

Some articulatory correlates of perceptual isochrony

BETTY TULLER and CAROL A. FOWLER
Haskins Laboratories, New Haven, Connecticut 06510

Given sequences of digits with temporally equidistant acoustic onsets, listeners do not perceive them as isochronous (Morton, Marcus, & Frankish, 1976). In order for the sequences to be perceptually isochronous, systematic departures from acoustic isochrony must be introduced. These acoustic departures are precisely those that talkers generate when asked to produce an isochronous sequence (Fowler, 1979), suggesting that listeners judge isochrony based on acoustic information about articulatory timing. The present experiment was an attempt to test directly whether perceptually isochronous sequences have isochronous articulatory correlates. Electromyographic potentials were recorded from the orbicularis oris muscle when speakers produced sequences of monosyllables "as if speaking in time to a metronome." Sequences were devised so that lip-muscle activity was related to the syllable-initial consonant, the stressed vowel, or the stressed vowel and final consonant. Results indicate that isochronous muscular activity accompanies both isochronous and anisochronous acoustic signals produced under instructions to generate isochronous sequences. These results support an interpretation of the perceptual phenomenon reported by Morton et al. to the effect that listeners judge isochrony of the talker's articulations as they are reflected in the acoustic signal.

Morton, Marcus, and Frankish (1976) report that sequences of digits presented with temporally equidistant acoustic onsets are not perceptually isochronous. Moreover, when listeners are allowed to adjust the intervals between successive digits, they introduce systematic departures from acoustic isochrony before judging the sequence of digits to be isochronous. These departures are such that the temporal alignment of a word relative to its neighbors varies with the duration of acoustic energy prior to the acoustically defined onset of its vowel.¹ Thus, a word pair such as "eight-six," repeated in alternation, tends to be arranged in time so that the acoustic onset-to-onset time is less for "eight-six" than for "six-eight."

Based on these findings, Morton et al. proposed that listeners judge the timing of the word sequences based on reference points within each word. Their data suggest that the longer the acoustic duration of the initial consonant, the longer the interval between the acoustic onset of the word and the location of its reference point. Morton et al. call the reference point the "P-center" of a word, and describe it as the word's "psychological moment of occurrence."

The technique of allowing a listener to adjust the intervals between words only establishes the relative

temporal alignments of successive words in a perceptually isochronous sequence; it does not establish the location of the P-center within each word. Moreover, further investigation by Morton et al. did not reveal any acoustic marker of a P-center. Specifically ruled out by their investigations are the stressed vowel's acoustic onset, the word's onset, and the vowel's peak intensity.

Two investigations (Allen, 1972; Rapp, Note 1) apparently do locate the P-center in a word, but both fail to discover how it is marked acoustically. Allen's subjects tapped "on the beat" of a designated syllable in a sentence, whereas Rapp's subjects repeated nonsense words in time with a regularly occurring pulse. Both studies revealed the same relationship between pulse or tap and the onset of a stressed vowel. The pulse or tap was located near the acoustically defined onset of the stressed vowel, but preceded it by a duration that correlated positively with the duration of the prevocalic consonant or consonant cluster.

In Rapp's data, it is evident that as the acoustic duration of the prevocalic consonant(s) increased, the beat tended both to precede the acoustic onset of the vowel by longer intervals and to follow the acoustic onset of the syllable by longer intervals. If stress beats are P-centers, and if P-centers occur at isochronous intervals in perceptually isochronous sequences, then Rapp's data predict the acoustic anisochronies reported by Morton et al. That is, syllables with acoustically long prevocalic consonants

The research reported here was supported by the following grants to Haskins Laboratories: NINCDS Grants NS-13617 and NS-13870 and BRSG Grant RR-05596. B. Tuller is also at University of Connecticut, Storrs, Connecticut, 06268; C. A. Fowler is also at Dartmouth College, Hanover, New Hampshire, 03755.

should be located temporally closer to their predecessor than syllables with acoustically short prevocalic consonants.

The three investigations by Morton et al., by Allen, and by Rapp concur in finding that departures from acoustic isochrony are systematic, and that the psychological moment of occurrence, P-center, or stress beat of a word varies in locus with initial consonant duration, but does not correspond to any obvious acoustic marker.

Based on a series of experiments Fowler (1979) proposed an articulatory rationale for the P-center phenomenon. She found that when talkers are asked to produce an isochronous sequence of stressed monosyllables, they generate an acoustic signal in which the anisochronies are just those that listeners require in order to hear a sequence as isochronous. In itself, this may mean only that talkers and listeners agree in mistaking anisochrony for isochrony. However, an additional study suggests a different interpretation of the findings, to the effect that the acoustic anisochronies arise as a byproduct of a fairly straightforward strategy for producing isochronous sequences. The study obtained vocal reaction times (as measured from spectrographic displays) for reading visually presented CV syllables. Reading reaction times correlated negatively with the measured acoustic duration of the initial consonant ($r = -.56$); that is, reaction time was short to syllables with long initial consonants. This is comparable to the finding by Morton et al. that digits with acoustically long initial consonants follow their predecessors more closely in time than digits with short-duration initial consonants. Moreover, differences in reading times between pairs of CVs differing in initial consonant correlated highly with the departures from isochrony of syllables differing in the same consonants observed in the first study ($r = .86$).

It seemed likely in the reaction time study that talkers initiated the production of each CV at nearly equivalent intervals following the visual presentation of the CV to be produced. Therefore, the differences in vocal reaction time among the CVs are probably due, for the most part, to the different times after articulatory onset that different consonants have acoustic consequences other than silence. These differences in reaction time correlate highly with departures from acoustic isochrony in perceptually isochronous utterances; hence, it seems likely that these latter departures, too, are products of isochronous articulatory gestures.

A plausible explanation of the talker's behavior when producing sequences of uniformly stressed isochronous monosyllables, then, is that he initiates some aspect of the production of each stressed syllable at isochronous intervals. When initial consonants of successive syllables differ, particularly in manner class, the resulting acoustic signal may be anisochronous. So, for example, in the production

of the stop consonant /b/, bilabial closure is maintained to build intraoral pressure for an explosive release. The acoustic correlate of this occlusion typically is silence. In contrast, in the production of /f/, the vocal tract is obstructed by the lip/teeth approximation, but it is not occluded entirely. Therefore, a period of silence need not precede the acoustic onset of an /f/ (or other fricative). Thus, even if lip gestures were initiated at temporally equidistant intervals in a sequence of syllables alternately starting with /b/ and /f/, the acoustic product would be anisochronous.

If psychological isochrony is indeed based on isochrony of articulatory gestures, then we may ask to what articulatory marker the P-center corresponds. Most obviously, it may correspond to a syllable's articulatory onset. However, this account probably fails to explain the data of Rapp and of Allen, cited earlier, which locate the stress beat within the acoustic realization of a consonant. Their findings may signify that articulatory gestures within the syllable, perhaps those associated with the vowel, are the gestures that a talker controls and that a perceiver listens for. Indeed, due to coarticulation among adjacent segments, the gesture for a vowel begins during the production of a prevocalic segment (see, e.g., Carney & Moll, 1971; Gay, 1977). Hence, a proposal that stressed vocalic segments are the timed articulatory gestures underlying the P-center or stress-beat effect is not implausible so long as the controlled vocalic gestures are time-locked to those of the consonant.

The present experiment has as its major aim to test directly the proposal that psychologically isochronous sequences have isochronous articulatory correlates, and in doing so, to uncover the articulatory correlate itself. To these ends, we examine muscle activity at the lips during the production of uniformly stressed isochronous sequences of monosyllables, and compare the relative degrees of isochrony of the muscle activity and of the consequent acoustic signal. In some utterances, the lip activity is associated with the production of the initial consonants in the sequence; in other utterances, it is associated with the production of the stressed vowel or the vowel and final consonant. Utterances are selected based on earlier studies so that some are expected to be acoustically isochronous, while other tokens are expected to show marked departures from acoustic isochrony. Our concern is to discover whether any part of the production of psychologically isochronous utterances is isochronous even when the acoustic product departs from isochrony.

METHOD

Subjects

The subjects were four adult females and one adult male. All were native speakers of English; four were naive to the purposes of the experiment.

Stimuli

Two kinds of CVC nonsense-syllable sequences were devised. Alternating sequences included two stressed monosyllables produced in alternation approximately 20 times each. Homogeneous sequences included a single stressed monosyllable repeated about 40 times.

Alternating sequences were devised in accordance with two criteria. First, based on earlier studies, the spoken sequences were expected to be acoustically anisochronous; second, both component syllables of a sequence involve the orbicularis oris muscle for articulation either of the initial consonant or of the stressed vowel. [The orbicularis oris acts to close or protrude the lips, and is associated with the production of labial consonants and lip rounded vowels (e.g., Bell-Berti & Harris, 1979).] Three alternating sequences were selected:

(1) /bak fak bak fak . . . /—In this sequence, the orbicularis oris is associated with the gestures for /b/ and /f/. Because /f/ is a fricative and /b/ is a stop, the acoustic departures from isochrony should be such that the interval from /bak/ to /fak/ is less than that between /fak/ and /bak/.

(2) /duk suk duk suk . . . /—In this utterance, orbicularis oris activity is associated with lip rounding for the vowel /u/. Acoustically, the /duk/-/suk/ interval should be shorter than the /suk/-/duk/ interval.

(3) /dup sup dup sup . . . /—In this utterance, orbicularis oris activity is associated with the vowel and the final consonant. Acoustically, the /dup/-/sup/ interval should be shorter than the interval between /sup/ and /dup/.

Homogeneous utterances were selected to match the alternating sequences. They were /bak bak . . . /, /duk duk . . . /, /suk suk . . . /, /dup dup . . . /, /sup sup . . . /. In initial acoustic measurements of /bak fak . . . /, the low amplitude of /f/-frication made it very difficult to obtain accurate measurements of acoustic onset; hence, the homogeneous sequence /fak fak . . . / was not used in the experiment. However, /fak/ was used in the alternating utterances to provide a labial fricative of long acoustic duration to contrast to the stop, /b/.

Measures

Electromyographic (EMG) activity was recorded from the orbicularis oris muscle using paint-on surface electrodes (Allen & Lubker, 1972) spaced at approximately ½ cm on the vermilion border of the lower lip. The EMG potentials and acoustic recordings of the subjects' utterances were recorded onto separate channels of a multichannel tape.

Procedure

The subjects were asked to produce six sequences of nonsense syllables, to stress every syllable equally, and to produce the syllables at isochronous intervals—that is, as if speaking in time to a metronome. Each subject produced all three alternating sequences and three matching homogeneous sequences (/bak bak . . . /, /duk duk . . . / or /suk suk . . . /, and /dup dup . . . / or /sup sup . . . /). The order of the syllables in alternating sequences was varied across subjects, some subjects beginning the sequences with the stop- and some with the fricative-initial syllable. Although subjects were not instructed to relax the rounding at the lips between syllables in sequences containing the vowel /u/, activity in orbicularis oris did return to baseline between syllables.

Data Analysis

Each homogeneous and alternating sequence was considered to contain two interval types, odd and even. Intervals of both types extended from the acoustic onset of the final consonant release of one syllable (the "left margin" of the interval) to the acoustic or EMG onset of the immediately following syllable (the "right margin"; see Figure 1). Among the alternating sequences, odd and even intervals differ in respect to their left and right margins. For example, in /bak fak bak fak . . . /, odd intervals begin at the release of /k/ in /bak/ and end at the acoustic or electromyographic onset of /f/ in /fak/. Even intervals begin at the release of /k/ in /fak/ and end at the onset of /b/ in /bak/. Homogeneous utterances likewise were partitioned into odd and even intervals, but now the intervals were identical in respect to their left and right margins.

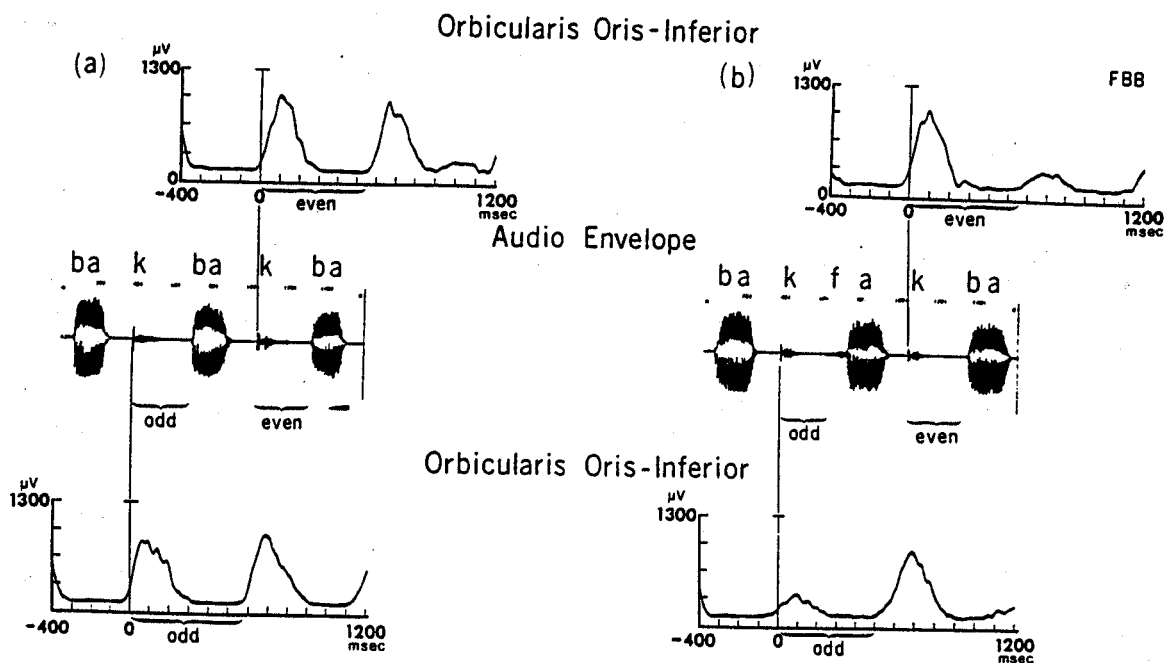


Figure 1. "Odd" and "even" intervals in acoustic and EMG records for both (a) homogeneous and (b) alternating sequences.

In the case of two subjects, sequences ending in /p/ were not released reliably. For these utterances, the offset of periodicity in the preceding vowel was used as the left margin of odd and even intervals.

To avoid contamination of the isochrony measurements by utterance-initial and -final lengthening effects (Klatt, 1976; Oller, 1973; Lindblom & Rapp, Note 2), the first two syllables and the last two syllables of all sequences were excluded from the analyses. The number of tokens of each interval type, odd and even, ranged from 15 to 20 for each subject.

Acoustic measurements were made from sound spectrograms. The EMG data were rectified, software integrated with a time constant of 35 msec and averaged using the Haskins Laboratories EMG system described by Kewley-Port (Note 3, Note 4). The left margin of each interval, as defined above, was used as the point around which tokens of each interval type were averaged.

The averaging program provides a listing of the mean amplitude of the EMG signal in microvolts during successive 5-msec intervals throughout the averaged interval. Using this listing, the onset of EMG activity in the orbicularis oris following consonant release was defined as the point in time at which 5% of the amplitude range between baseline and peak activity on the averaged signal had been traversed. Typically, no reversals of the EMG signal occurred beyond that point.

We emphasize that the EMG onsets were taken from signals averaged across all of the odd or even intervals within an utterance as produced by a single talker. That is, they were not taken from individual odd or even intervals. Although onsets are easily identified from averaged signals, they cannot be identified reliably from raw signals (Baer, Bell-Berti, & Tuller, 1979). This procedure of identifying onsets from an averaged signal has the disadvantage that, in contrast to our acoustic measures, variances cannot be obtained on individual utterances spoken by individual talkers. Instead, they must be obtained across utterances or, as we have done, across talkers.

RESULTS

For both acoustic and EMG measurements, difference scores were calculated between the means of the two interval types. (The difference scores for /suk suk.../ and for /duk duk.../ did not differ from each other and therefore were combined across subjects, as were those for /sup sup.../ and /dup dup.../.) Let us first consider the acoustic analyses indicated in Table 1.

For all homogeneous utterances, the difference scores are not significantly different from zero. In contrast, among the alternating sequences, the difference scores depart from zero in all instances. As

Table 1
Mean Difference Scores and SDs Between
Odd and Even Intervals of Each Sequence

Sequence	Acoustic		EMG	
	Mean	SD	Mean	SD
/bak bak.../	2.00	3.79	13.17	7.71
/bak fak.../	-95.00*	15.27	-3.00	11.08
/duk duk.../	-12.67	9.66	1.33	15.06
/duk suk.../	-84.17*	17.20	-27.00	24.29
/dup dup.../	-5.67	5.01	-5.50	11.58
/dup sup.../	-99.17*	14.81	1.17	26.73

* $p < .001$

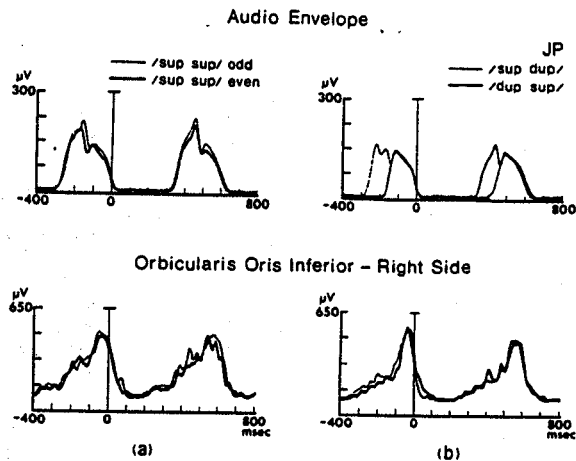


Figure 2. Audio envelope and EMG records for the /sup sup.../ and /dup sup.../ sequence pair for Subject J.P.

predicted, the differences are such that intervals from release of /k/ or /p/ to onset of the fricative are shorter than those between /k/ or /p/ release and stop onset. These results hold for each subject considered individually as well as for the pooled data of Table 1. These two results confirm the earlier work on isochrony in utterances of this type: homogeneous utterances are acoustically isochronous, but alternating utterances depart from isochrony in predictable ways (Fowler, 1979).

The results of the analyses on the EMG data are also given in Table 1. In these data, no difference scores deviate significantly from zero either among the homogeneous or among the alternating sequences.

Figure 2 provides an example of a pair of homogeneous and alternating sequences that display the difference in outcome for the acoustic and EMG measures of isochrony. The figure shows clearly that large acoustic departures from isochrony are obtained in the utterance /dup sup.../, but not in /sup sup.../. The EMG measure of lip rounding and of lip activity for the final /p/ are isochronous in both utterances.

A three-way repeated-measures analysis of variance [measure (EMG vs. acoustic) by sequence type (homogeneous vs. alternating) by utterance (bak vs. duk vs. dup)] was computed on the difference scores. The effects of measure [$F(1,4) = 13.9$, $p < .025$], of sequence type [$F(1,4) = 15.75$, $p < .025$], and of their interaction [$F(1,4) = 9.29$, $p < .05$] were significant. No other effect reached statistical significance. As Table 1 makes clear, the interaction is significant because acoustic difference scores are substantially larger than EMG difference scores for the alternating utterances, but are not larger for homogeneous sequences.

The absence of an effect of utterance type indicates that for the utterances of this experiment, articulatory

activity for the initial consonant is neither more nor less isochronous than activity related to the production of the stressed vowel or stressed vowel and final consonant.

DISCUSSION

These results support an interpretation of the P-center phenomenon that invokes the production strategy of the talker as it is reflected in the acoustic signal. In particular, they suggest that, when asked to produce isochronous monosyllabic utterances, talkers comply by producing isochronous articulatory gestures. Due, at least in part, to differences in the times after articulatory onset that different consonants have acoustic consequences other than silence, the acoustic product of isochronous gestures may depart substantially from isochrony.

The present experiment did not achieve its aim to identify an articulatory P-center. That is, it did not distinguish the production of the initial consonant, the vowel, or even the final consonant in degree of articulatory isochrony. The articulation of all of these segments was isochronous in the simple utterances of the present experiment. We suspect, however, based on some hints in the literature and on a recent study (Fowler & Tassinary, Note 5) that further investigation will distinguish them. In particular, as noted, Allen (1972) and Rapp (Note 1) both place the regulated "beat" within the acoustic realization of a consonant. Unless there is some consistent tendency to postpone the beat or tap involved in these techniques, the results they provide would tend to rule out the initial consonant of a syllable as the timed event detected by listeners in isochronous productions.

In addition, a recent investigation suggests that the initial segments of alternating utterances are isochronous only if they participate in consonant clusters of the same size (Fowler & Tassinary, Note 5). In that study, acoustic measurements were made of alternating sequences in which the initial consonantal context included clusters and single consonants (e.g., /sad strad sad strad . . . / or /tad trad tad trad . . . /). These sequences showed systematic departures from acoustic isochrony. In particular, syllables starting with a cluster were abutted more closely to their predecessors than those starting with a single consonant. Since the syllables beginning with clusters in this study had the same initial consonants as those starting with single consonants, it is unlikely that the articulatory onsets of these syllables were temporally equidistant. In contrast to the initial consonants, the vowels apparently remained isochronous independent of cluster size.

These preliminary findings suggest that further investigation of syllable internal gestures, perhaps those associated with the stressed vowel, are most likely to uncover the regulated events in isochronous productions.

In conjunction with the present findings, the perceptual data of Fowler (1979) and Morton et al. (1976) suggest that listeners' judgments of isochrony are, in some sense, articulation-referential. We briefly consider alternative explanations for these kinds of judgments.

One view of speech perception that explains the articulation-referential character of speech perception is the motor theory (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Indeed, the motor theory was devised to explain data of precisely the present sort in which descriptions of the articulatory generation of speech segments seem to correspond more closely to a listener's identifications of the segments than do descriptions of their acoustic realizations. These findings include the perceptual equivalence of the rising and falling F2 transitions in synthetic versions, respectively, of /di/ and /du/ (Liberman, Delattre, Cooper, & Gerstman, 1954), the perceptual nonequivalence of acoustically identical stop bursts in the synthetic (Liberman, Delattre, & Cooper, 1952) and real (Schatz, 1954) syllables, /pi/, /ka/, and /pu/, various trading relations that obtain among acoustically quite different "cues" for the same phonetic segment (Fitch, Halwes, Erickson, & Liberman, Note 6), and the role of silence in the appropriate contexts as information for stop and affricate manners of articulation (see Liberman & Pisoni, 1977, for a review). The motor theory explains these findings by proposing that listeners access their own production systems in order to identify the sounds of speech.

The motor theory is perhaps unique among views of speech perception in accounting for findings of the foregoing sort. However, because it invokes access to mechanisms for producing the stimulation being perceived, the motor theory implies that the processing underlying speech perception is unlike that underlying auditory nonspeech perception or perception in any other sensory modality. In respect to this proposal, the theory is somewhat unparsimonious and unattractive.

In fact, the data that the motor theory was devised to explain provide no evidence at all relating to *how* articulation-referential percepts are achieved by the perceptual system. That is, they do not support a theory that invokes access to the production system over one that does not. Instead, they reveal only *what* in an acoustic speech signal is perceptually significant—namely, acoustic patterns that reflect articulation, rather than any other acoustic "patterns" that might appear salient to an investigator.

Moreover, in contrast to the view of motor theorists that the articulation-referential character of speech perception makes it unique among perceptual modes, it is possible, instead, that articulation-referential perception is symptomatic of the close alliance between the perception of speech and other kinds of perception.

That is, insofar as it indicates a sensitivity on the part of perceivers to information in proximal stimulation (the acoustic signal in speech or in nonspeech audition, reflected light in vision, patterns of deformation on the skin in touch) that corresponds to or reflects properties of the distal source of the stimulation (the moving vocal tract in speech, a visible or tangible object in vision and touch, respectively), the articulation-referential character of speech perception may be analogous to the object-referential character of vision, touch, and perhaps, nonspeech audition. That is, independent of modality, perception seems generally to involve the pickup of patterns in proximal stimulation that correspond to or reflect aspects (audible, visible, tangible) of distal sources of stimulation.

Explanations as to how this is achieved in perception differ in fundamental ways (e.g., Gibson, 1966; Lindsay & Norman, 1977), but no explanation other than the motor theory invokes access by the perceiver to the means of producing the proximal stimulation. We presume, therefore, that other explanations for articulation reference in speech perception are possible.

Given that the present data do not provide evidence relating to the means by which perceivers achieve their articulation-referential percepts, we will not consider this issue further. However, the data do provide some indications concerning what in an acoustic speech signal provides information about isochrony to a perceiver and what does not. We consider that issue briefly.

We presume, based on the present findings, that the controlled intervals in isochronous productions are articulatory, and hypothesize that the edges of these intervals are the P-centers of successive syllables, as first described by Morton et al. (1976). If the P-center corresponds to an articulatory event to which a perceiver is sensitive, then some acoustic evidence of the P-center must mark its location for him. For example, if the P-center corresponds to the articulatory onset of a stressed vowel, then further investigation should show that it corresponds also to some of the detectable acoustic correlates of the onset in the speech signal. One correlate of the articulatory onset of a vowel is the onset of its influence on the acoustic realization of the consonant. Similarly, if the P-center is found to correspond to the articulatory onset of the syllable-initial or -final consonant, or to the closest approximation of a vowel to its "target" shape of the vocal tract (cf. MacNeilage, 1970), then whatever the acoustic correlates of those articulatory events might be, they should be found to correspond to the perceived locus of the P-center.

We should point out that none of these proposed articulatory markers has been shown to correspond to the P-center, and, indeed, the literature offers apparent counterevidence to some of them. For ex-

ample, as we have pointed out, the articulatory onset of the syllable-initial consonant occurs consistently earlier than the syllable's P-center. For its part, the onset of the syllable-final consonant occurs consistently later than the syllable's P-center. The most likely possibilities may be the vowel's attainment of its target shape of the vocal tract, which is usually achieved at or before the acoustically defined onset of a stressed vowel, or its articulatory onset. Although the articulatory onset of the vowel occurs earlier in the consonant than the P-center (as inferred from the data of Carney & Moll, 1971; Gay, 1977; see also, Bell-Berti & Harris, 1979; Bell-Berti, Harris, Tuller, Baer, & D'Antonio, 1978), any *detectable* acoustic correlates of the vowel's onset may occur somewhat later.

We point out in conclusion that we have demonstrated articulatory isochrony, but only under restricted conditions. First, we have varied only initial consonants in the alternating sequences, and have considered only the stop and fricative manner classes of consonant. To verify that some aspect of articulation remains isochronous in any psychologically isochronous sequence of stressed monosyllables, other variations in initial consonant (including clusters, for example), in the stressed vowel, and in the final consonant require examination. Second, the acoustic departures from isochrony that we have considered are only those due to P-center effects (and, thus, only segmental effects). Larger departures from acoustic isochrony occur in perceptually stress-timed utterances that contain both stressed and unstressed syllables (e.g., Lehiste, 1973; see also Donovan & Darwin, 1979; Lehiste, 1977). The conclusions of the present study cannot be generalized to these departures without investigation of utterances that include unstressed syllables. Despite these limitations in scope, however, we have demonstrated the occurrence of invariant articulatory timing that covaries with the judgments of listeners under conditions in which intervals based on articulation-free acoustic measures show significant variability.

REFERENCE NOTES

1. Rapp, K. *A study of syllable timing* (Papers from the Institute of Linguistics, 8, 14-19). Stockholm: University of Stockholm, 1971.
2. Lindblom, B., & Rapp, K. *Some temporal regularities of spoken Swedish* (Papers from the Institute of Linguistics, 21, 1-59). Stockholm: University of Stockholm, 1973.
3. Kewley-Port, D. *Computer processing of EMG signals at Haskins Laboratories* (Status Report on Speech Research, SR-33, 173-183). New Haven, Conn: Haskins Laboratories, 1973.
4. Kewley-Port, D. *An experimental evaluation of the EMG data processing system: Time constant choice for digital integration* (Status Reports on Speech Research, SR-37/38, 65-72). New Haven, Conn: Haskins Laboratories, 1974.
5. Fowler, C. A., & Tassinari, L. *Stress-timing in speech production and perception*. Manuscript in preparation.
6. Fitch, H., Halwes, T., Erickson, D., & Liberman, A. *Per-*

ceptual equivalence of two acoustic cues for stop-consonant manner (Status Reports on Speech Research, SR-57, 183-200). New Haven, Conn: Haskins Laboratories, 1979.

REFERENCES

- ALLEN, G. The location of rhythmic stress beats in English, I. *Language and Speech*, 1972, 15, 72-100.
- ALLEN, G. D., & LUBKER, J. F. New paint-on electrodes for surface electromyography. *Journal of the Acoustical Society of America*, 1972, 52, 124(A).
- BAER, T., BELL-BERTI, F., & TULLER, B. On determining EMG onset time. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication preprints, 97th Meeting of the Acoustical Society of America, Cambridge, Massachusetts, 1979*. New York: The Acoustical Society of America, 1979.
- BELL-BERTI, F., & HARRIS, K. S. Anticipatory coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America*, 1979, 65, 1268-1270.
- BELL-BERTI, F., HARRIS, K. S., TULLER, B., BAER, T., & D'ANTONIO, L. Some experiments on anticipatory coarticulation. *Journal of the Acoustical Society of America*, 1978, 64(S), S92(A).
- CARNEY, P., & MOLL, K. A cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica*, 1971, 23, 193-202.
- DONOVAN, A., & DARWIN, C. The perceived rhythm of speech. *Ninth International Congress of Phonetic Sciences*. Copenhagen: University of Copenhagen, 1979.
- FOWLER, C. "Perceptual centers" in speech production and perception. *Perception & Psychophysics*, 1979, 25, 375-388.
- GAY, T. Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America*, 1977, 62, 183-193.
- GIBSON, J. *The senses considered as perceptual systems*. Boston: Houghton-Mifflin, 1966.
- KLATT, D. The linguistic uses of segment duration in English. *Journal of the Acoustical Society of America*, 1976, 59, 1208-1221.
- LEHISTE, I. Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America*, 1973, 51, 2018-2024.
- LEHISTE, I. Isochrony revisited. *Journal of Phonetics*, 1977, 5, 253-263.
- LIBERMAN, A., COOPER, F., SHANKWEILER, D., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-461.
- LIBERMAN, A., DELATTRE, P., & COOPER, F. The role of selected stimulus variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, 1952, 65, 497-516.
- LIBERMAN, A., DELATTRE, P., COOPER, F., & GERSTMAN, L. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, 1954, 68(Whole No. 379).
- 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.
- LINDSAY, P. H., & NORMAN, D. *Human information processing* (2nd ed.). New York: Academic Press, 1977.
- MACNEILAGE, P. Motor control of serial ordering of speech. *Psychological Review*, 1970, 77, 182-196.
- MORTON, J., MARCUS, S., & FRANKISH, C. Perceptual centers. *Psychological Review*, 1976, 83, 405-408.
- OLLER, D. The effects of position in utterance on speech segment duration. *Journal of the Acoustical Society of America*, 1973, 54, 1235-1246.
- SCHATZ, C. The role of context in the perception of stops. *Language*, 1954, 30, 47-56.

NOTES

1. In this paper, we will use expressions such as "acoustic onset of the vowel" or "acoustic onset of the consonant." These phrases are loose ways of characterizing acoustic, but not necessarily linguistic or articulatory, boundaries in an optical display of speech. For example, "acoustic onset of the vowel" typically means the onset of periodic energy for the vowel. This is a conventional but, in fact, an incorrect way to mark a vowel's onset. Due to coarticulation, the articulatory onset of a vowel may occur in a preceding consonantal segment, and its acoustic influence on the signal will begin there. The so-called "vowel onset," defined as the onset of periodic energy, is the time at which the vowel predominates in its influence on the character of the acoustic signal—an event that may be of no psychological significance whatever. We begin by using these phrases, and by taking our measures with reference to the boundaries they indicate because they are the measures that have been used to characterize the P-center or stress beat. Eventually, however, we will want to revise the descriptions and the measurements that we take to make them compatible with the acoustic markers that are psychologically significant.

(Received for publication August 27, 1979;
revision accepted February 11, 1980.)