable-initial phonetic segments. The initial consonants vary in voicing, place of articulation, and most importantly, in manner of articulation. The acoustic phenomena observed in the data from the two experiments most probably arise in part from differences in the manners of articulation of the consonants and in part from differences in other aspects of their articulatory character, including the velocities of their closing gestures. Let us consider how these variables might produce the observed acoustic anisochronies.

Some of the consonants, both in Experiment 1 and in Experiment 3, are stops. Stops are produced, in part, by occluding the vocal tract when the place of articulation is reached. Thus, in the production of /b/ and /p/, the lips are shut temporarily; in /d/ and /t/, the tongue tip occludes the vocal tract at the alveolar ridge; and in /g/ and /k/, the tongue body occludes the vocal tract at the velum. In all cases, the acoustic correlate of the occlusion is silence.

In contrast to the stops are the fricatives (e.g., /s,z,f,v/) and the nasals (/m,n/). In the production of fricatives, the vocal tract is obstructed, but it is not occluded entirely. Thus, in producing /s/ and /z/, the tongue tip approaches the alveolar ridge, but does not prevent the passage of air through that location. Likewise, for /f/ and /v/, the air passage is restricted at the lips, but is not closed. Therefore, there is no reason, having to do with vocal tract obstruction anyway, for a period of silence to precede the production of a fricative. Similarly for the nasals, /m/ and /n/. Here the oral cavity *is* occluded; however, the nasal cavity is open and allows the air to escape through the nose.

On these articulatory grounds, other things being equal, one would expect a relatively long interval between the offset of a syllable and the onset of acoustic energy for a syllable-initial stop relative to a fricative or nasal. Consequently ISIs ending with a stop should be longer than those ending with a segment of any other manner class. Experiments 1 and 3 verify this prediciton.

In addition to this major effect that the manner class of a consonant may have on an ISI, there is another set of possible production-based influences, including the articulatory velocity of a consonant's closing gesture and, for the stops, the closure interval. These variables are not independent of the linguistic variable, manner class; nor are they entirely redundant with it. Therefore, they are considered separately from manner class here.

The recent data of Kuehn and Moll (1976) indicate that vocal tract closing gestures by the primary articulators for consonants differ in velocity as a function of several variables. Among the critical variables affecting articulatory velocity are the distance that the articulator has to move to achieve closure or near closure (the larger the distance, the faster the movement), and possibly the manner class of the consonant (some fricatives may be produced more slowly than stops).

Other data suggest that the voicing characteristics of the stop consonants also affect articulatory velocity, the closing gestures for the voiceless stops being produced more rapidly than those for the voiced stops (see MacNeilage & Ladefoged, 1976, for a review). [Generally, the voiceless/voiced comparison has been made between stops sharing place of articulation, /p/ being faster than /b/, /t/ faster than /d/, and /k/ faster than /g/. However, the data of Kuehn and Moll (1976) indicates little if any difference among the voiceless stops in a variable related to time to closure—that is, articulatory velocity with the effect of displacement on velocity subtracted out. Therefore, we will assume that voiceless stops as a class are produced more rapidly than voiced stops.]

Other things being equal, these differences in articulatory velocity should lead to differences in the time after articulatory onset that the different consonant segments have acoustic consequences. Syllable-initial consonants that are produced slowly will have acoustic consequences late relative to rapidly produced segments and therefore will terminate an ISI relatively late. In consequence (other things being equal), ISIs ending in voiced stops will be longer than ISIs ending in voiceless stops.

Other things are not entirely equal, however. The closure interval itself is longer in voiceless stops than it is in voiced stops (e.g., Lisker, 1957). This offsets the effect of articulatory velocity on ISI duration. However, data presented in Kozhevnikov and Chistovich (1965) and the acoustic data in Rapp (Note 1) both agree that in some contexts the differences between voiced and voiceless stops in articulatory velocity exceed the reverse differences in closure duration. Thus, the summed effects of the two variables, articulatory velocity of the closing gesture (or, better, time to closure) and closure duration, should lead to an earlier onset of acoustic energy (relative to articulatory onset) for voiceless stops than for voiced stops. Port's (Note 7) data disagree with those of Kozhevnikov and Chistovich and Rapp. He finds that differences among voiced and voiceless stops in respect to closure duration *exceed* their differential effects on preceding vowel duration—a measure of articulatory velocity of the closing gesture for the stop. However, his data are on stops that follow stressed /i/ or /I/ and that initiate an unstressed syllable. The data of Rapp and of Kozhevnikov and Chistovich are on stressed syllable-initial stops following stressed (Kozhevnikov & Chistovich) or unstressed (Rapp) /a/. Evidently, the variables of articulatory velocity of the closing gesture and of closure interval will affect the time after articulatory onset that acoustic consequences occur other than silence. However, the particular effect that they jointly have may differ depending on variables such as the identity of any preceding vowel and on the stress patterning of the utterance.

Experiments 1 and 3 each offer only a single assessment of the prediction that time to closure and closure duration of a segment ending an ISI will affect its duration, since, among the stops, only /b/ and /t/ were included as stimuli. In both experiments, the stops initiated a stressed syllable, as in the studies of Rapp and Kozhevnikov and Chistovich. Therefore, the appropriate prediction would be that the preacoustic articulatory duration of /b/ should exceed that for /t/. In consequence, ISIs ending in /b/ should exceed those ending in /t/. In Experiment 1, ISIs ending in /b/ averaged 486 msec while those ending with /t/ averaged 430 msec [$t(11) = 2.35$, $p < .05$]. In Experiment 3, the interval between the acoustic onsets of "black" and "bats" averaged 403 and 392 msec for Subjects 1 and 2, respectively, while that between "black" and "tacks" averaged 370 and 377 msec. These differences are as predicted, but are probably exaggerated relative to expected differences due to the articulatory variables, time to closure and closure duration.

Since the consonant durations themselves were measured by Morton et al. and, following them, in the present experiments from the onsets of their acoustic energy to vowel onsets, voiced stops are measured to be acoustically shorter than voiceless stops (that is, their voice-onset times are shorter). When consonant durations (and ISIs) are measured relative to the onsets of acoustic energy for the consonants, even within the stops, the kind of negative correlation observed in Experiment 3 may be seen, but for reasons having to do with articulatory velocity independently of concerns for rhythmicity.

Very little is known about the relative times to closure of other acoustic segments (but see Kuehn & Moll, 1976). It is possible that this variable, together with that of segmental manner class, can account for all of the observed variation in ISI durations in Experiments 1 and 3. This cannot be determined at present, however.

The foregoing discussion makes plausible the following hypotheses about the talkers' performances in Experiments 1 and 3 and the listeners' performances in Experiment 2:

(1) When a talker is asked to produce a stress-timed utterance, he acquiesces by initiating the *production* of stressed syllables at temporally equidistant intervals. For utterances composed of ISIs with nonidentical syllable-initial consonants, the consequence of *articulatory* isochrony is *acoustic* anisochrony; the deviations from acoustic isochrony can be predicted on the basis of differences in the manner classes of the consonants or in their respective times to closure (minus closure time for the stops). A talker does not try to compensate for the anisochronies in the acoustic signal that these variables introduce.

(2) Comparable differences in the temporal alignments of syllables may be observed in natural speech.

Their occurrence here does not necessarily indicate that talkers are working to produce stress-timed utterances under ordinary conditions of speech production. By hypothesis, the anisochronies arise for reasons relating to the articulatory properties of the syllable-initial consonants themselves, and provide no information concerning strategies of suprasegmental speech production.

(3) When listeners judge the rhythmicity of an utterance, their judgments are based on information about *articulatory* timing; they are not based on judgments of the intervals between acoustic syllable onsets, but rather (on a first approximation) on judgments of the intervals between articulatory syllable onsets. This implies that, in the appropriate context, acoustic silence provides information to a perceiver about articulatory activity, and thus about the occurrence of a particular class of phonetic segment—a conclusion already reached on independent grounds by Liberman and his colleagues. (See Liberman & Pisoni, 1977, for a review of some of this evidence.) More precisely, when an intraphrase syllable begins with a stop, its articulatory and *perceptual* onset may begin at the onset of the silent period or within the silent period that precedes the stop.

Together, the first two hypotheses listed above suggest that the acoustic anisochronies that are observed when talkers intend to produce stress timing are unrelated to any peculiar strategies for producing the prosodics of speech. That is, they do not arise because the talker *intentionally* causes the onsets of acoustic energy for successive stressed syllables to occur when they do. Instead, the anisochronies are a by-product of the talker making articulatory gestures at a stress-timed rate. And they are a by-product due to articulatory properties of individual phonetic segments, not to properties of articulation having to do with speech prosody.

If these proposals are correct, the acoustic anisochronies should show up even when concerns for stress timing or other aspects of speech prosody are irrelevant. In fact, they should be evident when a talker is trying to produce an *isolated syllable* as rapidly as he can. *Ceteris paribus*, a talker's vocal reaction time measured from the onset of an optical signal to produce a CV syllable to the onset of acoustic energy of the syllable, should be shorter for a syllable starting with a fricative or nasal than for one starting with a stop, and should be shorter for a voiceless stop-vowel syllable than for a voiced stop-vowel syllable. Moreover, these differences in vocal reaction time should correlate with measures of the manner class and articulatory velocity of an initial consonant and with the acoustic deviations from isochrony found in Experiments 1 and 3. Experiment 4 tests these predictions.

EXPERIMENT 4

Method

Stimuli and Procedure. A list of 20 CV syllables constituted the stimulus set. The vowel in every case was /a/. The consonants were the voiced stops /b,d,g/, the voiceless stops /p,t,k/, the nasals /m,n/, the fricatives /f,v,s,z,ʃ,θ/, the affricates /č,ǰ/ and the semivowels and -consonants /w,r,l,y/. The latter will be called collectively semivowels. Affricates and semivowels were added to provide additional information on the way in which articulatory variables may influence acoustic syllable—onset time and therefore vocal reaction time. Affricates, like stops, require an acoustic silent period during which the vocal tract is occluded. Semivowels are like fricatives and nasals in this respect; they do not occlude the vocal tract or require an acoustic silent period.

The stimuli to be produced by the subjects were presented visually one at a time on a CRT screen, and their sequencing was controlled by a computer program. Each syllable occurred exactly once in each block of 20 trials; ten blocks were presented in all. The syllables were randomized within each block, and each block presented a different random order. Likewise, different subjects received different randomizations of the stimuli.

On each trial, subjects first heard a warning bell; 495 msec later, a CV syllable (e.g., "BA") appeared on the viewing screen. The screen was covered with opaque paper except for a slit the width of a single line of print. Subjects were instructed to fixate the slit at a location indicated by an arrow drawn on the paper. The CV syllable appeared in this location on each trial. Intertrial intervals were selected randomly from the values 2, 2.5, 3, 3.5, and 4 sec, so that trials were not rhythmically sequenced.

The subjects were instructed to read each syllable aloud as quickly as possible after it had appeared on the screen. Reaction times were obtained in two ways. One way was immediate, but somewhat inaccurate, and served only as feedback for the subject. The second way was substantially more accurate. The first method of obtaining reaction times was via a microphone, a voice relay, and a millisecond counter interfaced to the computer. The reaction times obtained in this way were printed on the viewing screen after each trial. Subjects were instructed to maintain their reaction times below 500 msec, and this method of feedback was intended to facilitate their doing so.

Vocal reaction times obtained in this way are inaccurate because of the different energy levels at which different consonants may be produced. A CV syllable, acoustically defined, that starts out low in intensity (e.g., /s/) may not trigger the voice relay until the vowel onset, while one that starts out high in intensity (e.g., /b/) may trigger the relay at once.

To get around this difficulty, reaction times were also obtained by recording the warning bells and the subjects' responses on audio tape. Sound spectrograms were made of each trial, and reaction times were measured as the time between the onset of the bell and the onset of the acoustic signal for the syllable minus 495 msec.[3] For each subject, the first two blocks of trials were treated as practice trials and were excluded from the analysis of the data. Also excluded were trials on which misarticulations occurred or reaction times were above 600 msec. These averaged 4% across subjects and ranged between 1% and 7%. They will not be discussed further.

Subjects. The subjects were five students in the Introductory Psychology course at Dartmouth College. They received course credit for their participation.

Results

Table 2 gives the mean reaction times and standard deviations for each of the five manner classes of consonant (/w,r,l,y/ are treated together). The affri-

Table 2
Mean Vocal Reaction Times (in Milliseconds) for the Five Manner Classes of Consonant in Experiment 4

	stop	affricate	fricative	nasal	semi-vowel
Mean	349	374	337	323	333
SD	12.6	41.2	15.7	8.7	12.9

cates gave the longest reaction times followed by the stops. These two manner classes were expected to give the longest reaction times, although a difference between them was not expected. An analysis of variance to test the differences among the five means yielded a significant F value [$F(4,16) = 6.67$, $p = .002$]. However Scheffé's tests showed that the significance was due only to the difference between the affricate class and each of the other four groups and that between the stops and the nasals. The differences among the other classes, including most notably between the stops and the fricatives and between the stops and the semivowels, are not significant. Nonetheless, the rank ordering of the differences among the means is in the predicted direction, showing that consonants that require an initial period of silence (the affricates and the stops) have longer vocal reaction times than those which do not (nasals, fricatives, semivowels, and consonants). A Scheffé test comparing the affricates and stops against the other three manner classes of consonant yields a significant outcome [$F(4,16) = 4.97$, $p < .01$].

The results for the voiced and voiceless stops were also as expected based on the variables of time to closure and closure duration. Table 3 presents the reaction time means and standard deviations for these classes of stop each partitioned into the three places of articulation. Across all three places of articulation, the voiced stops gave longer reaction times than the voiceless stops. Moreover, except for the /d/-/t/ comparison, this direction of effect was perfectly consistent across subjects. Thus, all five subjects showed faster reaction times for /p/ than for /b/, and all showed faster times for /k/ than for /g/. Three of the five subjects had faster times on /t/ than on /d/. A two-way repeated measures analysis of variance (voicing by place of articulation) revealed a significant effect of voicing [$F(1,4) = 8.86$, $p < .05$], but no effect of place and no interaction.

The duration of the prevocalic acoustic signal of these syllables yields a crude, continuous measure of consonant manner class. This measure correlates negatively ($r = -.56$, $p < .05$) with vocal reaction time, indicating that long reaction times tend to occur for CV syllables whose consonant phonemes are short in acoustic duration, while short reaction times occur for long-duration consonants. (This correlation measure was computed on 16 syllables excluding the semivowels and consonants. For these latter syllables, the measure of prevocalic duration was difficult to make reliably.) The correlation parallels the outcomes of Experiments 1 and 3 to the effect that long ISIs precede short-duration consonants and short ISIs precede long-duration consonants.

When the correlation is restricted to the consonants b, m, n, t, f, and s (the phoneme set of Experiment 3), the r value is $-.72$ ($p < .05$). This value is somewhat less than the values $-.96$ and $-.75$ observed for the subjects of Experiment 3, but it suggests a relationship between vocal reaction time and consonant manner class that is of the same sort as that between ISI and consonant manner class as observed in Experiment 3.

Finally, the latency *differences* among the possible pairings of b, t, m, n, and s (the consonant phonemes of Experiment 1) in the present experiment are highly correlated with the deviations from isochrony of the corresponding alternating utterances of Experiment 1 ($r = .86$, $p < .01$).

Discussion

It seems quite likely in this experiment that the time to *start* producing a response was fairly constant across the different CV syllables. [Nonetheless, there is a nearly significant correlation of $-.44$ between vocal reaction time and phoneme frequency as measured by the tables of Dewey (1970).] Therefore, differences in reaction time have primarily to do with any differences in the manner class or the articulatory characteristics of a consonant that affect the time at which the segment has some acoustic consequence other than silence. Since variability due to these factors accounts for 74% of the variance obtained in Experiment 1, it seems likely that these factors are the essential ones that give the P-center phenomenon its character.

In themselves, the observations of Experiments 1, 3, and 4 are not surprising once they have been made. They indicate only that talkers generate systematic temporal alignments of syllables when they talk by adopting very simple and obvious articulatory strategies (i.e., stress timing in Experiment 1 as instructed, either stress timing or, even more simply, producing a syllable immediately on completing the preceding one in Experiment 3; producing a syllable as quickly as possible in Experiment 4, as instructed). It happens

Table 3
Mean Vocal Reaction Times (in Milliseconds) for Voiced and Voiceless Stops in Experiment 4

	Voiced		Voiceless	
	Mean	SD	Mean	SD
bilabial	364	15	338	15
alveolar	351	21	338	27
velar	360	13	344	13

that these strategies have complex, though systematic, acoustic consequences.

The findings of these three experiments are of interest primarily in conjunction with the observations of Morton et al. (1976) and their verification in Experiment 2. Together, the set of findings indicate that listeners track acoustic information specifying *articulatory* strategy when they perceive an utterance; they do not treat the acoustic signal as a signal that is independent of its vocal-tract source.

The final experiment in this series reexamines the P-center or stress beat itself, and seeks to establish a relationship between it and the underlying articulation of a syllable.

EXPERIMENT 5

Acoustically defined, the stress beat precedes the onset of a stressed vowel by an amount of time that increases with the duration of the prevocalic consonant. This acoustic description is compatible with the data of Allen (1972), Morton et al.(1976) and Rapp (Note 1) as well as with that of Experiments 1-3.

A simple, but speculative articulatory proposal that also fits these data is that the stress beat or P-center is time-locked to (but not necessarily identical to) the onset of articulatory activity for the syllable. (This proposal will be refined in the General Discussion.) On this view, the correlations found by Rapp and Allen, and the acoustic anisochronies reported by Morton et al., occur because different manner classes of consonants tend to have acoustic consequences at variable lags with respect to their articulatory onsets.

These two descriptions of P-center locus, the one a source- or articulation-free acoustic description, and the other articulation-based, are indistinguishable in most utterances. However, they can be distinguished by careful selection of prevocalic consonants. In the following experiment, utterances are selected that allow the two descriptions to be tested separately.

Here, as in Experiment 1, sets of homogeneous and alternating utterances were constructed. Homogeneous utterances consisted of the same CV syllable repeated rhythmically six times. Alternating utterances consisted of two CV syllables that were alike in place of articulation, but differed in respect to voicing lead. These syllable types were repeated in alternation three times each. In both utterances types, the component syllables rhymed with /ad/ and the initial syllables were /b,d,g/ either prevoiced or voiced (no voicing lead). Thus, there were three homogeneous utterances in which no syllable was prevoiced, three in which all were prevoiced, and three alternating utterances in which prevoicing and voicing alternated.

Acoustically, prevoiced consonants differ from their voiced counterparts in having a substantial duration of acoustic energy due to voicing preceding the stop-consonant release. In the voiced stops, the release is preceded by silence.

In contrast to this, articulatorily, the two segments are alike except that the vocal cords vibrate during vocal-tract closure in the first case, but not in the second. That is (presumably), the onsets of vocal-tract closure for the two classes of stop occur at the same or nearly the same temporal distances from the stop releases. In the prevoiced consonant, closure is accompanied by voicing; in the voiced consonant, it is not.

The initial, acoustic description of the P-center or stress beat suggests that the P-center of a syllable with a prevoiced consonant should precede its vowel onset by a longer duration than the P-center of a syllable with a voiced consonant because the prevoiced CV syllable has the longer prevocalic acoustic duration. More particularly, in Rapp's data, the stress beat nearly coicided with the release of the voiced stop /d/. If this represents the general case for voiced stops, then the stress beat for voiced /b,d,g/ will coincide, or nearly coincide, with the stop release. That for the prevoiced stops will precede it.

In contrast, if P-centers are time-locked to the *articulatory* onsets of CV syllables, regardless of the consonants' voicing class, they should be located at the same relative temporal distances from any acoustic markers that the two syllable-types share, including the stop releases.

In summary, then, the acoustic description of the P-center leads to the prediction that intervals in the homogeneous utterances, measured from stop release to stop release, should be isochronous, because the P-centers of their component syllables all have the same locations relative to these acoustic markers (or any others). But alternating utterances should be anisochronous on this measure because the P-center of the voiced stop will coincide with the stop release, while that of the prevoiced stop will precede it. The articulatory description suggests that both sets of utterance types will be isochronous when intervals between stop releases are measured.

Method

The nine utterances, six homogeneous and three alternating, were presented in a typed list in random order to two subjects. Both subjects were native speakers of English and both were phoneticians.[4] (Prevoiced stops are not distinct phonemes from the voiced stops in English. Therefore, untrained speakers would find it difficult to produce the voicing difference systematically.) The subjects were asked to read through the list twice, producing each utterance at a slow stress-timed rate. Their utterances were recorded on audio tape in a soundproofed booth.

Sound spectrograms were made of each utterance, and two types of measurements were made on each one: (1) the intervals between the acoustic-syllable onsets of Syllables 2 and 3, 3 and 4, and 4 and 5 (ISIs 2, 3, and 4 of Experiment 1; (2) the intervals between the stop bursts of Syllables 2 and 3, 3 and 4, and 4 and 5. The first set of intervals should be isochronous only for the six homogeneous utterances if the usual P-center results are obtained. Among the alternating utterances, intervals between a prevoiced and a voiced stop (ISI_3) should be longer than those between a voiced and a prevoiced stop (ISI_2 and ISI_4). The second set of intervals, according to the acoustic description of the P-center, should be isochronous only for the homogeneous utterances. The alternating utterances should be anisochronous in the same way as those of Measure 1, although the anisochrony should be less pronounced. According to the articulatory proposal, both utterance types should be isochronous.

Results

Table 4 presents the outcome when acoustic syllable onsets delimit the relevant intervals. On this measure, for both subjects, homogeneous utterances are isochronous and alternating utterances are not. An analysis of variance with ISI and utterance type (voiced, prevoiced, and alternating) as repeated measures factors showed both main effects to be significant [$F(2,2) = 76.97$, $p = .01$, and $F(2,2) = 21.08$, $p < .05$, respectively] as was their interaction [$F(4,4) = 47.36$, $p = .003$]. Scheffé's tests on the individual means attribute the significant interaction to differences between the alternating utterances, on the one hand, and the prevoiced and voiced utterances, on the other. In particular, for voiced and prevoiced homogeneous utterances, the three ISIs do not differ significantly. In contrast, for the alternating utterances, ISI_3 is longer than ISI_2 and ISI_4 [$F(2,4) = 135.5$, $p = .008$]; ISI_2 and ISI_4 do not differ one from the other. In addition, ISI_2 and ISI_4 of the alternating utterances are shorter than their prevoiced and voiced counterparts [$F(5,4) = 37.89$, $p = .003$], while ISI_3 in the alternating utterance is longer than its homogeneous counterparts [$F(2,4) = 24.83$, $p = .007$]. In short, the alternating utterances deviate from isochrony in the predicted way, in that intervals starting with prevoiced stops and ending with voiced stops are long relative to intervals that are the reverse of this.

Table 5 presents the comparable values on the burst-to-burst measure. On this measure, all utterance sets for both subjects are isochronous. (Although the alternating utterances show a slight cyclicity in ISI duration, it is contrary to the predicted direction and in any event is nonsignificant.) The analysis of variance yielded nonsignificant outcomes on both main effects and on the interaction term.

Table 4
Durations (in Milliseconds) of Interstress Intervals of Homogeneous and Alternating Utterances in Experiment 5 Measured Onset to Onset

	ISI_2		ISI_3		ISI_4	
	AA	LL	AA	LL	AA	LL
prevoiced	700	698	704	686	696	684
voiced	689	645	676	654	698	665
alternating	576	563	797	731	580	555

Note—Each value is the average of six tokens.

Table 5
Durations (in Milliseconds) of Interstress Intervals of Homogeneous and Alternating Utterances in Experiment 5 Measured Burst to Burst

	ISI_2		ISI_3		ISI_4	
	AA	LL	AA	LL	AA	LL
prevoiced	698	696	692	670	726	690
voiced	689	645	676	654	698	665
alternating	700	660	690	639	706	649

Note—Each value is the average of six tokens.

Discussion

The experimental outcome supports the articulation-based description of the P-center and fails to support an articulation-free acoustic description. It should be noted, of course, that the experimental design stacked the deck somewhat in favor of the articulatory description. Its prediction was that on the critical burst-to-burst measure, the null hypothesis would fail to be rejected. This is a weak prediction.

Two kinds of consideration support the argument that there are *truly* no differences among levels of the independent variable ISI (rather than that, there are differences to which the experimental design is insensitive.)

First, it is clear that the data are not very noisy. The average deviation from isochrony for homogeneous utterances measured from acoustic syllable onset to acoustic syllable onset averages 12 msec ($/ISI_2$-$ISI_3/$ and $/ISI_3$-$ISI_4/$ averaged) for one subject and 9 msec for the other. Moreover, the patterning of these data on this measure is similar to that of Experiment 1 and of the other P-center-related studies.

Second, on the burst-to-burst measure, the prevoiced to voiced interval does not even tend to be long. Instead, it is slightly shorter than the other intervals, but not reliably so. (This may signify that, like voiceless and voiced stops, prevoiced and voiced stops differ somewhat in time to closure.)

We conclude, *tentatively,* therefore, that the P-center is time-locked to the articulatory onset of the prevocalic consonant in a monosyllabic stressed utterance.

GENERAL DISCUSSION

This final discussion will consider three questions: (1) What is a P-center (and, closely related to this, what does it correspond to in a speech event)? (2) How do listeners track P-centers when they make rhythmicity judgments and on other occasions (as in the phoneme-targeting experiments)? (3) What implications does the P-center phenomenon have, if any, for devising and evaluating a plausible proposal about natural suprasegmental speech rhythms, stress timing in particular?

What is a P-center?

Curiously, P-centers are phenomena that talkers can regulate and listeners can track, but which investigators are unable to *see* in any optical representation of an utterance (see Marcus, Note 5).

Were it not for the Allen and Rapp studies, the P-center might be supposed to correspond to the articulatory onset of a stressed syllable. This would have different acoustic correlates, depending on what type of gesture is initiated. However, the Rapp and Allen studies place the stress beat very often *within* the acoustic realization of the stressed syllable's prevocalic acoustic signal, and these mutually reinforcing results require an accounting. A speculative account is suggested here.

The P-center may correspond to articulatory activity relating to the stressed *vowel* itself (perhaps to its articulatory onset, or to the attainment of the nearest approximation to its "target" vocal-tract shape). Anticipatory coarticulation in CV syllables is very well documented. Due to coarticulation, the articulatory vowel onset would tend to occur *during* the production of a preceding consonant. Gay (1977) reports that the initiation of movement towards V_2 in a V_1CV_2 utterance coincides with, or occurs just after, the attainment of consonantal closure. The consonants in his study were the voiceless stops. This would locate the P-center no earlier than the onset of the consonant's silent period if it corresponds to the vowel's articulatory onset or somewhat thereafter if it corresponds to the vowel's nearest approximation to its target. Rapp's Figure I-B-6 includes one voiceless stop, t. The stress beat in /atád/ was located somewhat after t's release, suggesting that the target view of the P-center locus, suggested above, may be the more accurate of the two proposals.

Interestingly, Carney and Moll (1971) report that, for hV_1CV_2 nonsense utterances in which the consonant is a fricative rather than a stop, movement toward the second vowel begins *before* the attainment of consonant near-closure for the fricative. This difference between the studies of Gay and Carney and Moll is consistent with the evidence (Kuehn & Moll, 1976; also see MacNeilage & Ladefoged, 1976) that, in general, fricatives have slower articulatory velocities than stops. If consonants and their following vowels have articulatory *onsets* at some constant temporal delay with respect to each other regardless of the manner class of the consonant, then articulatory reference points for the vowel should appear sooner in a fricative's production than in the more rapid production of a stop. This would place a vowel-related P-center relatively earlier in a fricative than in a stop. It is difficult to say whether this is the case articulatorily. However, it seems generally to be compatible with the acoustic data. In Allen's acoustic measurements, taps to syllables preceded by voiced stops were located closest to the acoustic vowel onset (and thus to the consonant offset). Next were voiceless and voiced fricatives. These data are compatible with the foregoing considerations. However, farthest away of all were the voiceless stops, which should have the highest articulatory velocity going into the closure period. Rapp's data offer only limited relevant information, but it is similar to Allen's. Among the stops and fricatives that she examined, the stress beat is closest to the vowel onset for /d/; it is farther away for /t/ and /s/, which are quite similar to each other. /t/'s stress beat precedes that for /s/ slightly.

It is also interesting in regard to this vowel-based proposal that, in Rapp's data, talkers located the stress beat at a nearly invariant interval *after* the first (unstressed) vowel's acoustic onset in the various /a$C_n$ád/ utterances, even though the beat occurred at a variable locus relative to acoustic markers for the consonants and second, stressed vowel.

How Do Listeners Track P-centers?

Morton et al. failed to uncover an acoustic marker for the P-center. Of course, this failure does not imply that the P-center has no acoustic marker. But, coupled with the present findings, it suggests a need to reconsider the kind of acoustic marker that one can reasonably expect to find. If, indeed, the P-centers are acoustic correlates of some *abstract* gesture-*type* (for example, the onset of articulatory activity for a vowel, any vowel), then the acoustic correlates may not be invariant in any superficial sense. There is no reason to expect the P-center to correspond, for instance, to an intensity peak or to an abrupt change in fundamental frequency if these are not acoustic correlates of its underlying gesture type. Instead, the acoustic correlates of the P-center may be that (very large) *class* of acoustic signals that, in the appropriate context, *signify* (for example) "the onset of articulatory activity for a vowel." The problem is to explain what about these signals, other than some simple shared acoustic property, endows them with that significance for a listener.

Even if simple acoustic invariants are excluded as plausible markers of P-centers, complex acoustic correlates probably are not. Vowel onsets (and asymptotic attainments of "target" vocal-tract shapes) may well be marked by their acoustic coarticulatory influences on consonants. Due to coarticulation, consonants vary acoustically with their segmental context. This means that (among other segments) a postconsonantal vowel's "anticipatory" articulation during the consonant is marked *by* its effect on the acoustic signal for the consonant. Its articulatory onset, then, may be marked by the initiation of its effect on the consonant. Its attainment of a target shape of the vocal tract may be signaled by a stabilization of the vowel's effect on the acoustic signal for the consonant. Possibly this information marks the P-center for a listener.

Suprasegmental Speech Rhythms

A convenient aspect of the proposal that P-centers correspond to articulatory *onsets* is that it enables an intuitive view of speech timing to be preserved—namely that timed events begin and end at the edges of linguistic units. If, instead, P-centers correspond to some within-segment locus—e.g., to vowel-target near-attainments or to some variable locus within a prevocalic consonant—and if intervals between P-centers are regulated in speaking (as they clearly *can* be)—current models of speech production based on linguistic segmental and suprasegmental timing units are disconfirmed. Moreover, devising a new model would entail a major overhaul in our views of speech production. Nonetheless, the data do seem to support the articulatory targets view more than the vowel onsets view of P-center locus.

In respect to stress-timing itself, as noted in the introductory section, the P-center findings suggest a need to reevaluate the tests of the claims that speech is stress-timed. One necessary adjustment to the procedures cited earlier (among other adjustments; see Fowler, Note 6) is to measure the intervals between P-centers rather than those between acoustic onsets of stressed syllables. But given the difficulty of locating P-centers precisely, and more importantly, given that the stress-timing proposals are not properly tested by measuring interstress intervals and looking for significant differences (see Footnote 1), other kinds of tests that avoid ISI measurements should be adopted (e.g., Fowler, Note 6; Allen, Note 8).

REFERENCE NOTES

1. Rapp, K. A study of syllable timing. *Papers from the Institute of Linguistics*, University of Stockholm, 1971, **8**, 14-19.
2. Lea, W. *Prosodic aids to speech recognition: IV. A general strategy for prosodically guided speech understanding.* (Advance Research Projects Agency Report No. PX10791), March 1974.
3. Shen, Y., & Peterson, G. Isochronism in English. *Studies in Linguistics, Occasional Papers*, University of Buffalo, 1962, *9*.
4. Lindblom, B., & Rapp, K. Some temporal regularities of spoken Swedish. *Papers from the Institute of Linguistics*, University of Stockholm, 1973, **21**, 1-59.
5. Marcus, S. *Perceptual centers*. Unpublished fellowship dissertation, Kings College, Cambridge, 1975.
6. Fowler, C. *Timing control in speech production*. Bloomington, Indiana: Indiana University Linguistics Club, 1977.
7. Port, R. *The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words.* Bloomington, Indiana: Indiana University Linguistics Club, 1977.
8. Allen, G. The place of rhythm in a theory of language. *UCLA Working Papers*, 1968, **10**, 62-84.

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, I. *Language and Speech*, 1972, **15**, 72-100.

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

Coleman, C. *A study of acoustical and perceptual attributes of isochrony in spoken English*. Unpublished doctoral dissertation, University of Washington, 1974.

Cutler, A. Rhythmic factors in the determination of perceived stress. *Journal of the Acoustical Society of America*, 1975, **57**, S25. (Abstract)

Dewey, G. *Relative frequency of English spellings*. New York: Teachers College Press, 1970.

Duckworth, J. *An inquiry into the validity of the isochronic hypothesis*. Unpublished doctoral dissertation, University of Connecticut, 1965.

Gay, T. Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America*, 1977, **62**, 183-193.

Klatt, D. The linguistic uses of segment duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 1976, **59**, 1208-1221.

Kozhevnikov, V. A., & Chistovich, L. A. *Speech: Articulation and perception*. Moscow-Leningrad, 1965. (English translation: J.P.R.S., Washington, D.C., No. JPRS 30543.)

Kuehn, D., & Moll, K. A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*, 1976, **4**, 303-320.

Lehiste, I. Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America*, 1973, **51**, 2018-2024.

Liberman, A., & Pisoni, D. Evidence for a special speech-perceiving subsystem in the human. In T. Bullock (Ed.), *Recognition of complex acoustic signals*. Berlin: Dahlem Konferenzen, 1977.

Lisker, L. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, 1957, **33**, 42-49.

MacNeilage, P., & Ladefoged, P. The production of speech and language. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 7). New York: Academic Press, 1976.

Martin, J. Rhythm-induced judgments of word stress in sentences. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 627-633.

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

Oller, D. The effects of position in utterance on speech segment-duration in English. *Journal of the Acoustical Society of America*, 1973, **54**, 1235-1246.

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

Shields, J., McHugh, A., & Martin, J. Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, 1974, **102**, 250-255.

Trager, G. L., & Smith, H. L. *Outline of English structure. Studies in linguistics* (No. 3). Norman, Oklahoma: Battenburg, 1951.

NOTES

1. For brevity, an inaccuracy has been introduced into the discussion. To my knowledge, in every instance in which a

stress-timing proposal has been made, it has taken a very weak form, to the effect that the intervals between stressed-syllable onsets *tend* towards isochrony. They are held to deviate from strict isochrony to a degree that varies with the compositional (phonological, syllabic, grammatical) heterogeneity of its component interstress intervals. However, although the stress-timing claims have invariably taken this weakened form, the experimental tests cited above are uniformly of the stronger (and easier to test) claim that interstress intervals are isochronous. The literature suggests that a fair test of the weak form of the hypothesis yields more hospitable data (see Fowler, Note 5, for a review).

2. The extant evidence makes it clear that this adjustment in measurement strategy would not reduce the variability in interstress interval duration to zero. Some more of the variability in the studies of Duckworth (1965) and Shen and Peterson (Note 3) may be put down to an infelicitous assignment of major stresses to syllables. Both adopted the conservative view of Trager and Smith (1951), to the effect that only one major stress may occur between two terminal junctures in a sentence. Nonetheless, substantial variability in interval durations was also reported by Lehiste (1973) and by Lea (Note 2), not all of which would disappear were the intervals between P-centers measured. However, it may be argued that the remaining variability is compatible with the weak version of the stress-timing view (see Fowler, Note 6).

3. I thank Don Nemcek for making the spectrograms.

4. I thank Arthur Abramson and Leigh Lisker for serving as subjects in this experiment.

Appendix A
Absolute Values of Durational Differences (in Milliseconds) Among the Interstress Intervals of Homogeneous and Alternating Utterances in Experiment 1

Initial Segment	$\lvert ISI_2\text{-}ISI_3 \rvert$	$\lvert ISI_3\text{-}ISI_4 \rvert$	$\lvert ISI_2\text{-}ISI_4 \rvert$
		Homogeneous	
a	37	37	0
b	18	23	5
m	0	15	15
s	22	23	45
t	23	7	16
n	23	60	37
		Alternating	
b,a	7	67	60
m,a	150	150	0
s,a	204	157	38
t,a	60	105	45
n,a	179	172	7
b,s	278	285	7
m,s	60	45	15
t,s	150	165	15
n,s	37	52	15
b,t	60	68	8
n,t	45	30	15
t,m	135	135	0
b,n	157	120	37
m,n	22	52	30
m,b	195	195	0

Appendix B
Mean Vocal Reaction Times (in Milliseconds) for the 20 Consonants in Experiment 4

Consonant	Mean	SD
b	364	9
d	351	15
g	360	22
p	338	34
t	338	45
k	344	48
v	349	37
z	343	47
f	332	46
s	313	100
ʃ	310	100
Θ	377	36
j	374	42
c	373	67
m	332	43
n	314	42
w	343	8
r	338	16
L	320	22
y	332	14

(Received for publication October 2, 1978; revision accepted January 17, 1979.)