Reprinted From
Attention and Performance IX, J. Long & A. Baddeley (EDS.)
Hillsdale, N. J.: Lawrence Erlbaum Associates

439

# 30

# Natural Measurement Criteria for Speech: The Anisochrony Illusion

**Carol A. Fowler and Louis G. Tassinary**
*Dartmouth College, Hanover, New Hampshire*
*and*
*Haskins Laboratories, New Haven, Connecticut*
*U.S.A.*

## ABSTRACT

When talkers are asked to produce isochronous sequences of monosyllables, they fail in systematic ways, as measured by conventional syllable onset–onset intervals. Listeners, however, hear these attempts as isochronous and hear sequences with equal onset–onset intervals as anisochronous. The agreement between talkers and listeners and the discrepancy between their results and conventional measurements are apparently symptomatic either of a symmetrical production–perception illusion or of systematic measurement error. We examine the second possibility, establishing, in the light of information in the phonetics literature, that the demarcation of syllable onset is flawed if it is intended to mark acoustic reflections of articulatory onset. Moreover, it is flawed in systematic ways that explain many of the talkers' apparent deviations from isochrony. Next we attempt to identify what talkers align temporally in isochronous productions and show that it is a syllable-internal event (as other investigators have also recognized), and not syllable onset. The syllable-internal event appears to be an acoustic reflection of vowel production.

## INTRODUCTION

A sequence of spoken digits with isochronous acoustic energy onsets[1] does not sound evenly timed to listeners. Moreover, when listeners are asked to adjust the intervals between a pair of alternating digits until the sequence sounds isochron-

---

[1] We use awkward locutions such as "acoustic energy onset" or, worse yet, "durations of acoustic energy prior to the measured onset of the vowel" ("prevocalic acoustic duration," for short)

ous, they introduce systematic departures from measured isochrony (Morton, Marcus, & Frankish, 1976). Talkers behave compatibly. When asked to utter an isochronous sequence of uniformly stressed monosyllables differing in initial consonant, they generate sequences that, by conventional acoustic measures, are not evenly timed. However, they reproduce just the departures from measured isochrony that listeners require in order to hear even timing (Fowler, 1979). These departures are characterized by a strong positive relationship between onset–onset intervals among syllables and durations of acoustic energy preceding the measured vowel onset in the initial monosyllables of the intervals.

This cluster of findings is interesting in revealing both a close agreement between listener judgments and talker behavior and a clear discrepancy between measures of these behaviors and measures of the acoustic signal. One explanation for the discrepancy ascribes it to an illusion characteristic of language users both as listeners and as talkers. An alternative explanation ascribes it, complementarily, to systematic failures of conventional measures to reflect the natural constituency of the acoustic speech signal. The first explanation is compatible with accounts of a related phenomenon observed when listeners judge sentence timing. Several studies have found that listeners hear interstress intervals in more-or-less naturally produced sentences as isochronous, even when the intervals depart markedly from measured isochrony. Because this apparent insensitivity to anisochrony is absent when nonspeech intervals are judged, investigators ascribe the judgments to an illusion peculiar to speech (Darwin & Donovan, in press; Lehiste, 1973).

Some reason to doubt this account, at least in its application to the less natural sequences of Morton et al. (1976) or of Fowler (1979), is provided by a study of articulatory timing. Tuller and Fowler (1980) obtained concurrent acoustic and electromyographic records during productions of stressed sequences of monosyllables that talkers had intended to be isochronous. In the study, muscle activity underlying production both of consonants and of vowels was isochronous despite substantial measured acoustic anisochrony. Apparently then, talkers followed instructions to produce even timing by regulating articulatory timing, and their impressions of isochrony were not illusory. Given the close agreement between talkers and listeners regarding isochrony, it is likely that listener judgments of isochrony in the studies of Fowler and Morton et al. are based on acoustic reflections of articulatory timing.

If the acoustic signal does reflect articulatory timing in such a way that listeners can detect it, then we must ask why our acoustic measures do not reflect it accurately. In fact, the reason can be discovered in the phonetics literature

---

to describe parts of the acoustic signal. In the isochrony literature, the more convenient terms, "sound onset" or "phonetic-segment onset" have been used as if they were coextensive with "acoustic energy onset." Likewise, "consonant duration" is used as if it were coextensive with "prevocalic acoustic duration" in a CV syllable. However, a major point we hope to make is that the terms are not coextensive. Because it is critical to keep the referents of these terms separate, it is necessary to keep the terminology distinct at the cost of some elegance.

(Fant, 1962; Lisker, 1957). Fundamentally, it is that the criteria for partitioning the acoustic signal into consonants, vowels, and intersyllable intervals used in the isochrony literature violate a natural partitioning of the acoustic signal into those units in two major ways. First, they equate onsets of acoustic energy for a phonetic segment with the phonetic-segment onset itself; second, they ignore the fact of coarticulation among neighboring segments. The research just cited highlights the first of these errors of measurement, as we explain now, and the results of the experiment we report in the following suggest the second.

## Phonetic Segment Onsets

The criterion for syllable onset used in our previous studies and in those of Morton et al. (1976) and Lehiste (1973) was the onset of acoustic energy for the syllable-initial segment. Intuitively, there is no more reasonable criterion than this because it represents the first visible evidence of syllable onset in a spectrographic or oscillographic display. However, the criterion is flawed as a measure of *articulatory* onset and hence as a measure of phonetic-segment onset for a talker, because it leads the measurer to designate as onset different phases in the production of different manner classes of consonant.[2] In effect, this measurement criterion implies that, for a talker, achieving isochrony requires certain anisochronies of articulation to line up the release phases of oral stops with the closure phases of other syllable-initial segments. In our production experiment just described, talkers did not behave this way. Instead, they more closely aligned closing phase with closing phase, closure with closure, and release with release. Nor would there be any point in a talker's regulating the acoustic-energy onsets of stops in view of the finding that perceivers require evidence of a closure interval (silence in the studies that have investigated the phenomenon) in order to hear later acoustic evidence of release as a stop consonant (Dorman, Raphael, & Liberman, 1979; Liberman & Pisoni, 1977). These observations support fairly strongly the second proposed interpretation of the discrepancy between speaker/listener behavior and acoustic measures. The discrepancy is due to an optical illusion, so to speak, of the measurers rather than to an acoustic illusion of the speaker–listeners.

## Vowel Onsets

Although the study of Tuller and Fowler (1980) revealed articulatory isochrony in utterances with measured acoustic anisochrony, it failed to isolate a particular phonetic segment or any other aspect of the syllable as a possible focus of the

---

[2] Consonants are produced in three broad phases: a closing phase during which an articulator approaches a point of constriction, a closure interval during which the constriction is maintained, and a release phase. The acoustic energy onsets of segments other than oral stops occur at or near the onsets of the closure intervals. But the acoustic energy onsets of syllable-initial oral stops occur at stop release.

talker's timing strategies. An obvious proposal is that the talker intends to initiate the production of syllables at temporally equidistant intervals under isochrony instructions. Our experiment tests this proposal but, before describing the test, we should explain why we expected to *disconfirm* the hypothesis and why we entertained a different one—namely that vowels would be evenly timed under conditions in which the initial segments of a syllable would not.

Our alternative hypothesis was based on the following set of observations:

1. Some investigators have proposed that listeners refer to a syllable-internal location when they make isochrony judgments (Morton et al., 1976). The critical location shifts inward from the measured onset of the syllable as the duration of the prevocalic acoustic signal increases. This hypothesis explains the listener's isochrony judgments equally as well as one assuming that it is the articulatory-syllable onsets to which the listener attends. It has the added advantage of generalizing to some other findings. In particular, listeners' taps to stressed syllables in naturally produced sentences are syllable-internal (Allen, 1972) and Swedish talkers, producing iambic disyllables, align the stressed syllable such that the pulse is syllable-internal (Rapp, 1971).
2. The syllable-internal event has been identified by some investigators with the stress beat of a syllable (Rapp, 1971; Allen, 1972). Allen's conclusion is based on his finding that taps to unstressed syllables have a far less consistent locus than taps to stressed syllables. In Rapp's study, Swedish talkers producing a disyllable in time to a pulse aligned the stressed syllable with the pulse, skipping over the unstressed syllable (Rapp, 1971). Our own studies have included only uniformly stressed syllables and, therefore, cannot in themselves be taken to address the phenomenon of stress. However, in relevant respects, our findings and those of Morton et al. (1976) are compatible with Allen's using spontaneously generated sentences and Rapp's using iambic disyllables. Apparently, phonetic-segmental sources of departure from measured isochrony are not strongly affected by the presence or absence of unstressed syllables in the utterance.
3. Stress is fundamentally a property of a vowel and is not a property of syllable-initial consonants. Linguistic rules of stress placement select syllables for stressing based only on properties of the syllable "rhyme" (that is the vowel and any syllable-final consonants). The syllable "onset" (any prevocalic consonants) is irrelevant in all the dozen or so languages that have been studied in detail (Prince, 1980). In some languages, only the vowel is relevant (Hayes, 1980); in none is the vowel irrelevant.

Based on these considerations, we expected that a study separating the articulation of syllable onset and rhyme would reveal the rhyme as the focus of a talker's timing strategies. A natural hypothesis is that vowel onsets would be regulated by talkers—a hypothesis consistent with the outcome of Tuller and

Fowler (1980). This does not mean that the regulated event would coincide with the conventionally measured vowel onset, however. Instead, due to anticipatory coarticulation of a vowel with a consonant in a CV (Carney & Moll, 1971; Gay, 1977; Ohman, 1966), it should precede the measured vowel onset in any syllable beginning with a consonant.

The experiment had as its major aim to test the hypothesis that talkers regulate the onset of a syllable-initial segment against a hypothesis that they regulate a syllable-internal event. A second aim was to generate information enabling localization of any regulated event if it were to be syllable-internal.

## EXPERIMENT

The first intent of the experiment was addressed by designing two types of utterance. One utterance type included the monosyllable /ad/ produced in alternation with rhyming consonant-initial syllables. For continuant consonants, using conventional acoustic measures, we would expect isochrony or near isochrony if the talker regulates the syllable-initial segments of these syllables but not if he or she regulates a syllable-internal event that shifts in relation to the measured syllable onset with different initial segments. The same logic may be applied to utterances of the second type, in which initial consonants are the same across syllables in the utterance, but the size of a prevocalic cluster is varied (e.g., /sad strad . . . /).

The second aim of the study, to locate and identify any syllable-internal event that might be regulated, was addressed by replicating Rapp's procedure in which talkers produce syllables to a metronome.

### Method

*Subjects.* Subjects were three graduate students, two male and one female, enrolled at Dartmouth College. All three are native English speakers.

*Stimuli and Materials.* We created 33 unique utterances, each a seven-syllable nonsense sentence. As in earlier studies, the monosyllables in the sentences all rhymed with /ad/ but differed in initial consonant or cluster. The orthographic (rather than phonetic) spellings of the sentences were printed in uppercase primer type on 5 × 8 index cards.

Each nonsense sentence was either "homogeneous" or "alternating" in composition. Homogeneous sentences were composed of the same monosyllable repeated seven times (e.g., MOD MOD MOD MOD MOD MOD MOD). Nine were generated from the monosyllables: /ad/, /bad/, /dad/, /fad/, /mad/, /nad/, /pad/, /sad/, and /tad/. Alternating utterances were composed of two different nonsense syllables in alternation. Twenty-four alternating sentences were generated, including: (1) each single-consonant monosyllable alternating with /ad/

(e.g., /sad ad . . . /), plus its mirror reversal (e.g., /ad sad . . . /); and (2) all the following: /strad stad . . . /, /strad sad . . . /, /stad sad . . . /, and /trad tad . . . /, and their mirror reversals.

Two copies of each alternating sentence and four copies of each homogeneous sentence were made. The 84 stimulus cards were split into four groups of 21 with the constraint that groups 2 and 4 and groups 1 and 3 together contained one copy of each alternating sentence and two copies of each homogeneous sentence. Sentences within each group were randomly ordered and this ordering was maintained for all subjects. The groups of 21 sentences were arranged in different pseudorandom orders for each subject with the constraint that either groups 1 and 3 or groups 2 and 4 appear in either of the middle two positions.

*Procedure.* Subjects were instructed to read aloud the sentence on each stimulus card in as rhythmic a fashion as they could, stressing every syllable. "Rhythmic" was defined as onset–onset isochrony. They were asked to repeat each sentence until they felt they had succeeded. All of their productions, practice and final, were recorded, but only the last version of each sentence was used in the analysis.

When the 84 nonmetronome trials were completed, the subject was asked to repeat the procedure with the help of a metronome. The subject listened to the metronome and adjusted it to a comfortable speaking rate. The instructions were the same as before only this time the subject was told to "say each syllable on the beat."

*Measurements.* Spectrograms were made from the recorded sentences generated on trials 22–63 during both phases of the experiment (84 per subject). To avoid obscuring the rhythmicity effects by utterance-initial and -final lengthening (Lindblom & Rapp, 1973; Klatt, 1976) all measurements were confined to intervals including syllables 2 through 6. The low-amplitude frication of /f/ in /fad/ proved difficult to see on many spectrograms. Therefore we did not make measurements on utterances including that syllable.

The measurement criteria adopted for all the intervals measured were as follows:

1. Acoustic onset: The onset was defined as the onset of frication in fricative-initial syllables, the initial release burst in stop-initial syllables, and the onset of nasal resonance in /nad/ and /mad/. Finally, for the syllable /ad/, the glottal stop (if any) served as onset; otherwise, the prevocalic acoustic onset and the acoustic vowel onset were the same.

2. Acoustic vowel onset: For all the syllables, with the exception of /strad/ and /trad/, the vowel onset was the point where the glottally excited, full formant pattern was first evident. For /strad/ and /trad/ the vowel onset was measured as the start of the rising third formant. When this was not observed, the point at which the third formant approximated that of the alternate syllable in both frequency and amplitude was used.

3. Syllable offset: For all the syllables this was taken to be the offset of regular glottal pulsing. The final /d/ release was not used due to the infrequent acoustic evidence of its occurrence.

## Results

The measurements made on each syllable type were averaged across its repetitions. This resulted in 11 numbers per subject for each measure (i.e., one number for each syllable type).[3]

*Analysis of Nonmetronome Utterances.* As a check that subjects were able to produce the homogeneous utterances in a stress-timed rhythm, absolute mean differences in syllable onset asynchrony (SOA) were calculated. This involved averaging the odd SOAs—that is, the onset-onset intervals between syllables 2 and 3 and between syllables 4 and 5—and the even SOAs and taking the absolute value of the difference between odd and even intervals. A mean difference score was derived by averaging across all tokens of a given utterance. The mean deviations from isochrony ranged between 3 msec and 31 msec with a mean deviation of 15 msec.

In contrast to this are the corresponding values for the alternating utterances. In these utterances, odd and even SOAs are different in that their initial and final consonants are reversed with respect to one another. Here the mean deviations from isochrony ranged between 21 msec and 187 msec with an average of 62 msec. Figure 30.1 plots this difference in SOA as a function of differences in the durations of acoustic energy prior to the measured onsets of the vowels (henceforth, "prevocalic acoustic durations"). The figure reveals a linear relationship between these two measures ($r = .88$, $p < .001$). All three subjects showed a relationship of approximately this magnitude. This is similar to the relationship reported in Fowler (1979), but in the present data it is shown to hold even for alternating utterances with identical syllable-initial consonants. If talkers were regulating articulatory syllable onsets, these anisochronies, as well as those of the continuant-vowel alternating utterances, would not be expected.

*Analysis of Metronome Utterances.* As for the nonmetronome utterances, the relationship between SOA differences and differences in prevocalic acoustic

---

[3] Our criterion for overall agreement was based on a comparison of the matrices of intercorrelations among measures for each subject. We counted as an agreement among subjects any instance in which the correlations between a pair of measures were uniformly nonsignificant or were uniformly significant and compatible in sign. We used the obtained frequencies of nonsignificant outcomes and of significant positive and negative correlations to compute the probability of an agreement. That is, we asked, given the obtained probabilities of each outcome for each subject, what is the probability that across subjects the outcomes would assort themselves as they did into 16 agreements and 7 disagreements or into any combination more extreme than this. The computed probability is vanishingly small. Only one of the seven disagreements was due to significant correlations differing in sign.

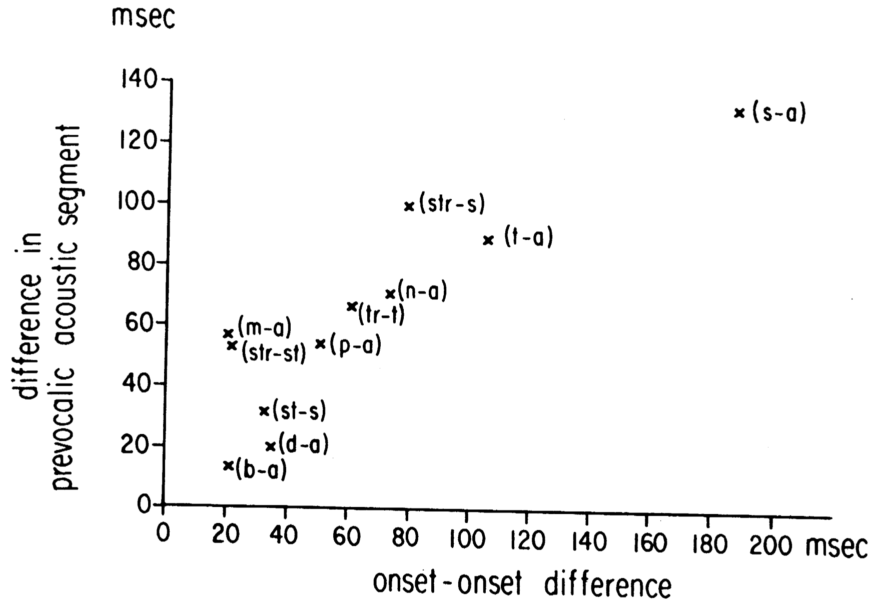

FIG. 30.1. Relationship between onset-onset differences and differences in prevocalic acoustic durations of syllables in alternating utterances.

duration in the metronome utterances is linear ($r = .91$, $p < .001$). Again, the relationship was characteristic of all three talkers. The mean deviations from isochrony ranged from 10 msec to 209 msec, averaging 73 msec.

Figure 30.2 represents the relationship of the metronome pulse to certain intervals within and across syllables. In the figure, the vertical line at zero on the horizontal axis represents the metronome pulse. The different syllable-initial consonants or clusters are listed on the vertical axis. Reading from left to right, points in the figure represent, respectively, the vowel offset of the preceding syllable, the prevocalic acoustic onset of the syllable, the acoustically defined vowel onset, and the vowel offset. These measures are collapsed over talkers and over homogeneous and alternating utterances. The numbered horizontal lines represent intervals that we discuss.

The figure closely resembles Rapp's (1971) Fig. I-B-6 of disyllables produced repeatedly to a metronome by native Swedish speakers. One notable outcome, also evident in Rapp's data, is the locus of the metronome pulse, which tends *not* to coincide with the boundary of any acoustic segment. In itself, this does not rule out acoustic segments as critically timed events in isochronous speech, because we cannot be sure that the pulse location overlays a critical locus of timing control in the syllable. For some perhaps psychoacoustic reason, the pulse might be consistently offset from such a locus. However, even if it were, the figure makes clear that no shift forward or back would cause the pulse to coincide

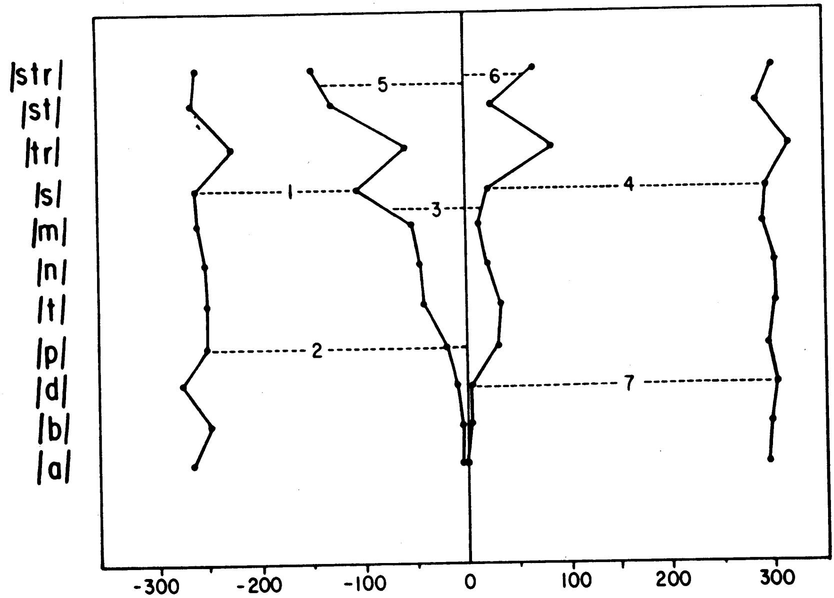

FIG. 30.2. Syllables of the experiment averaged around the metronome pulse. The pulse is the vertical line at zero on the horizontal axis. Along the vertical axis are the different syllable-initial consonants or clusters arranged approximately according to measured duration. Measure 1: intersyllable interval; Measure 2: preceding syllable offset to pulse; Measure 3: prevocalic segment duration; Measure 4: acoustically defined vowel duration; Measure 5: prevocalic acoustic onset to pulse; Measure 6: pulse to measured vowel onset; Measure 7: pulse to measured vowel offset.

consistently with any acoustic-segment boundary except the vowel offsets 260 msec to 300 msec away. Thus, to the extent that we expect to find a particular locus of temporal control in the talker's isochronous productions, the event probably is not coincident with any conventionally measured acoustic segment.[4]

In view of these considerations, the most conservative way to make use of the metronome pulse is in aligning the different syllables with respect to each other as we have done in the figure. From this alignment, several relationships among measures are apparent. In particular, several measures appear to vary systematically with changes in syllable-initial segment and segment duration. As the prevocalic acoustic duration (measure 3) increases, the metronome pulse falls progressively inward from the acoustically defined syllable onset (measure 5) and progressively back from the acoustically defined vowel onset (measure 6).

---

[4] There is also the possibility that there is no special locus of timing control. Isochrony among syllables may be achieved by cyclic articulations in which no one part of the cycle is special. We consider this possibility in Discussion, p. 532.

Separate planned comparisons show that the effect of syllable type, ordered according to measure 3, on measures 5 and 6 is linear, $F(1, 20) = 193.3$, $p < .001$, and $F(1, 20) = 35.2$, $p < .001$, respectively. For its part, measured vowel duration (measure 4) decreases with increases in prevocalic-segment duration. Compression of a vowel in the context of a preceding or following consonant or cluster is commonly reported in the literature (Rapp, 1971; Lehiste, 1970). Vowel duration also is linearly related to syllable type ordered by prevocalic segment duration, $F(1, 20) = 78.172$, $p < .001$. It is interesting that both the vowel's compression and the decrease in preceding vowel offset to syllable onset (measure 1) are just complemented by increases in measures 5 and 6 so that the intervals from preceding vowel offset to pulse (measure 2) and pulse to vowel offset (measure 7) remain invariant across changes in syllable-initial consonant. Planned comparisons on these measures yield $F$ values less than 1. (Planned comparisons performed on the data of individual subjects also failed to reach significance.)

## Discussion

The results of the study allow us to reject the hypothesis that talkers regulate articulatory syllable onsets when instructed to produce isochronous sequences. Systematic departures from measured acoustic isochrony were observed for alternating pairs of syllables beginning with the same consonant and for alternating syllables with initial continuant consonants and vowels.

However, isochrony was observed for vowel offsets, implicating the vowel as a regulated segment.[5] This does not necessarily imply that a talker's strategy for producing isochronous monosyllables is one of timing vowel offsets. If vowel offsets are isochronous, so are any points temporally equidistant from the offset. The offset was simply a point that we were able to identify consistently. It may or may not be significant in talking and listening.

In fact, we tend to discount the offset as a likely focus of regulation for two reasons. First it is implausible a priori, particularly in view of the linguistic theories of prosody cited earlier. In these theories, the syllable rhyme, consisting of vowel and postvocalic consonants, is central to explanations of metrical structure in speech. The vowel offset breaks up the rhyme and groups the vowel with the prevocalic consonants, which have no role in explanations of prosody.

---

[5] Given that the criterion for vowel offset was termination of voicing signaling closure for the following /d/, it is possible that this location in the syllable does not correspond to vowel offset for a talker or listener, who may be sensitive to the carryover coarticulation between vowel and final consonant. However, because the final consonant was invariably /d/ and the vowel /a/, the extent of carryover coarticulation should be invariant over syllables. Therefore, the measured vowel offset should be temporally equidistant from the true offset and the fact remains that vowel offsets were isochronous in these productions.

A second reason for discounting the vowel offset as a regulated event is that the metronome pulse fell about as far as possible from the vowel offsets. It was not even the case that the pulse was midway between offsets. Rather it consistently followed an offset by about 260 msec and preceded the next one by 300 msec.

A different proposal is that vowels are produced invariantly and cyclically in these sequences and may or may not have a focal regulatory locus such as vowel onset or offset. It may seem as if this account is somewhat infirmed by the measured shortening of the vowel (measure 4) in the context of long prevocalic acoustic segments. However, as noted earlier, the articulatory onsets of vowels occur within the production of preceding consonants so that their produced extent exceeds their conventionally measured extent. For this anticipatory coarticulation to explain the vowel shortening as apparent—that is, as unaccompanied by an articulatory decrease in vowel duration—it would be necessary for more of a vowel's produced extent to coarticulate with consonants and clusters having longer measured durations.

The speech production literature does not offer either confirming or disconfirming evidence on this point. However, bringing what evidence there is together with a pair of assumptions entailed by the current proposal, we can reproduce the major findings of our study. The evidence concerns timing of the different phases of consonant production and the relative timing of consonants and vowels in CV syllables. It is most complete for voiced and voiceless stops and fricatives, so our description is confined to these segments. The relevant findings from the literature are first that time to closure for voiceless stops is less than for a voiced stop (Chen, 1970). In a VCV this difference is counteracted by differences in closure duration, where voiceless stops have longer closure intervals. However, in syllable-initial position, the closure intervals are equal (Stathopoulos & Weismer, 1979). This implies that voiceless stops will have acoustic consequences relatively sooner after articulatory onset than voiced stops. This is verified in a study of vocal reaction time for CVs (Fowler, 1979) where latency for voiced stops was 358 msec and that for voiceless stops, 340 msec. The direction of difference was consistent across the three voiced-voiceless pairs and significant across the group of five subjects.

There is some evidence that, intervocalically, fricatives have a generally slower closing velocity than stops (see MacNeilage & Ladefoged, 1976, for a review), but the difference apparently is absent in syllable-initial position (Kuehn & Moll, 1976). In any event, fricatives will have acoustic consequences near the beginning of their closure phases and hence, compared to stops, a relatively short time after their articulatory onsets. This difference too is reflected in the vocal reaction time for fricatives relative to stops (Fowler, 1979).

The literature on coarticulation in CVs (in fact, generally, in VCVs) shows evidence of vowel initiation during the closing phase of voiced stops (Ohman,

1966) and fricatives (Carney & Moll, 1971) but at the onset of the closure interval for voiceless stops (Gay, 1977).

All of these facts may be accommodated in Fig. 30.2 if we suppose that consonants in CVs (but not CCVs or CCCVs) are uniformly initiated say, 175–200 msec prior to the metronome pulse and that vowels are invariantly initiated 50–75 msec later. Initial consonants in cluster-vowel syllables must be initiated relatively earlier than those in CVs. Thus, the proposal that vowels are generated cyclically in these productions is compatible with known facts of consonant timing and CV coarticulation and is not infirmed by the occurrence of measured vowel compression.

A remaining issue to consider is whether or not vowels have a focal regulatory locus. As we pointed out in Footnote 4 (p. 529), there need not be any special regulatory focus in the cycle. However, if there is none, we have no way to account for the consistent alignment of the syllables with the metronome pulse. Yet the alignment characteristic of the composite shown in Fig. 30.2 was characteristic of each talker.

Possible explanations for the location of the metronome pulse—and also, we think for the stress beat—would invoke a suprasegmental marker such as an intensity peak or a pitch indicator. We have ruled out most intensity-based explanations by showing that listeners make the same pattern of isochrony judgments to syllable sequences having undergone infinite peak clipping as to sequences with a normal intensity contour (Tuller & Fowler, 1981; see also Morton et al., 1976). Examinations of the syllables used in our studies have not revealed any pitch marker corresponding to the metronome-pulse location or to any point temporally equidistant to it.

An alternative is suggested by the finding that for all three talkers, the metronome pulse fell directly on the vowel's measured acoustic onset (within ±6 msec) in the syllable /ad/. This suggests a possible alignment of the metronome pulse with achievement of the vowel's target vocal-tract shape, a proposal that later studies will test.

In our view, the hypothesis to pursue is that vowels are rhythmically timed in sequences of monosyllables believed by talkers and listeners to be isochronous. We have attempted to obtain evidence that the metronome pulse marks a perceived correlate of the vowel by obtaining measures of perceived relative vowel duration in syllables having different initial consonants. If the metronome pulse does mark a vowel correlate for talkers and listeners, then vowels of the same measured duration should seem relatively longer in the syllable /sad/, say, than in /ad/. We tested this (Fowler & Tassinary, in preparation) using a procedure developed by Raphael (1972). He showed that voicing of a final consonant may be cued by preceding vowel duration—a long-duration vowel cuing final voicing. We used the syllables /sad/, /mad/, /bad/, and /ad/ produced by the three talkers of the present study. From each syllable, we constructed a continuum of 10 different measured vowel durations by removing pitch pulses from each

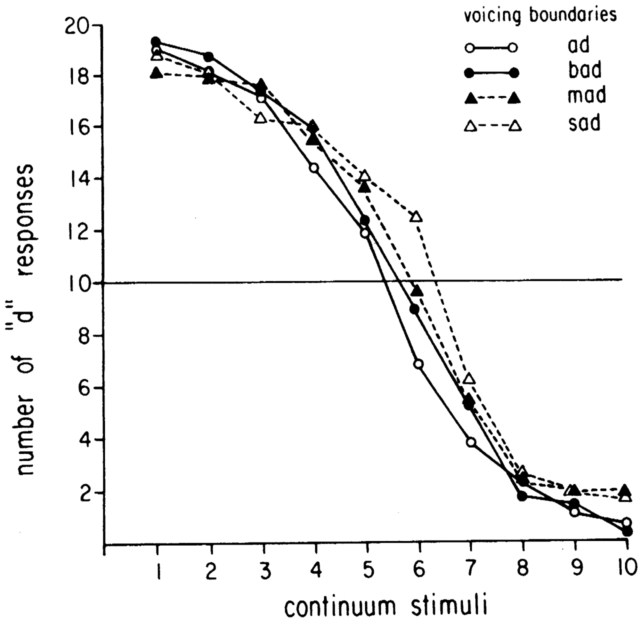

**FIG. 30.3.** Labeling functions for the four vowel-duration continua: /ad/, /bad/, /mad/, and /sad/ collapsed over two talkers (Fowler and Tassinary, in preparation).

vowel's steady states. Subjects were asked to label the final consonant of each syllable as /d/ or /t/. The hypothesis was that subjects would resist hearing /t/ increasingly in the series: /ad/, /bad/, /mad/, and /sad/ compatible with the increasing metronome precession of the measured vowel onset in the series (see Fig. 30.2).

Figure 30.3 shows the outcome for two of the three talkers. Listeners were unable to label the final consonants of syllables spoken by the third talker consistently enough to enable us to use their responses. Although the effects on syllables of the remaining talkers were smaller than those predicted based on the metronome precession of the measured vowel onset, the ordinal relationships among the labeling curves are exactly as predicted, $F(1, 3) = 21.6$, $p = .02$, according to a planned comparison.

Together with the results of the metronome study, these results suggest that the vowel onset, as it is generally located in a spectrographic display, has no particular status in perceiving or producing isochronous speech. Except for its salience in displays of speech, it is not surprising that it has no special status, because it does not correspond to an occasion when a phonetic segment begins or reaches its articulatory target. It corresponds only to a point in time when a consonant ceases to predominate in its effects on the acoustic signal visibly displayed. The best evidence we have now suggests that an acoustic measure of intersyllable interval that provides the most natural measure of syllable timing for talkers and listeners will be one time-locked to vowel production.

## ACKNOWLEDGMENTS

The research reported here was supported by NINCDS Grant NS 13617 and NIH Grant HD 01994 to Haskins Laboratories and by Biomedical Research Support Grant RR 05392 from the National Institute of Health to Dartmouth College. We thank Scott Orr, Angelo Strenta, and Kathy Vaughan for their patient participation as talkers in the experiment and George Wolford for advice on analysis and for reviewing an earlier version of the manuscript.

## REFERENCES

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 cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica,* 1971, *23,* 193-202.

Chen, M. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica,* 1970, *22,* 129-159.

Darwin, C., & Donovan, A. Perceptual studies of speech rhythm: Isochrony and intonation. In J. C. Simon (Ed.), *Language generation and understanding,* in press.

Dorman, M., Raphael, L., & Liberman, A. Some experiments on the sound of silence in phonetic perception. *Haskins Laboratories Status Reports on Speech Research,* 1979, *SR-58,* 105-137.

Fant, G. Descriptive analysis of the acoustic aspect of speech. *Logos,* 1962, *5,* 3-17.

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

Fowler, C., & Tassinary, L. Manuscript in preparation.

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

Hayes, B. *A metrical theory of stress rules.* Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1980.

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

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

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

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

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

Lindblom, B., & Rapp, K. Some temporal regularities of spoken Swedish. *Papers in Linguistics from the University of Stockholm,* 1973, *21,* 1-59.

Lisker, L. Linguistic segments, acoustic segments and synthetic speech. *Language,* 1957, *33,* 370-374.

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.

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

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

Prince, A. A metrical theory for Estonian quantity. *Linguistic Inquiry,* 1980, *11,* 511-562.

Raphael, L. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America,* 1972, *51,* 1296-1303.

Rapp, K. A study of syllable-timing. *Papers in Linguistics from the University of Stockholm,* 1971, *8,* 14-19.

Stathopoulos, E., & Weismer, G. The duration of stop consonants. In J. Wolf & D. H. Klatt (Eds.), *Speech Communication Papers.* New York: Acoustical Society of America, 1979.

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

Tuller, B., & Fowler, C. The contribution of amplitude to the perception of isochrony. *Haskins Laboratories Status Reports on Speech Research*, 1981, *SR-65,* 245-250.