Perceived P-Center Location in English and Japanese*

Robert Allen Fox
Speech & Hearing Science, The Ohio State University

1.0 Introduction

To most speakers/hearers, both linguistically trained (Abercrombie, 1964; Classe, 1939; Pike, 1945) and naive (Donovan & Darwin, 1979; Lehiste, 1972), languages sound rhythmical, that is, the occurrence of spoken elements in time seems to be organized in terms of some potentially specifiable set of principles. Three classes of rhythmic organization have been proposed for languages: stress-timing (e.g., English, German), syllable-timing (e.g., French, Spanish), and mora-timing (Japanese). In stress-timing, for example, the temporal regulation of the spoken utterances should make the intervals between stressed syllables approximately isochronous. However, research directed at the nature of the temporal characteristics of speech, particularly for English (the most studied language in terms of its rhythmical properties), has failed to discover strict regularity or isochrony between acoustically defined intervals--such as stressed syllable onset--in either spontaneous or more constrained utterances (e.g., Classe, 1939; Shen & Peterson, 1962; Bolinger, 1965; Lehiste, 1972; see discussion in Lehiste, 1982).

The perception of rhythmicity in speech also does not seem to arise from the presence of isochronous acoustic onsets of linguistic elements (such as stressed syllables). For example, if listeners hear acoustically isochronous sequences of monosyllables (i.e., equal temporal intervals between the syllables' acoustic onsets) whose initial consonants differ in terms of manner of articulation these sequences will sound irregular. Listeners will hear these sequences as being "regular" only if systematic deviations from acoutic isochrony are introduced (Morton, Marcus, & Frankish, 1976; Fowler, 1979, 1981; Fowler & Tassinary, 1981). Fowler (1979) found that the temporal 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 thus apparent that listeners and talkers are capable of focusing on <u>some</u> aspect of orally produced speech when required either produce speech or to make timing judgments. A question that remains is determining upon what basis listeners/speakers on making their timing judgments. Morton et al. (1979) introduced the term "perceptual center" or "P-center" which was defined as the <u>perceptual moment of occurrence</u> of a monosyllabic token [1]. Regular sequences of speech tokens have, by definition, perceptually isochronous P-center. The P-center thus defined presumably corresponds to the locus of the "stress beat" [2] (Allen 1972; Rapp, 1972). The P-center, however, does not seem to correspond to any commonly measured acoustic event such as the onset of measurable acoustic energy, the onset of the periodic energy of the stressed vowel, or the energy peak (Rapp, 1971; Fowler, 1979; Tuller & Fowler, 1981). Rather, the P-center in stressed syllables corresponds to some event in the signal which can be affected by the duration of the initial consonant (Fowler, 1979), the durations of the medial vowel and final consonants (Marcus, 1981; Smith & Fowler, 1984; Fox & Lehiste, 1985a,b,c), as well as by the addition of unstressed prefixes and/or suffixes (Fox & Lehiste, 1986).

Fowler and her colleagues (e.g., Fowler, 1979; 1983; Fowler & Tassinary, 1981; Tuller & Fowler, 1980; Smith & Fowler, 1984) have suggested that the P-center may correspond to an articulatory event, such as the onset of the vowel. Since coarticulatory phenomena may blend the acoustic characteristics of the vowel with surrounding consonants this articulatory onset may not line up conveniently with commonly used acoustic measurements (such as onset of vocalic periodicity). As Fowler argues, this may produce the situation in which the acoustic measures deviate from isochrony, even in the event of articulatory isochrony. In particular, the articulatory onset of the vowel may occur during the production of the preceding consonant (particularly with segments such as fricatives, see discussion coarticulatory overlap in Fowler, 1983). This hypothesis corresponds well to the findings of experiments which required subjects to mark perceived stress beats in repeated syllable sequences by finger taps (Allen, 1972a,b; Van Katwijk & van den Berg, 1968) or click location manipulations (Eggermont, 1969; Rapp, 1971). When the stimulus syllable began with a stop, the listeners tended to mark the stress beat as occurring at or near the onset of the vowel. However, when the initial consonant was a fricative and longer in duration, the beat was perceived as occurring earlier in relation to the onset of the vowel's periodicity.

The perceived stress beat does not seem to be related to articulatory onset in a simple manner, however. For example, Marcus (1981) demonstrated that increasing the duration of the [t] closure in the token eight--which would presumably not affect the perception of the articulatory onset of the vowel--shifted the perceived location of the token's stress beat. For the purposes of this paper it is sufficient to state that the phonetic structure of the entire word may contribute to the location of the stress beat.

Given that stress-timing is but one possible principle in the organization of speech rhythm, one obvious concern is with the status and/or nature of stress-beat (or P-center) location in languages using different timing principles. In particular, is the P-center a universal phenomenon? If so, is the location of the P-center determined by the same set of acoustic and/or articulatory cues?

The suggestion that the P-center phenomenon was universal in spoken language behavior was made by Hoequist (1983a) who conducted a study examining the P-center effect in the production of English, Spanish, and Japanese monosyllables. Hoequist required sets of subjects to produce a series of rhythmic utterances. Each utterance was composed of 10 alternating monosyllables that differed in terms of their initial consonant (an experimental design similar to Fowler, 1979 and others). The stimulus syllables included a, ma, ba, pa, and sa, although only the pairs a-ba, ma-ba, and pa-sa (in both orders) were used in the test utterances. Subjects uttered these alternating sequences in time to a metronome for practice (no information about rate was given) and in the test condition uttered the sequences without external timing cues. The utterances were analyzed in terms of both the durations of the nine intersyllabic intervals (ISIs) in each sequence and the duration of any portion of the syllable preceding vocalic periodicity.

Hoequist (1983a) compared average difference in duration for adjacent ('different onset') ISIs (e.g., pa-sa vs. sa-pa) with the difference for non-adjacent ('same onset') ISIs (e.g., pa-sa vs. pa-sa). The pattern of results indicated that the P-center came after the acoustic onset of the syllable. Examination of the different-onset ISIs in terms of the onset of vocalic periodicity showed that the P-center came before, although much closer to, the onset of periodicity. Analysis of the duration differences showed a significant effect of Onset Type (same vs. different) but no effect associated with Language (English vs. Japanese vs. Spanish). There was also no significant Language X Onset interaction. Hoequist suggested that the P-center effect was present in all three languages investigated, apparently to the same degree. In general, any speaker who attempted to produce isochronous syllables aligned some point in the token which did not correspond either to the acoustic onset of the syllable or the onset of the periodicity.

The question which this paper poses is whether the perceived location of the P-center or stress beat is also generalizable across distinct language groups. To address this question a perceptual experiment was conducted comparing the responses from a group of functionally monolingual Japanese speakers with a group of monolingual American English speakers. In particular, Smith & Fowler (1984), and Fox & Lehiste (1985a,c) demonstrated that the nature of the final consonant in CVC monosyllables affected the location of the stress beat (or P-center) when subjects were required to produce sequences of monosyllabic tokens in both metronome and non-metronome conditions. The present experiment examines whether analogous syllable-final variations can shift the location of the P-center in monosyllables in a perceptual task, and whether such shifts are the same for both American English and Japanese speakers.

2.0 Method

2.1 Subjects

There were 29 monolingual American English subjects. These subjects were undergraduate students at The Ohio State University who participated to fulfill a course requirement in Speech & Hearing Science. There were 31 native Japanese subjects. These subjects were second-year students in the Domestic Science Department at a women's junior college in Tokyo. The instructions for the Japanese subjects were in Japanese and the test was administered by a native Japanese Professor (Dr. Morio Kohno).

2.2 Stimulus Materials

Eleven stimulus tokens were constructed, each of which had the form [da__]. Ten tokens ended in a coronal consonant and one token ended with the vowel [a]. The stimuli consisted of the following: dah, dot, dodd, doss, dozz, dosh, dotch, dodge, don, doll, and dar. A male talker (RAF, a phonetician) produced several examples of each token in time with a metronome pulse which occurred every 1000 ms. The tokens were recorded with a high-quality cassette recorder (Sony TC-FX705) using a condenser microphone (Sony ECM-170) while the talker sat in a sound-conditioned booth (IAC). The metronome pulse was used as an organizing cue and was not recorded. These productions were then low-pass filtered at 4800 Hz and digitized at a 10 kHz sampling rate using the ILS waveform analysis programs implemented on a PDP 11/23 computer. One example of each token was selected for editing. For each token, all acoustic energy prior to the release of the initial [d] consonant was eliminated and the durations of the medial vowel and final consonant were measured. Final stops were released and their durations were measured from consonant closure to closure release. The overall amplitudes of the tokens were then equalized. The vowel, consonant, and vowel+consonant durations for these 11 tokens appear in Table 1.

Table 1. Acoustic measurements, including medial vowel and final consonant duration, and probit-determined means for listener-perceived isochronous ISIs for each of the 11 stimulus tokens, in ms. (Note, the sonorant consonants [r] and [l] are considered as part of the vowel in the following table, and in the accompanying analyses.)

Token	Vowel Duration	Consonant Duration	Probit-determined ISIs	
			English Subjects	Japanese Subjects
dah	451	0	975	965
dodd	388	144	999	974
dot	248	98	1047	1071
dozz	408	124	996	982
doss	272	237	1006	1020
dosh	303	251	996	1020
dotch	226	269	1060	1077
dodge	342	194	1003	999
don	388	147	993	961
doll	466	0	977	960
dar	404	0	1010	951

2.3. Procedure

The experimental procedure utilized was based on that used by Halpern & Darwin (1982). In each separate experimental trial there were four experimental tokens. The first three tokens were dah while the fourth token was one of the 11 tokens listed above. On each trial the intersyllable interval (the syllable-onset to syllable onset interval) between the first, second, and third tokens was 1000 ms. The ISI between the third and fourth tokens varied from trial to trial. This interval deviation amounted to 0%, 3%, 6%, and 9% of the basic 1000 ms ISI. Since each deviation could be either longer or shorter than the basline, there was a total of 11 different ISIs for the final interval. The deviation increments were based on difference limens estimated by Halpern & Darwin (1982). The presentation order of these sequences were then randomized, example stimulus sequences and fillers were added and the stimului were converted into analog form, filtered at 4800 Hz and recorded on a high-quality stereo cassette recorder (Sony TC-FX705).

For each trial, listeners were required to listen to the four tokens presented in sequence and to respond whether the final token occurred 'too early' or 'too late.' The experiment was conducted in one session which lasted about 25 minutes. This procedure will not determine the absolute location of a token's P-center but rather will allow a determination of each token's relative P-center location using probit analysis of the resulting psychometric function as compared with the other 10 tokens. This procedure will thus allow us to compare whether vowel and final consonant durations affect the relative location of the P-center in both American English and Japanese speakers.

3.0 Results and Discussion

The data for each stimulus token were collapsed over listeners in each of the two language groups and psychometric functions were derived for each token by plotting the number of sequences in which the fourth token was judged 'late' as a function of the variable ISI interval. These data were then submitted to probit analysis (Ray, 1982) which fitted a normal ogive to each different function. Shown in Table 1 are the means of the fitted distributions for each of the 11 stimulus tokens for each of the two language groups. These means represent an estimate of the ISI required between the third and fourth token so that all four tokens are perceived as occurring isochronously. If we assume that subjects are making their judgments on the basis of aligning the P-center of the four stimulus tokens in time, then the longer the estimated mean ISI to produce isochrony, the earlier the location of the P-center in the fourth stimulus token. These data will be further analyzed first by separate language group to determine the best predictor(s) of estimated isochronous ISIs and then together using analysis of variance to determine whether the two different groups produced significantly different responses.

The English data were analyzed using step-wise multiple linear regression analysis with estimated isochronous ISI values as the dependent variable and vowel duration, consonant duration, and vowel+consonant duration as the independent variables. Regression analysis showed that the ISI values were significantly predicted by vowel duration (r=0.844, $F(1,9)$=22.2, p<.002). The slope of the regression line was -0.27. This suggests that as vowel duration increases by 100 ms, the ISI duration needed to produce a perceptually isochronous sequence decreases by 27 ms. This value is only slightly smaller than as those obtained by Smith & Fowler (1984) and by Fox &

Lehiste (1985b,c) who examined the effect of medial vowel duration on P-center location. The present regression results support the conclusion that as the vowel duration increases--as a function of the final consonant--the P-center location moves to a later point in the token. ISI means were also significantly related to final consonant duration ($r(11)=0.62$, $p<.02$), but final consonant duration is also significantly related to vowel duration ($r(11)=-0.90$, $p<.001$). If the contribution of vowel duration is partialled out from the consonant duration variable, consonant duration is only marginally related to mean ISI ($t(10)=-2.2$, $p<.067$).

The estimated isochronous ISI data for the Japanese subjects were also analyzed using step-wise multiple linear regression. Analysis showed that estimated ISIs were significantly predicted by vowel duration ($r=0.93$, $F(1,9)=57.4$, $p<.001$). The slope of the regression line was -0.50. This suggests that as vowel duration increases 100 ms, the ISI duration needed to produce an isochronous sequence decreases by 50 ms. This value is greater than that obtained both for the American English group and by Smith & Fowler (1984) and Fox & Lehiste (1985b,c). Estimated ISIs were also significantly related to final consonant duration ($r(11)=0.81$, $p<.001$), but when the contribution of vowel duration is partialled out, consonant duration is not even a marginally significant predictor of mean ISI ($t(10)=-0.56$, $p>.59$). The basic pattern of results is the same between the two language groups, namely, as vowel duration increases, the P-center location moves to a later point in the token. The similarity between the two groups is best illustrated by the fact that the estimated ISIs between the English and Japanese groups are significantly correlated ($r(11)=0.85$, $p<.001$) although there seems to be some difference between the groups in terms of the contribution of final consonant duration to the estimated ISI means.

Since the estimated ISI values have been calculated on the basis of responses collapsed over subjects within each of the two language groups, they cannot be easily used to determine differences between the two groups. To examine such differences, the number of 'late' responses for each subject for each stimulus token were calculated--that is, the responses were collapsed over the nine experiment ISI durations. The more 'late' responses a token receives, overall, the earlier in the token the P-center occurs. To balance the number of subjects within each language group the responses from two Japanese subjects were not included. The two subjects chosen had participated in a rhythmic production test (utilizing Japanese stimuli only) prior to the perceptual test. These responses were then submitted to a mixed-design, repeated-measures analysis of variance (ANOVA) with the factors Stimulus Token and Language [3]. The cell means for number of 'late' responses in each language for each stimulus token appear in Table 2. The ANOVA showed significant main effects of both Stimulus Token ($F(10,280)=21.08$, $p<.001$) and Language ($F(1,28)=4.68$, $p<.05$). In addition, there was a significant Stimulus Token x Language interaction ($F(10,280)=2.21$, $p<.05$).

First and as expected, these results demonstrate that the number of 'late' responses given to a stimulus token seems to vary as a function of its final consonant/medial vowel durations. Second, these results show that there is a slight difference in the mean number of 'late' responses overall between the two language groups. Third, these results show that the two language groups tend to have a different pattern of 'late' responses across different stimulus. The difference is small, but with this number of subjects, significant. This difference is very likely related to the differential effect of final consonant duration on the perception of isochronous sequences in the two language groups. It is tempting to speculate that the response differences between the two language groups are related to the differences between English and Japanese in phonetically acceptable syllable structures--particularly with regard to syllable-final consonants. However, such speculation would obscure the more interesting discovery that the perceptual responses of English and Japanese subjects are very similar, despite phonological-phonetic and/or timing differences between the languages.

Table 2. Cell means for number of 'late' responses by language groups and stimulus token.

Stimulus Token	Subject Group	
	English	Japanese
dah	4.41	3.55
dodd	4.07	3.72
dot	5.90	6.24
dozz	4.41	3.76
doss	4.69	5.00
dosh	4.41	5.10
dotch	5.97	5.79
dodge	4.59	4.59
don	4.03	3.10
doll	3.48	3.31
dar	4.31	2.90

In summary, the data support the hypothesis by Hoequist (1983a) that the P-center effect is a universal phenomenon. In both groups the estimated value of the ISI between the third and fourth tokens required to produce an isochronous sequence was significantly related to the vowel duration of the fourth token. These data also show that there are some differences between the perceptual responses of the American English group and the Japanese group. In particular, the P-center locations estimated for the Japanese subjects do not seem to have been significantly affected by final consonant duration; only medial vowel duration. The P-center locations for the American English subjects were significantly affected by medial vowel duration, and additionally affected by final consonant duration at at least a marginally significant level.

The results presented here complement those presented by Hoequist (1983) and support the contention that the P-center phenomenon might be found in speakers/hearers of all languages, but many questions remain.

How might the P-center effect operate in the production or perception of Japanese stimuli having either a light (one-mora) or a heavy (two-mora) syllable. Is the P-center related only to single syllable production/perception or does it also relate to a language's more global rhythmic organization? It thus goes without saying that much work remains to be done in understanding the organization of timing in both the perception and timing of speech. However, in this volume dedicated to Ilse Lehiste, we should take the space to briefly acknowledge the numerous contributions which Ilse has made to field in the areas of speech timing and prosodic phenomena in particular, and to the understanding of linguistic phenomena in general. Ilse has provided many important experimental and theoretical contributions, of course, but an even greater contribution is her insistence upon scientific rigor in the study of language behavior. She continues to provide our field with an example of the fertile scientific mind at work, and remains a scholar who is warmly appreciated by her colleagues who will value their less frequent interactions with her after her retirement.

Notes

*This research was supported by Grant 1 R01 NS21121-01 from NINCDS, National Institutes of Health. Ilse Lehiste is my co-investor on this grant and as such should be considered the 'covert' co-author of this paper. Of course all the mistakes are mine. I wish to thank Professor Morio Kohno of the Kobe City University of Foreign Studies who collected the Japanese data for me.

1. Morton et al. (1975) used spoken digits as their stimuli, and their precise definition of P-center more properly refers to the moment of occurrence of a spoken digit.

2. The terms P-center and stress beat, as used in the relevant experimental literature, seem to refer to the same linguistic phenomenon and the reader should assume that these terms are interchangeable in this paper.

3. Technically this analysis may violate one assumption underlying the use of parametric statistic analytic techniques. In particular, use of ANOVA assumes that the data analyzed are interval in nature. One could argue that since the calculated responses can only assume the values from 0-9, that they better represent ordinal level data. This type of violation is probably not very significant and actually is actually relatively common in psychological research. However, following Hays (1973), I will here caution that the ANOVA results may not accurately reflect the magnitude of the differences between language and/or stimulus tokens, but should tell us something about the quality differences between them.

References

Abercrombie, David (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), 216-22. London: Longmans.

Allen, George (1972a). The location of rhythmic stress beats in English: An experimental study, Part I. Language and Speech 15.72-100.

Allen, George (1972b). The location of rhythmic stress beats in English: An experimental study, Part II. Language and Speech 15.179-195.

Bolinger, Dwight (1965). Pitch accent and sentence rhythm. In Forms of English: Accent, morpheme, order (I Abe & T. Kanekiyo, editors), 163. Cambridge, MA: Harvard University Press.

Carney, P. & Moll, K.A. (1971). A cineflourographic investigation of fricative consonant-vowel coarticulation. Phonetica 23.193-202.

Classe, Andre (1939). The rhythm of English prose, Oxford: Blackwell.

Compton, Andrew (1980). Timing patterns in French. Phonetica 37.205-234.

Donovan, A. & Darwin, C. (1979). The perceived rhythm of speech. Proceedings of the ninth international congress of phonetic sciences 2.268-74.

Eggermont, J. (1969). Location of the syllable beat in routine scansion recitations of a Dutch poem. IPO Annual Progress Report (Institute for Perception Research, Eindhoven) 4.60-4.

Fowler, Carol A. (1979). "Perceptual centers" in speech production and speech perception. Perception & Psychophysics 25.375-88.

Fowler, Carol 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, Carol A. & Tassinary, Louis G. (1981). Natural measurement criteria for speech: The anisochrony illusion. In Attention and performance, IX (J. Long and A. Baddeley, editors.), 521-536. Hillsdale, N.J.: Erlbaum.

Fox, Robert A. & Lehiste, Ilse (1985a). The effect of final consonant structure on syllable onset location. Journal of the Acoustical Society of America 77.S54 (Abstract).

Fox, Robert A. & Lehiste, Ilse (1985b). The effect of vowel quality variations on stress-beat locations. Journal of the Acoustical Society of America, 78.S21 (Abstract).

Fox, Robert A. & Lehiste, Ilse (1985c). Syllable structure and its effects upon stress-beat location. Presented to the 1985 National Convention of the American Speech-Language-Hearing Association, Washington DC, 21-25 November.

Fox, Robert A. & Lehiste, I. (1986). The effect of unstressed affixes on stress-beat location in English. In Proceedings of the Montreal Symposium on Speech Recognition, 36-37.

Halpern, A.R. & Darwin, C.J. (1982). Duration discrimination in a series of rhythmic events. Perception & Psychophysics 31.86-9.

Hoequist, Charles E. (1983a). The perceptual center and rhythm categories. Language and Speech 26.367-76.

Hoequist, Charles E. (1983b). Syllable duration in stress-, syllable-, and mora-timed languages. Phonetica 40.203-37.

Lehiste, Ilse (1972). Rhythmic units and syntactic units in production and perception. The Journal of the Acoustical Society of America 54.1228-34.

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

Lehiste, Ilse (1982). Isochrony Reconsidered. Journal of Phonetics 5.253-63.

Katwijk, A. van, & Berg, B.L.S. van der (1968). Perceptual and motoric synchronization with syllable beats. IPO Annural Progress Report (Institute for Perception Research, Eindhoven) 3.35-9.

Marcus, Stephen (1981). Acoustic determinants of perceptual centers (P-center) location. Perception & Psychophysics 30.247-56.

Morton, John, Marcus, Stephen & Frankish, C. (1976). Perceptual centers (P-centers). Psychological Review 83.405-08.

Rapp, K. (1971). A study of syllable timing. Quarterly progress report, Vol. I, Speech transmission laboratory, University of Stockholm, 14-19.

Ray, A.A. (ed.), (1982). SAS user's guide: Statistics, Cary, NC: SAS Institute Inc.

Shen, Y. & Peterson, Gordon (1962). Isochronisms in English Studies in Linguistics, Occasional Papers, No. 9.

Smith, Mary R. & Fowler, Carol A. (1984). The perceptual centers of alliterative monosyllables. Journal of the Acoustical Society of America, 76.S81 (Abstract).

Tuller, Betty & Fowler, Carol A. (1980). Some articulatory correlates of perceptual isochrony. Perceptions & Psychophysics 24.277-83.