Journal of Phonetics (1987) **15**, 1–13

The effect of vowel quality variations on stress-beat location

Robert Allen Fox

Speech & Hearing Science, Department of Communication, 324 Derby Hall, 154 N Oval Mall, Columbus, OH 43210, U.S.A.

and

Ilse Lehiste

Department of Linguistics, The Ohio State University, 190 North Oval Mall, Columbus, OH 43210, U.S.A.

Received 27th November 1985, and in revised form 5th September 1986

This study examines the effect which medial vowel variations have upon the rhythmic production and perception of stressed CVC monosyllables. Experiment 1, a production test, demonstrated that as a talker produces a longer vowel, the stress beat of the monosyllable moves away from the vowel onset, all other things being equal. Experiment 2 demonstrated that the perceived stress-beat location of a stressed monosyllable also moves away from the vowel onset as a function of the vowel duration and that the stress beat occurs later in tense (long) vowels than in lax (short) vowels. Experiment 3 demonstrated that this difference between tense and lax vowels disappears when the durations of the vowels are equalized; thus the effect is due to actual duration rather than vowel quality. Thus although tense vowels may be expected to have greater length, listeners did not react to vowel quality cues in terms of greater length anticipation. The data support the contention that stress-beat location is determined on the basis of the structure of the entire token.

1. Introduction

Naturally-spoken language is perceived as rhythmical when heard by native speakers (Classe, 1939; Abercrombie, 1964; Lehiste, 1972; Donovan & Darwin, 1979). However, research directed at the discovery of the nature of the timing characteristics of speech has failed to discover strict regularity or isochrony between acoustically defined intervals. For example, in production experiments on English (usually classified as a "stress-timed" language) one often finds considerable durational variation among interstress intervals (e.g. Classe, 1939; Shen & Peterson, 1962; Bolinger, 1965; Lehiste, 1972).

Experiments on the perception of speech have also shown that listeners' judgments of rhythmicity are not based upon intervals between the onset of acoustic energy of successive syllables (or perhaps stressed syllables) in a sequence. For example, Morton, Marcus & Frankish (1976) showed that sequences of spoken digits were judged to be

0095–4470/87/010001 + 13 $03.00/0

irregularly spaced when presented with regular intervals between onsets of acoustic energy of the digits. When allowed to adjust the intervals between the digit onsets in order to make them sound regular, listeners introduced systematic deviations from isochrony. Fowler (1979) found that the acoustic deviations from isochrony that appeared in the speech of talkers attempting to produce isochronous sequences of speech were precisely those anisochronies required by listeners to perceive the utterances as regular.

It is apparent that listeners and talkers are capable of focusing on some aspect of orally produced speech when required to produce regular sequences of speech or to make timing judgments. Morton *et al.* (1976) introduced the term "perceptual center" or "P-center" to reference that phenomenon in a word or syllable that must be regularly spaced in time with respect to other P-centers in a sequence of words or syllables in order for the sequence to sound isochronous. The P-center presumably corresponds to the locus of the "stress beat" (Allen, 1972*a*; Rapp, 1971). This stress-beat does not seem to correspond to any obvious acoustic event such as acoustic onset, acoustic energy peak, or vowel onset (Morton *et al.*, 1976; Tuller & Fowler, 1980; Marcus, 1981; Fowler & Tassinary, 1981) but rather can be affected by the nature of the initial consonant (Fowler, 1979) and the duration of the medial vowel and final consonants (Marcus, 1981) in CVC syllables. The P-center, or stress beat, may be a universal phenomenon, since at least one study (Hoequist, 1983) has shown that talkers from each of three languages (English, Spanish, and Japanese) behave in a similar way when required to produce isochronous syllables. That is, they align some point in the syllables which does not correspond either to acoustic onset or vowel onset, but lies somewhere in between. The similarity among these languages with regard to stress-beat location is all the more interesting since these languages have been identified as having different isochronous units (stress foot, syllable, and mora).

Fowler and co-workers (e.g. Fowler, 1979; 1983; Fowler & Tassinary, 1981; Tuller & Fowler, 1980) have suggested that the stress beat may correlate better with some articulatory event, such as the onset of the production of the medial vowel in CVC syllables. Tuller & Fowler (1980), for example, made EMG measurements of lip activity by talkers during the production of regular sequences of tokens beginning with different consonants. They found that isochrony could be found in muscle activity in the lips, even though the acoustic signal displayed deviations from isochrony. Fowler argues that the acoustic measures often considered as corresponding to vowel onset (e.g. onset of periodicity in the acoustic signal) will often deviate from isochrony, even in the event of articulatory isochrony, because of anticipatory coarticulation. In particular, the articulatory vowel onset may tend to occur during the production of the preceding consonant (see Carney & Moll, 1971; Gay, 1977; and the relevant discussions in Fowler, 1983).

The perceived stress beat does not seem to be related to the articulatory onset in a simple manner; however, since Marcus (1981) demonstrated that modification of relatively peripheral aspects of the speech signal, such as the duration of the /t/ closure in the word "eight" (which would presumably not necessarily affect the perceived articulatory onset of the vowel) affected the listeners' perception of the location of the stress beat. From his results (and those of Smith & Fowler, 1984) it seems clear that aspects of the entire signal, and not just the onset of a particular acoustic or articulatory event, may be involved in the perception and production of timing in speech and in the location of the stress beat.

Recent accounts of how stress-beat locations are determined have tended to emphasize either acoustic parameters (e.g. Marcus, 1981) or articulatory parameters (e.g. Fowler, 1979, 1983; Tuller & Fowler, 1980). However, as Marcus has stated, acoustic events have their origin in articulatory gestures and the organization of speech production plans must take into account such perceptual consequences. Thus there is little point in advocating either a strictly acoustic or strictly articulatory account. What is important is the determination of the extent to which each separate aspect of a speech token contributes to the token's stress-beat location and how the structure of the entire token affects both the production and perception of the stress beat.

The basic purpose of the present study was to examine the effect of vowel quality variations (and associated vowel duration variations) on the rhythmic production and perception of stressed CVC monosyllables. The data obtained by Marcus (1981) and Fox & Lehiste (1985) suggest that as the duration of the vowel increases, the stress-beat location should move away from the beginning of the monosyllable further into the vowel. However, both these studies examined the perceptual and/or production effects of vowel duration on stress-beat location without varying the phonetic quality of the vowel. The present study attempts to determine whether there might be strictly phonetic effects (i.e. effects associated with phonetic quality alone or with expected intrinsic vowel duration differences) in addition to acoustic effects (i.e. actual durational differences). As is well known, tense vowels for English have greater intrinsic duration than lax vowels (Peterson & Lehiste, 1960), and speakers of English may be expected to associate phonetic vowel quality with the appropriate intrinsic duration.

The first experiment, a production experiment, required talkers to produce a series of monosyllabic tokens which differed only in terms of the identity of the medial vowel. Experiment 1 is similar to a study reported by Smith & Fowler (1984)[1] except that a more representative sample of American English vowels is included (they used five vowels, this study uses 15 and includes both monphthongs and diphthongs), and both metronome and non-metronome conditions are utilized.

2. Experiment 1

2.1. *Method*

2.1.1. *Talkers*

Three talkers, two female and one male, native to the purposes of the experiment, were paid for their participation. All were students at the Ohio State University with no known hearing impairment. Although naive to the goal of the present experiment, all three subjects had been participants in a long series of experiments requiring them to produce rhythmic sequences of speech stimuli under conditions similar to those found here. They should be considered, therefore, well-practiced.

2.1.2. *Stimuli*

The basic stimuli consisted of sets of seven-token sequences similar to those used by Fowler (1979) and Fowler & Tassinary (1981). Each sequence was composed of seven

[1] Experiment 1 was first designed in the Autumn of 1982. The data were collected in Autumn 1984 as a part of a project approved for funding by NINCDS in the Spring of 1984. The Smith & Fowler (1984) paper only came to our attention in the Winter Quarter of 1985.

monosyllabic tokens, most of which corresponded to real English words. There were 15 different tokens, each of which had a CVC syllable structure. All the stimuli began with /s/ and ended in /t/. The medial vowel was one of the following: /i, I, eI, ε, ae, a, ʌ, ɔ, oU, u, aI, oI, au, ər/. This set of vowels consists of six nonrhotic tense monophthongs (/i, eI, a, oU, ɔ, u/), five nonrhotic lax monophthongs (/I, ε, ae, ʌ, U/), the rhotic vowel /ər/, and three true diphthongs (/ai, oI, aU/).

Each seven-token sequence was either "homogeneous" or "alternating" in nature. Homogeneous sequences were composed of a token repeated seven times (e.g. *sit sit sit sit sit sit sit*). There were 15 different homogeneous sequences. Alternating sequences were composed of two different tokens in alternation (e.g. *sit sight sit sight sit sight sit*). These sequences were constructed by having each monosyllabic token alternate with the *sight* token. Twenty-eight different alternating sequences were generated. Fourteen sequences began with *sight* (e.g. *sight sit sight* . . .), the Type 1 alternating condition, and 14 were their mirror-reversals (e.g. *sit sight sit* . . .), the Type 2 alternating condition. The stimulus set included four occurrences of each homogeneous sequence (eight occurrences of the *sight* homogenous sequence) and two occurrences of each alternating sequence, for a total of 120 sequences. The sequences were put into random order and presented to talkers one at a time on a CRT screen under the control of a computer.

2.1.3. *Procedure*

Talkers were instructed to read the seven-token sequence which appeared on the screen and to produce it in a rhythmic fashion. Rhythmicity was described in terms of isochronous onset-to-onset intervals. If a talker was dissatisfied with his/her production on any trial, the talker was instructed to repeat the sequence. All of a talker's productions were recorded, but whenever a talker repeated a sequence, only the last production was measured.

The talkers read the sequences under two different conditions: metronome and non-metronome. In the metronome condition talkers were instructed to produce the tokens on the beat of the metronome. The metronome pulse was a 1000 Hz square wave 100 ms in duration. A metronome pulse occurred every 1000 ms. The talkers heard the metronome pulses continuously throughout the session. In the non-metronome condition talkers were only instructed to be rhythmic in their productions and were given no external timing cues. For all talkers, the metronome condition preceded the non-metronome condition (similar, though not identical, in design to that used by Hoequist, 1983), but both recordings were made during the same session. Note that this design may prejudice talkers to produce tokens more isochronously than they would in normal speech, but this experiment is designed to examine whether talkers introduce anisochronies even under such controlled conditions.

2.1.4. *Measurements*

For each sequence the following measurements were made from graphs produced using a Siemens mingograf (50 cm s^{-1}): duration of the initial fricative, duration of the medial vowel, and intervals between vowel onsets. The start of any detectable acoustic energy in the token was defined as the fricative onset. The start of periodicity was defined as the acoustic vocal onset (and the offset of the fricative). All tokens ended with the stop [t], so the end of the vowel was defined as the abrupt cessation of energy corresponding to the stop closure. In the metronome condition, the interval between the vowel onset and the metronome pulse onset was also measured. In this paper we will assume that any obtained shifts in the vowel onsets, relative to either the metronome pulse (in the metronome condition) or to adjacent vowel onsets (in the non-metronome condition) will be indicative of stress-beat location shifts. Since the stress-beat of a monosyllable does not seem to correspond to any single acoustic event (cf. Fowler, 1979, 1983; Marcus, 1981), its location cannot be measured directly. In this paper, therefore, we are assuming explicitly that the onset of the vowel gives a good indication of the stress beat's relative position. Note that since all tokens begin with the same segment (/s/), we should not encounter the type of measurement errors discussed by Fowler (1979, 1983) and Fowler & Tassinary (1981) associated with different manners of articulation.

2.2. *Results and discussion*

2.2.1. *Metronome data*

Figure 1 presents the results from the metronome condition averaged across all three talkers. This figure shows the mean locations of fricative onset, vowel voicing onset, and stop closure onset for each of the 15 vowels, relative to the metronome pulse. The vertical line in the middle of the figure indicates the onset of the metronome pulse. The bottom 11 vowels are the lax (short) and tense (long) nonrhotic vowels, while the top four vowels represent the rhotic vowel /ər/ and the three true diphthongs. In each section of the figure the vowels are ordered in terms of mean measured duration (short to long from bottom to top). In general, there is a trend for the onsets of the longer vowels to occur earlier than the onsets of the shorter vowels, although this pattern is disrupted in the true diphthongs and /ər/. Regression analysis of the mean intervals between the metronome

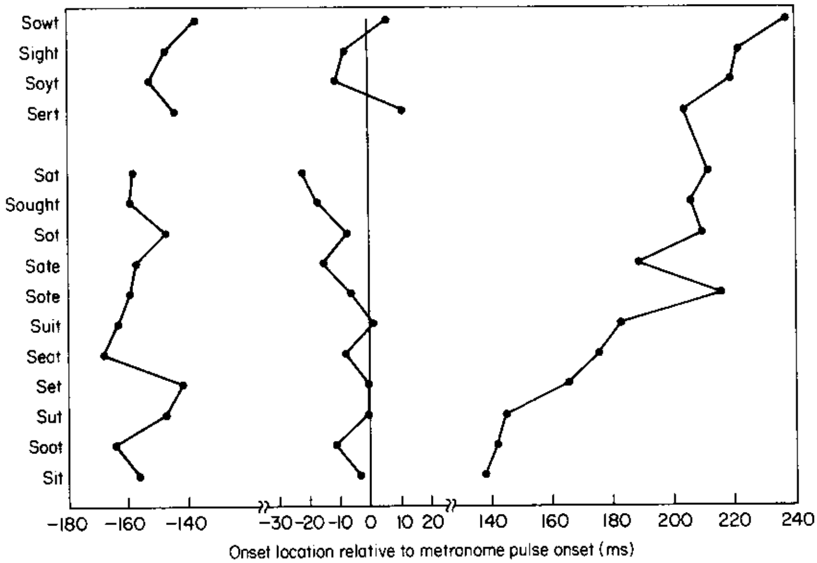

Figure 1. Mean locations of fricative energy onset, vowel voicing onset, and stop closure onset relative to the metronome pulse collapsed over talkers. The vertical line at 0 ms represents the onset of the metronome pulse. Different tokens are plotted bottom to top in the figure. The leftmost points represent fricative onset, the middle points represent vowel onset, and the rightmost points represent stop closure onset.

pulses and the vowel onsets for the 11 nonrhotic monophthongal tokens were significantly predicted by the mean vowel duration [$r(11) = -0.66$, $p < 0.03$]. Assuming that the talker aligns the production of the stress beat with the metronome pulse, it seems that when a talker produces a longer vowel, the stress-beat moves away from the vowel onset toward the end of the token, all other things being equal. This means that the longer vowel has to be started earlier than a shorter vowel, in order to keep the stress beat in the same position relative to the metronome pulse. The slope of the regression line (-0.16) is near the value of the parameter relating medial vowel and final consonant duration to the location of the perceptual center in digits obtained by Marcus (1981).

However, this tendency is not uniform across talkers. Although both talkers 1 and 3 showed significant negative correlations between vowel duration and vowel onset [talker 1: $r(11) = -0.67, p < 0.01$; talker 3: $r(11) = -0.59, p < 0.03$], talker 2 showed a significant positive correlation between vowel duration and vowel onset [$r(11) = 0.66$, $p < 0.013$]. A repeated measures analysis of variance using all 15 vowels [using a general linear model procedure, Ray (1982)] showed that there was a significant effect due to talkers [$F(2,28) = 14.1, p < 0.001$]. A Scheffé test showed that talker 2 was significantly different from both talkers 1 and 3. This suggests that talkers differ either in their ability to do the task or else utilize different rhythmic strategies. Although Marcus (1981) argued that such individual variations are only of peripheral importance to the issue of speech timing, our data demonstrate that such individual variation can significantly affect the evaluation of the data from a production experiment. When talker 2 was eliminated from the data, a planned comparison (trend analysis) showed that the effect of medial vowel, ordered according to mean vowel duration, was linear across all 15 different tokens [$F(1,14) = 7.25$, $p < 0.02$]. The same planned comparison is non-significant if talker 2 is included [$F(1,28) = 0.85, p > 0.30$].

2.2.2. *Non-metronome data*

We next turn to the non-metronome data. As one measure of the degree to which the token onsets were produced in an isochronous manner, the signed differences between the mean even and odd interval durations were calculated for each stimulus. This involved averaging the second, fourth, and sixth intervals (EVEN) and the third and fifth intervals (ODD) and then obtaining the ODD–EVEN difference (ODDEVEN). The first interval was not used in the ODD calculations because an earlier study using a similar task (Fox & Lehiste, 1985) found that these three talkers produced a first interval which was significantly shorter than intervals two and four. No shortening effect was found for any of the other five intervals. The ODDEVEN measure is similar to the absolute durational differences calculated by Fowler (1979) and the stimulus onset asynchrony (SOA) differences calculated by Smith & Fowler (1984). If the intervals between the vowel onsets in the sequences are isochonous, there should be no significant difference between the ODD and EVEN measures. However, if there is a systematic variation in the location of alternating vowel onsets, then we should find differences between the ODD and EVEN measures. In terms of stress-beat location, if a short vowel has its stress beat nearer the vowel onset than does a long vowel, then there should be a substantial difference in the ODDEVEN measure between the Type 1 and 2 alternating sequences. Figure 2 shows the mean ODDEVEN values for both alternating conditions for short (lax) vowels, long (tense) vowels, and the true diphthongs. As can be seen, as the vowel lengthens (and its length becomes more similar to that of the standard sight alternating token); the difference in ODDEVEN values is almost eliminated.

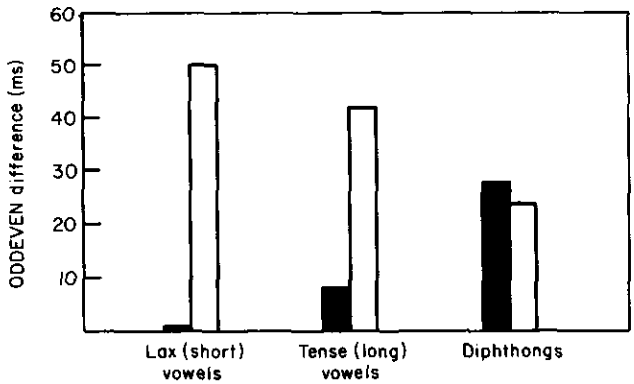

Figure 2. Mean ODDEVEN values for both alternation conditions for tense and lax monopthongs, and true diphthongs for the nonmetronome data. (■) Type 1 alternation; (□) type 2 alternation.

A second way used to analyze the non-metronome data was to regress the duration differences between adjacent intervals (interval n minus interval $n + 1$) against the vowel, fricative, and total durations of the tokens beginning the two intervals, plus the differences in their vowel, fricative, and total durations (e.g. vowel n duration minus vowel $n + 1$ duration). Regression analysis, using data from all three sequence types for each separate talker, generally showed that the difference in vowel duration between two adjacent tokens significantly predicted the interval difference. In particular, vowel duration differences were significantly related to interval differences in both talker 1 [$r = 0.35$, $F(1,576) = 78.7$, $p < 0.001$] and talker 3 [$r = 0.47$, $F(1,477) = 134.7$, $p < 0.001$][2]. For talker 2 the frication duration difference was the only parameter significantly related to interval differnce [$r = 0.15$, $F(1,597) = 14.3$, $p < 0.001$], but this accounted for very little of the variance (2.3%). Thus in both the metronome and non-metronome conditions, talker 2 behaved differently than talkers 1 and 3.

In summary, the data obtained demonstrate that, for two of the three talkers, as a talker lengthens his or her vowel durations, the stress beat of the monosyllable moves toward the end of the token. The data support the view that the stress-beat location of naturally produced monosyllabic tokens can be predicted on the basis of the acoustic (or phonetic) structure of the entire syllable. These data do not directly address the issue of how the articulatory events related to stress beats (or perceptual centers) are affected by vowel quality or duration variations. The next obvious question is whether the perceived stress-beat location is affected in the same manner by vowel quality variations.

In terms of the perception of the stress beat, Marcus (1981) showed that when the duration of the word "one" was modified by artificially lengthening the duration of the vowel listeners located the stress beat (or perceptual center) nearer the end of the token. This demonstrated that vowel length alone could affect the perceived location of the stress beat. Fox & Lehiste (1985) examined the effects of final consonant modification upon the perceived location of the stress beat in CVC tokens and found that as the duration of the medial vowel increased (the phonetic quality of the vowel remained the

[2]Although all talkers were recorded using the same stimulus materials, the degrees of freedom for individual talkers differ because of missing data. In particular, talkers on occasion produced an incorrect token (whose measurements were not included) or else reliable acoustic measurements could not be made on some portion or portions of their recorded production.

same for all tokens), the between-token interval duration required for perceived isochrony decreased. We interpret these results as showing that, again, as the duration of the vowel in a stressed monosyllable increases, the perceived stress beat location of the monosyllable moves away from the onset of the token. Experiment 2 was conducted to determine whether the same tendency could be observed when using naturally produced stimuli whose medial vowels differ in terms of both duration and phonetic quality. This is not a trivial question since it is possible that listeners could compensate for the expected durational differences associated with phonetic quality differences.

3. Experiment 2

3.1. *Method*

3.1.1. *Stimuli*

There were 15 different stimulus tokens of the form /s*V*t/. The set of tokens was identical to that used in Experiment 1. A male talker (RF) produced several examples of each token. The tokens were recorded on a Sony TC-FX705 stereo cassette using a Sony ECM-170 condenser microphone while the talker sat in an IAC sound-conditioned booth. The data were then low-pass filtered at 5 kHz and digitized at a 10 kHz sampling rate using the ILS waveform analysis and editing program. One example of each recorded token was selected for editing. For each token, the duration of the initial fricative noise was equalized at 150 m. The portion of the fricative noise removed during waveform editing was never near the end of the noise so that any formant energies relevant to the identity of the following vowel remained. That portion of the waveform corresponding to the voiced portion of the vowel was unmodified and retained the duration of the original production. The duration of the stop consonant closure and the release burst (and following frication) were equalized at 150 ms and 100 ms for all tokens, respectively. The durations and mean formant frequencies of these tokens appear in Table I.

3.1.2. *Procedure*

The experimental procedure utilized was based on that used by Halpern & Darwin (1982). There were four stimulus tokens in each separate experimental trial. The first three tokens were always *sight* while the fourth token was one of the 15 listed above. On each trial, the onset-to-onset interval between (OOI) the first, second, and third tokens was 1000 ms. The OOI between the third and fourth tokens varied from trial to trial. This interval deviation amounted to 0, 3, 6, 9 and 12% of the basic 1000 ms OOI. Since each deviation could be either shorter or longer than the basic OOI by a given percentage, there were a total of 9 different OOIs for the final interval.

For each trial, listeners were required to listen to the four tokens presented in sequence and were required to respond whether the final token occurred "too early" or "too late". Responses of "don't know" or "just right" were not permitted. The experiment was conducted in one session and lasted about 25 minutes. The stimuli were presented to subjects at a comfortable listening level via Boston Acoustics loudspeakers while the subjects were seated in a classroom.

3.1.3. *Subjects*

There were 28 listeners who participated in the experiment to fulfill a course requirement. All were speakers of American English and had no known speech or hearing impairments.

Table I. Mean formant frequencies and vowel durations for stimulus tokens used in Experiments 2 and 3. Formant measurements are in Hz; duration measurements are in ms

Token	F1	F2	F3	Vowel duration	
				Exp. 2	Exp. 3
seat	390	2449	2656	225	261
sit	453	1734	2546	180	264
sate	453–374	1562–2202	2687–2609	222	262
set	640	1640	2624	210	263
sat	687	1546	2624	253	260
sot	703	1249	2593	244	259
sut	656	1296	2562	195	261
sought	609	1109	2577	246	260
sote	468–390	1309–1140	2562–2590	245	259
soot	421	1296	2515	184	263
suit	390	1265	—*	229	262
sight	687–406	1265–2031	2656–2609	267	267
sowt	656–593	1468–1109	2924–2468	267	
soyt	468–421	1031–2608	2374–2390	258	
sert	656	1394	1874	220	

*Formant 3 amplitude for [u] was too low to allow accurate measurement.

3.2 *Results*

The data for each stimulus token were collapsed over listeners and psychometric functions were derived for each token by plotting the number of sequences in which the fourth token was judged as "late" as a function of the OOI deviation. The data were then submitted to probit analysis (Ray, 1982) which fitted a normal ogive to each different function. The means of the fitted distributions were then calculated for each of the 15 stimulus tokens. These means represent an estimate of the onset-to-onset interval which is required for the token to be perceived as isochronous with the first three tokens, i.e. the onset-to-onset interval which would produce an equal number of "too early" and "too late" responses. As in Experiment 1, the data from the monophthongal, nonrhotic vowels indicate that the perceived stress-beat location was affected by vowel duration. In particular, regression analysis of the data from the 11 nonrhotic monophthongs (the same vowels shown in the lower portion of Figure 1) showed that mean isochronous OOIs were significantly related to vowel duration [$r(11) = 0.60, p < 0.03$].[3] Figure 3 shows the probit-determined OOI means plotted against vowel duration (along with the regression line, slope = -0.40). The slope of the effect is the same as that obtained by Smith & Fowler (1984). In addition, there was a significantly greater number of "early" responses to the lax (short) vowels than to the tense (long) vowels [$t(27) = 3.25, p < 0.003$].

When the diphthongs and *sert* are included in the regression analysis, vowel duration ceases to be a significant predictor [$r(14) = 0.05$, ns]. A planned comparison on the number of "early" responses, ordered according to vowel duration, showed a significant linear trend when only the 11 monophthongal vowels are included [$F(1,270) = 9.44$,

[3]Lehiste (1975) demonstrated that listeners tended to underestimate the duration of the last interval of a four-token sequence when all intervals had objectively equal durations, perhaps because listeners normally expect the last interval to be longer (due to pre-boundary lengthening). However, for the purpose of Experiments 2 and 3 it should make no difference whether or not subjects consistently underestimate the last interval, since we are concerned with whether listeners respond differently to longer vs. shorter vowels.

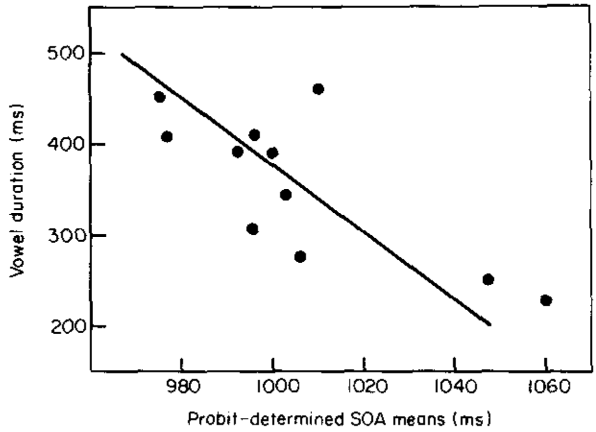

Figure 3. A plot of the probit-determined onset-to-onset interval means by vowel duration in Experiment 2.

$p < 0.002$], but the same planned comparison is nonsignificant when all 15 vowels are included [$F(1,378) = 0.41, p > 0.50$]. The same pattern as that observed in Experiment 1 was thus obtained, namely, that vowel duration does not seem to be related to stress-beat location in diphthongs as it does in monophthongs.

In order to determine whether or not there were strictly phonetic (rather than acoustic) effects related to expected intrinsic vowel duration (Peterson & Lehiste, 1960) a third experiment was conducted which was identical to Experiment 2 except that the vowel durations of all stimuli were equalized. If the results obtained are the same as those in Experiment 2, it may indicate that listeners are being significantly affected by anticipated differences in the vowel durations.

4. Experiment 3

4.1. *Method*

4.1.1. *Stimuli*

There were 12 different stimulus tokens including the 11 nonrhotic monophthongs from Experiment 2 plus the token *sight*. For all tokens, the duration of the signal corresponding to the voiced portion of the vowel was waveform edited to be close to 260 ms. This length was chosen because all original vowel durations were equal to or shorter than this length. Duration equalization was accomplished (to within one glottal pulse) by repeating selected glottal pulses. The vowel durations following waveform editing are shown in Table I. The vowel qualities produced were perceptually very similar to the original stimuli, including the slightly diphthongized /eI/ and /oU/. The vowel in *sight* was unmodified.

4.1.2. *Procedure*

The experimental procedure was the same as that used in Experiment 2.

4.1.3. *Subjects*

There were 21 listeners who participated in the experiment to fulfill a course requirement. All were speakers of American English and had no known speech or hearing impairments. None of these subjects had participated in Experiment 2.

4.2. *Results*

The data for each stimulus token were collapsed over listeners and psychometric functions were derived for each different token. The data were submitted to probit analysis and the means of the fitted distribution were then calculated.

Unlike Experiment 2, the data did not indicate that the perceived stress-beat location was affected by vowel quality variations alone. There was no difference in the number of "early" responses between the tense and lax vowels [$t(20) = 1.05$, $p > 0.30$]. There was also no significant correlation between the Experiment 2 and 3 probit-determined onset-to-onset interval means [$r(11) = -0.34$, $p > 0.15$]. These results demonstrate that the stress-beat location shifts obtained in Experiment 2 were not a function of acoustic variations in the fricative portions of the stimuli since those were unchanged in Experiment 3. The tense-lax distinction is phonetically manifested both in the spectral structure of the syllable nuclei and in their duration. The phonetic quality of the initial part of the vowel may be expected to create an anticipation of appropriate vowel duration. The listeners did not show any evidence of such an anticipation; they reacted to the duration of the whole syllable nucleus. This implies that the basic effect for the nonrhotic monophthongs is phonetic, rather than phonological in nature. We base this conclusion on the fact that the intrinsic duration differences are a characteristic of the English vowel system, rather than a universal phonetic regularity; if English listeners' judgments are not affected by differences in intrinsic duration they are making a phonetic rather than phonological decision.

5. General discussion

We have shown that, in general, the stress-beat location of a stressed monosyllable, for both talkers and listeners, is affected by the duration of the medial vowel. In particular, any increase in vowel duration shifts the stress beat to a later position in the syllable. In addition, this effect seems to be primarily phonetic, rather than phonological in origin since vowel quality variations alone do not produce comparable effects.

In the last decade, then, various studies have shown that the stress beat (or P-center) of a monosyllable (in both perception and production) can be shifted by variation of the duration of initial segments (Fowler, 1979; Fowler & Tassinary, 1981), medial vowels (Marcus, 1981, Smith & Fowler, 1984, in addition to the data presented here) and final segments (Smith & Fowler, 1984; Fox & Lehiste, 1985). Since perceived stress-beat location can be affected by variations in those parts of the syllable that follow the occurrence of the perceived stress beat, it seems clear that the listener reacts to the syllable as a whole, rather than to any single acoustic event. There is reason to be optimistic that an acoustically based model similar to that suggested by Marcus (1981) could be developed to predict the expected location of the stress beat of a monosyllable.

The results from Experiment 1 have shown that the talker's stress-beat locations also seem to be sensitive to the same variables. The acoustic events measured in this study, of course, have their origins in the timing of the speaker's articulations, but it is less clear just what our results mean for an articulatory model of stress-beat location. For example, Fowler (1979, 1983) has suggested that the stress beat (or perceptual center) corresponds to some portion of the vowel's articulatory trajectory and has explained stress-beat shifts produced by initial consonant variations in terms of vowel coarticulation and its subsequent effects upon the relevant acoustic measurements. In order to account for the data presented here one must also assume that such shifts are produced because of coarticulatory effects which result from vowel quality variations. Although perhaps not intuitively sensible, Fowler (personal communication) and her colleagues have collected data (as yet unpublished) which suggest that such coarticulatory effects may indeed be found. However, even if this is the case, it is unclear why the perceptual effect was eliminated when vowel durations were equalized since the same degree of coarticulation was presumably still present in the acoustic signal. Further research needs to be done to coordinate the acoustic and articulatory views of stress-beat location.

This research was supported by Grant 1 RO1 NS21121-01 from NINCDS, National Institutes of Health. A version of this paper was presented to the 110th Meeting of the Acoustical Society of America, November 4–8, 1985, Nashville, TN. We wish to thank Stephen M. Marcus, Vincent J. van Heuven and, especially, Carol Fowler who reviewed this manuscript and made many important suggestions as to revisions. We also would like to thank the statistical consultants at the Instructional and Research Computer Center for their help in analyzing the data. Reprints may be obtained from the first author.

References

Abercrombie, D. (1964) Syllable quantity and enclitics in English. In *In Honour of Daniel Jones* (Abercrombie, D., Fry, D. B., MacCarthy, P. A. D., Scott, N. C., & Trim, J. L. M., editors), pp. 216–22. London: Longmans.

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

Allen, G. (1972*b*) The location of rhythmic stress beats in English: An experimental study, part II, *Language and Speech*, **15**, 179–195.

Bolinger, D. (1965) Pitch accent and sentence rhythm. In *Forms of English: Accent, Morpheme, Order* (T. Abe & T. Kanekiyo, editors), pp. 163. Cambridge, MA: Harvard University Press.

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

Classe, A. (1939) *The Rhythm of English Prose*. Oxford: Blackwell.

Donovan, A. & Darwin, C. (1979) The perceived rhythm of speech, *Proceedings of the ninth international congress of phonetic sciences* 2, 268–274.

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

Fowler, C. A. (1983) Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet, *Journal of Experimental Psychology: General*, **112**, 386–412.

Fowler, C. A. & Tassinary, L. (1981) Natural measurement criteria for speech: The anisochrony illusion. In *Attention and performance, IX* (J. Long & A. Baddeley, editors), pp. 521–536. Hillsdale, NJ: Erlbaum.

Fox, R. A. & Lehiste, I. (1985) The effect of final consonant structure on syllable onset location, *Journal of the Acoustical Society of America*, **77**, S54 (abstract).

Gay, T. (1977) Articulatory movements in VCV sequences, *Journal of the Acoustical Society of America*, **51**, 2018–2024.

Halpern, A. R. & Darwin, C. J. (1982) Duration discrimination in a series of rhythmic events, *Perception and Psychophysics*, **31**, 86–89.

Hoequist, C. E. (1983) The perceptual center and rhythm categories, *Language and Speech*, **26**, 367–376.

Ladefoged, P. (1982) *A Course in Phonetics*, second Edition. New York: Harcourt Brace Jovanovich.

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

Lehiste, I. (1975) The perception of duration within sequences of four intervals. Paper presented at the 8th International Congress of Phonetic Sciences. Leeds, August 21.

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

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

Peterson, G. & Lehiste, I. (1960) Duration of syllable nuclei in English, *Journal of the Acoustical Society of America*, **32**, 589–596.

Rapp, K. (1971) A study of syllable timing. *Quarterly progress Report, Vol. I*, Speech transmission laboratory, University of Stockholm, pp. 14–19.

Ray, A. A. (editor) (1982) *SAS User's Guide: Statistics*, Cary, NC: SAS Institute Inc.
Shen, Y. & Peterson, G. G. (1962) Isochronisms in English, *Studies in Linguistics*, Occasional Papers, No. 9.
Smith, M. R. & Fowler, C. A. (1984) The perceptual centers of alliterative monosyllables, *Journal of the Acoustical Society of America*, **76**, S81 (abstract).
Tuller, B. & Fowler, C. A. (1980) Some articulatory correlates of perceptual isochrony, *Perception and Psychophysics*, **24**, 277–283.