Perception & Psychophysics
1981, 30 (3), 247-256

Acoustic determinants of perceptual center (P-center) location

STEPHEN MICHAEL MARCUS
Institute for Perception Research (IPO), Eindhoven, The Netherlands

Morton, Marcus, and Frankish (1976) defined "perceptual center," or "P-center," as a neutral term to describe that which is regular in a perceptually regular sequence of speech sounds. This paper describes a paradigm for the determination of P-center location and the effect of various acoustic parameters on empirically determined P-center locations. It is shown that P-center location is affected by both initial consonant duration and, secondarily, subsequent vowel and consonant duration. A simple two-parameter model involving the duration of the whole stimulus is developed and gives good performance in predicting P-center location. The application of this model to continuous speech is demonstrated. It is suggested that there is little value in attempting to determine any single acoustic or articulatory correlate of P-center location, or in attempting to define P-center location absolutely in time. Rather, these results indicate that P-centers are a property of the whole stimulus and reflect properties of both the production and perception of speech.

Morton, Marcus, and Frankish (1976) observed that although a sequence of digits may be produced by a human speaker such that they are perceived as isochronous, if tokens of each naturally spoken digit are presented with isochronous acoustic onsets, they are perceived as occurring irregularly. The perceptual center, or P-center, of each digit was defined as its *perceptual moment of occurrence*. Regular sequences thus have, by definition, perceptually isochronous P-centers.

We were initially interested in producing isochronous digit lists for memory experiments (see Morton, Marcus, & Ottley, in press). We found that although naturally spoken lists were perceived as regular, our attempt at automating this process using stored tokens of each digit was clearly perceptually irregular and unacceptable. Rather than beginning with any hypotheses about what points need to be regular in a regular sequence, we defined P-centers as such points, whatever their acoustic or articulatory correlates might be. In order to experimentally investigate these P-center locations, we made the simplifying assumption that, at least for isolated concatenated stimuli such as we were using, P-center location for a given stimulus is independent of the nature of adjacent stimuli. That is, if we know the relative temporal alignment of an alternating sequence of two tokens, say "one" and "two," which results in perceived isochrony, and similarly for the same token of "two" and one of "three," we can predict the correct timing for an alternating sequence of "one" and "three." This forms our null hypothesis of no context dependency, and will be termed the *independence hypothesis*. A paradigm was designed which both uses and tests this hypothesis in determining relative P-center alignment.

GENERAL METHOD

A token of each stimulus was sampled at 20 kHz, 8-bit samples, using the Applied Psychology Unit sample and display program, BARD, running on a CTL Modular One computer. After visual and auditory examination to determine start and end points, tokens were stored on disk.

During an experiment, stimulus pairs were presented in a fixed randomized order, controlled by a punched tape. For each experimental trial, a control program recovered pairs of stimuli from disk and presented one with onsets at regular temporal intervals, 2T, where T is the interstimulus interval (ISI) from onset to onset. After three presentations of this "fixed" stimulus, the other "movable" stimulus was added. Its onset time could be advanced or retarded relative to the mean ISI by rotating a knob (see Figure 1). The absolute positional location of the knob was randomized from trial to trial, and subjects had no feedback on the adjustments they made.

Subjects were tested individually and were instructed to adjust the movable stimulus until the sequence was perceived as isochronous. Examples of soldiers receiving marching orders ("left-right-left . . ."), together with a demonstration of the experimental setup itself, were given. The subject responded by pressing a button when she or he was satisfied with the regularity of the sequence, and stimulus presentation stopped immediately. The final chosen offset, t, was noted, and after a pause of 3 sec the following trial began. Subjects were encouraged to respond rapidly but accurately, and were informed at the beginning of the total number of trials in the experiment.

The experiments described here were carried out at the Applied Psychology Unit, Cambridge, England, while the author was a PhD student supported by a Medical Research Council fellowship. Thanks are due to John Morton for his inspiration and guidance, even from the confines of a Paris hospital bed, to Ray Bloomfield for the construction of specialized computer hardware, and to Ian Nimmo-Smith for statistical assistance. Finally, the comments and criticism of friends and colleagues in Cambridge and Eindhoven, and of Carol Fowler and an anonymous reviewer, have all contributed in some positive way to the ideas developed and presented here.

0031-5117/81/080247-10$01.25/0

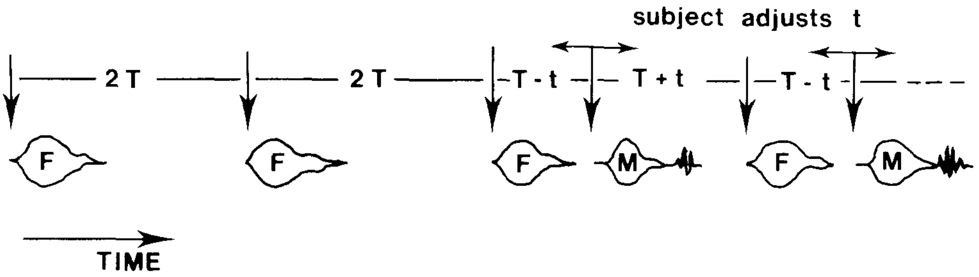

Figure 1. The stimulus sequence in on-line experiments. Stimulus F occurs at fixed temporal intervals. The relative timing of Stimulus M is controlled by the subjects.

An experimental run consisted of all pairings of all stimuli, excluding that of a stimulus with itself, as this was found to rapidly give rise to verbal transformation effects (Warren & Gregory, 1958), the stimuli often becoming perceived as nonspeech. For a set of nine digits there were thus 72 trials, and a typical run lasted 35 to 60 min, depending on the subject.

SOLVING FOR P-CENTERS

The result of a run with N stimuli is an N×N matrix with all nondiagonal cells filled. Let us term this matrix [t], and the observed offset of the i^th^ stimulus paired with the j^th^, t_{ij}. The independence hypothesis may be restated in the form that each stimulus may be represented by a unique P-center, located at time $p_i + k$ from the onset of the i^th^ stimulus, where k is an arbitrary constant fixed for all stimuli with which we are concerned. Stimulus i and stimulus j should then be perceived as regular when their P-centers are regular, and their onset asynchrony is $(p_i + k) - (p_j + k) = p_i - p_j$. The actual observed difference in onsets, t_{ij}, will comprise three components: (1) the relative difference in P-center location between the two stimuli, $p_i - p_j$; (2) an order effect bias, resulting from a tendency to rotate the knob further clockwise or anticlockwise than the desired position; and (3) a random error.

Let us make the simplifying assumption that order effect bias and random error are independent of i and j and may be represented by independent random variables. Since variability in order effect is confounded with the random error itself, we may conveniently combine these into a single error term, e_{ij}, with mean k and variance σ_e^2 independent of i and j. Thus,

$$t_{ij} = p_i - p_j + e_{ij}. \tag{1}$$

Since we know the values of both t_{ij} and t_{ji}, we may compute estimates, $\hat{k}$ and s_e, of both k and σ_e for each run. Subjects' instructions effectively request them to adopt a zero value of k, so we should expect $\hat{k}$ to be distributed normally around zero. The value s_e^2 gives a measure of the subject's variance in reproducing his own chosen set of offsets and has $\omega = (N-1)/2 - 1$ degrees of freedom.

Let

$$D_{ij} = (t_{ij} - t_{ji})/2 \tag{2}$$

and let

$$R^2 = \sum_{i=1}^{N} \sum_{j=1}^{i-1} (D_{ij} - p_i + p_j)^2. \tag{3}$$

A least squares P-center fit is the set of $\{p_i\}$ which minimizes R^2, and these may be determined by solving the set of simultaneous differential equations:

$$\frac{\partial}{p_i} R^2 = 0 \forall i. \tag{4}$$

These N equations determine $\{p_i\}$ except for an arbitrary constant, which was chosen such that $\Sigma p_i = 0$.

The minimized value of R^2 is the total square error between the data $\{t_{ij}\}$ and the P-center fit, $\{p_i\}$. It has $(\omega - N)$ free data points and so the residual variance per data point between the data and its corresponding least squares P-center fit is $R^2/(\omega - N)$ on $(\omega - N)$ degrees of freedom. It may be compared directly with s_e^2, and the significance of the difference in accuracy of the P-center fit and the subject's own replication of his adjustments estimated as the F-ratio $[R^2/(\omega - N)]/s_e^2$ on $(\omega - N, \omega)$ degrees of freedom. The independence hypothesis is the null hypothesis that the residual variance is not greater than s_e^2, and thus that this F-ratio does not differ from 1.00.

CENTRE, a FORTRAN subroutine for the solution of Equation 4 in the general case in which some entries in $\{t_{ij}\}$ may be either vacant or the average of a variable number of observations, is available from the author.

EXPERIMENT 1

The purpose of this experiment was to test the independence hypothesis and the paradigm described above, to look for individual differences in P-center location, and also, for practical purposes, to determine estimates of P-centers for a set of spoken digits.

Method

Stimuli. The stimuli were tokens of the nine English digits "one" through "nine" uttered in isolation by a female voice by a native speaker of British English. The digits ranged in length from 310 to 460 msec. They were sampled and stored as described under General Method.

Seventy-two pairs were presented in a predetermined random order.

Subjects. Four female members of the Applied Psychology Unit subject panel served as subjects for three experimental runs each. They were paid for their participation. The three runs were preceded by one practice run to accustom the subjects to the experimental setup, and each experimental run was preceded by 10 practice trials. There was an interval of 1 week between successive runs on the same subject.

The subjects were seated in a sound-damped booth, and stimuli were presented over Telephonics TDH-39 headphones with circumaural cushions at a comfortable listening level.

Results

Table 1 gives relative P-center locations for each run, together with residual variance, F-ratio, and order effect. A mean solution for all 12 runs is also given. In all cases, the F-ratios do not show significant deviation from the independence hypothesis ($p > .05$).

An analysis of variance cannot be directly performed, since the nine values for each run involve an arbitrary constant, here chosen such that the nine values total zero. The nine-dimensional data were therefore transformed by orthogonal projection into eight-dimensional space to account for this arbitrary constant. Then a multivariant analysis of variance procedure (Winer, 1971, pp. 232-240) was applied to obtain Wilk's Λ-statistic, which with Rao's transformation gave a quasi-F-ratio of 5.52 (df = 24,3). This is not significant ($p > .05$), and therefore the data provide no evidence to support between-subject differences in P-center location.

Discussion

This experiment shows the independence hypothesis to be a generally valid and useful empirical construct. The results have additionally provided no evidence for individual differences in relative P-center location, and this is as we should hope, since otherwise it would generally be impossible to produce stimulus sequences which would be perceived as regular by a whole group of listeners. Not only were there no significant differences in this experiment, but good consistency was also found between these results and a number of runs on the author and colleagues at the Applied Psychology Unit. Since variance was considerably smaller for colleagues than for housewives, and this resulted in stable P-centers being obtained with only a few runs, the author served as subject in the remaining experiments. During the experimental run, the only basis for a subject's responses is the perceived timing of the stimuli. The subject has no feedback on the adjustments they are making and no idea of, or control over, their absolute magnitude, and thus the only way he or she can produce consistent responses is by carrying out the task itself.

AN ACOUSTIC CORRELATE OF P-CENTER LOCATION

Morton et al. (1976) did not attempt to locate a single acoustic correlate of P-center location in the acoustic waveform. We did note, however, that relative P-center alignment for the set of digits "one" to "nine" illustrates a number of properties of P-center alignment. First, P-centers do not appear to correspond to any single clearly identifiable acoustic event, such as measured vowel onset, peak waveform

Table 1
Results of Experiment 1

	Stimulus											
Subject	1	2	3	4	5	6	7	8	9	Residual	F(27,35)	Order
1.1	−34	−19	4	9	−36	69	46	−48	9	4368	1.46	−24.5
1.2	−60	−10	−10	5	−17	34	66	−45	36	2511	1.01	− 2.4
1.3	−38	−12	− 3	9	−18	42	61	−27	−15	1415	.96	− 1.5
2.1	2	−12	−11	25	11	36	20	−50	−21	4799	1.54	−11.3
2.2	− 5	− 8	1	− 4	−11	48	18	−17	−22	4260	1.65	− 6.5
2.3	−21	−17	− 3	− 9	−16	47	26	21	−28	4436	1.04	− 7.6
3.1	−28	−31	11	−11	−12	50	15	−32	38	5403	.90	41.5
3.2	−33	−42	15	5	1	40	30	−25	9	4591	.58	14.5
3.3	4	−22	15	−23	− 7	41	39	−55	8	6160	.87	− 8.2
4.1	−29	−22	−25	13	−15	65	65	−47	− 5	2878	1.21	−13.8
4.2	−39	−12	15	− 9	−28	56	61	−30	−15	1605	1.08	− 9.8
4.3	−14	−27	7	− 8	−22	54	62	−66	14	3319	1.45	− 1.7
Mean	−25	−19	1	0	−14	48	42	−35	0			

Note—P-center values are in milliseconds. "Residual" is the residual variance of the P-center fit. The F ratio is the ratio of residual/cell variance and is in all cases not significant (p > .05). "Order" gives the order-effect bias in milliseconds.

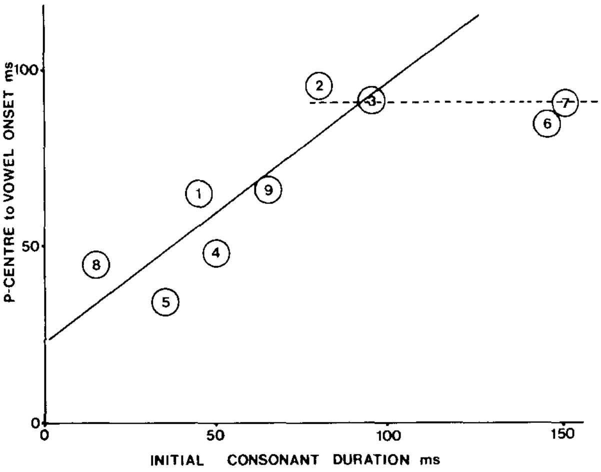

Figure 2. The relationship between relative P-center location and initial consonant duration for the set of digits used in Experiment 1.

intensity, or, of course, acoustic onset. Second, initial consonant duration appears to have a major effect on P-center location; the shorter the initial consonant in a digit, the longer the duration required between the onset of that digit and the onset of the preceding digit in a perceptually isochronous list. Conversely, the onset of a digit with a long initial consonant would need to be presented at a relatively shorter interval following the preceding digit.

Using a different paradigm, one in which subjects were required to produce nonsense words in time with an audible click, Rapp-Holmgren (Note 2) showed that the stress beat in a spoken word precedes its vowel onset by a duration positively correlated with the preceding consonant duration. She found a very high correlation using phonetically controlled nonsense stimuli in which only the consonant itself was varied. In yet another paradigm, one in which subjects tapped to the "beat" of a particular syllable, Allen (1972) looked for and found a similar, but much lower, correlation, this time with words of varying phonetic composition, in continuous speech. Figure 2, taken from Marcus (Note 1), illustrates the same correlation for the P-center-aligned digits in Morton, Marcus, and Frankish's Figure 1, based on the mean solution of Experiments 1 and the runs on APU colleagues. For all digits except "six" and "seven," there is an excellent straight-line fit with a correlation of .87 and a slope of .75.

It was initially supposed that "six" and "seven" might behave differently because of their initial fricative, and a dotted horizontal line representing constant time from P-center to vowel onset has been drawn giving a good fit to "six," "seven," "two," and "three." However, such a relation was not found here for "four" and "five" or other sets of stimuli (Marcus, 1976, Note 1) or in Rapp's or Allen's data. This led to a sequence of experiments investigating the effect of systematic modifications of the acoustic waveform on P-center location.

EXPERIMENT 2

It was first supposed that the starting points of the digitized waveforms for "six" and "seven" stored on the computer might contain initial irrelevant background noise.

Method

Stimuli. Additional versions of "seven" were constructed by progressively deleting sections of the acoustic waveform, beginning at the onset and proceeding in 30-msec steps. This resulted in a series of six stimuli, A, B, C, D, E, and F, whose endpoints were perceived as "seven" (the original stimulus) and "Devon" (in which all frication had been removed), some of the intermediate stimuli having the natural rise of /s/ replaced by an abrupt onset (Figure 3).

In addition to these stimuli, the digits 1, 2, 4, 5, 8, and 9 were included. In two parallel series of experiments, these stimuli were presented together with stimuli A, C, E or B, D, F. In each series, all pairs were presented except that of a stimulus with itself, or, in the case of "seven," with a modified version of itself. The author served as subject in three runs on each series, which were sufficient to produce stable and reliable P-centers.

Results

Table 2 gives the mean results for each subpart of the experiment, with the arbitrary constant in the solution adjusted such that the six digits 1, 2, 4, 5, 8, and 9 which appeared in both series are optimally aligned. Figure 4 presents the same information for

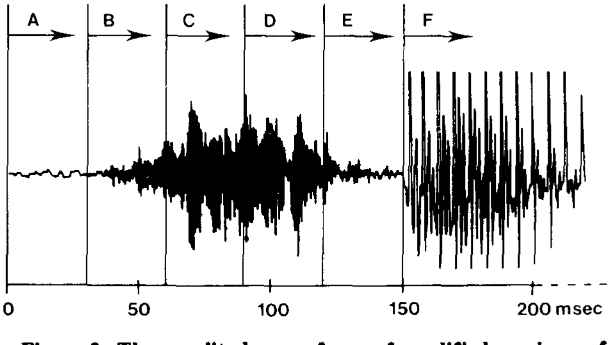

Figure 3. The amplitude waveform of modified versions of "seven" used as stimuli in Experiment 2.

Table 2
Results of Experiment 2

	Stimulus											
	1	2	4	5	8	9	A	B	C	D	E	F
Part 1	7	−20	42	4	−31	3	55		10		−24	
Part 2	7	−21	41	5	−32	1		29		−3		−37

Note—Mean P-center values in milliseconds are given for each subpart of the experiment, with the arbitrary constant in the solution adjusted to give an optimal match of the common stimuli.

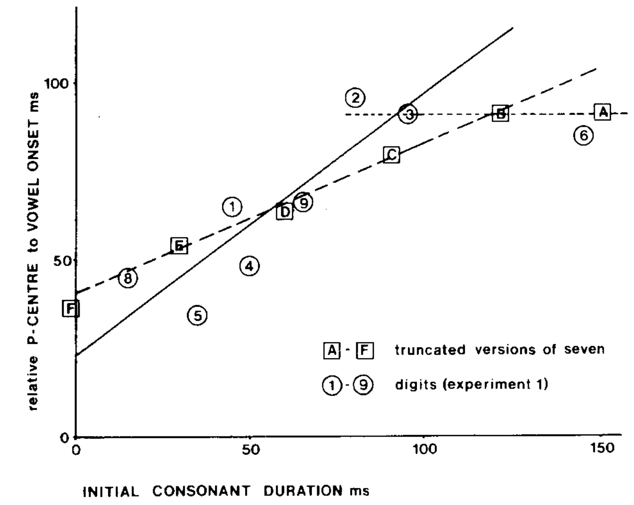

Figure 4. The results of Experiment 2 superimposed on Figure 2. Note that Stimulus A is identical to "seven."

the six modified versions of "seven" as is presented for all the digits in Figure 2. Only for the first two stimuli (A and B) is there no change in P-center location relative to vowel onset; thereafter, the results are well fitted by a straight line with a slope of .45.

Discussion

Despite large differences in the amount of stimulus energy removed in each "bite," and abrupt and unnatural onsets and distinct phonemic changes for some stimuli, a smooth and continuous change in P-center location resulted. Only the removal of the first 30 msec of "seven" produced no change in P-center location and could thus be attributed to onset sampling error. An alternative account for the discrepancy between "six" and "seven" and the other digits was thus sought.

EXPERIMENT 3

The most obvious parameter to investigate in addition to prevocalic consonant duration is the duration of the vowel itself. In this experiment, computer speech-editing techniques were used to pitch-synchronously extend vowel duration in a number of stimuli.

Method

The syllables /bæ, dæ, gæ, pæ, tæ, kæ/ were naturally spoken by the author and sampled and stored on disk. An additional version of each of the first three stimuli was produced by pitch-synchronously duplicating consecutive pitch periods in the stored waveform. Periods were manually selected on the computer display. The vowel portions of /bæ, dæ, *and* gæ/ were extended by 59, 58, and 61 msec, respectively. Extensions were made well after the formant transitions into the vowel were complete. These new stimuli will be denoted /bæ+, dæ+, and gæ+/.

There were three runs at each of two mean ISIs, 650 and 950 msec.

Results

Table 3 gives the results of each run. There are no consistent differences between the two ISIs, though the longer is typified by a higher residual variance. The change in vowel duration resulted in P-center shifts of 21, 17, and 21 msec, or 36%, 29%, and 34%, for /bæ+, dæ+, *and* gæ+/, respectively. The direction of the shift produced by vowel extension was towards the end of the word.

Discussion

P-center location was indeed found to vary with vowel duration, P-centers shifting by around a third of the change in vowel duration, longer vowels resulting in later P-centers, and vice versa. P-centers are thus a function not just of any one single stimulus event, such as, say, the large change in energy conventionally taken as vowel onset, but a property of much of the stimulus.

EXPERIMENT 4

This experiment aimed to take the results of the last further, by examining the effect of word-final acoustics on P-center location. Two parameters were varied: timing and acoustic energy.

Method

The digit "eight" from Experiment 1 was modified in two ways. In the first set of experiments, alternative versions were produced

Table 3
Results of Experiment 3

	Stimulus											
Run	bæ	dæ	gæ	bæ+	dæ+	gæ+	pæ	tæ	kæ	Residual	F(27,35)	Order
1	−29	−21	− 8	−3	0	10	7	29	15	464	.83	2.3
2	−22	−15	−15	3	− 3	9	8	34	2	616	1.80	.7
3	−18	−13	−13	−4	− 1	8	5	28	9	365	1.00	2.8
4	−26	−20	−17	−5	− 3	15	8	41	9	1241	.92	9.6
5	−33	−37	−11	5	3	6	15	44	8	629	.76	10.6
6	−11	−19	−11	1	−13	3	14	41	−5	1137	1.19	1.3
Mean	−23	−20	−12	−1	− 3	8	9	36	6			

Note—The first three runs were at an ISI of 650 msec, and Runs 4-6 were at an ISI of 950 msec.

in which the closure prior to the final /t/ release was extended by 30 msec and also shortened by the same duration. Since the closure is not totally silent, extension was produced by duplicating a portion of noise prior to the release. These will be termed "8+" and "8–," respectively. In the second set, timing remained unaltered, but two versions were produced with the released /t/ burst amplified by 4.5 and 9 dB, respectively; these will be termed "8*" and "8**." Subjectively, the temporal modifications were hardly detectable, while the amplitude changes were dramatic.

Each pair of modified stimuli was presented together with the original "eight," and also 1, 2, 3, 4, 5, and 9.

There were five runs on each set of stimuli. In each run, all 66 possible pairings, excluding versions of "eight" with themselves were presented at an ISI of 650 msec.

Results

Tables 4a and 4b give the results for each set of stimuli. 8+ and 8– showed mean shifts of +13 and –9 msec, or 43% and 30% of the change in closure duration, respectively. In contrast, the amplitude modifications are typified by the results of the first run, and overall a mean shift of only –1 and +2 msec was found for 8* and 8**.

Discussion

The previous experiment showed that P-centers are determined not by a single acoustic event, but by events occurring over a considerable span of the time course of the stimulus. This is even more clearly demonstrated in this experiment, in which the final consonant of CVC stimuli was modified in duration. Even when this modification was a barely noticeable 30 msec in the prerelease closure duration in the /t/ of "eight," as large a proportional change in P-center location resulted as from changes in nuclear vowel duration. In contrast, much more perceptible changes in /t/ burst amplitude (+9 dB) produced neither a significant nor a noticeable change in P-center location. Thus, P-centers seem to be critically determined by the temporal makeup of a stimulus, while being invariant of considerable changes in stimulus energy.

AN ACOUSTIC MODEL

These experiments led to the development of an acoustic model of P-center location in isolated monosyllables. Such a monosyllable may be schematically represented by C_1VC_2 in Figure 5, where C_1 and C_2 are single consonants, consonant clusters, or null and V is a vowel or diphthong. The durations x and y represent initial consonant or consonant cluster duration and vowel plus final consonant duration, respectively. Onsets and offsets were defined as arbitrary points at which stimulus energy exceeds a certain criterion and vowel onset by the point at which, visibly and audibly, the acoustic consequences of the vowel predominate over those of the preceding consonant. The model is:

$$P = \alpha x + \beta y + k, \tag{5}$$

where P is P-center location relative to stimulus onset, α and β are parameters of the model, and k is an arbitrary constant representing the fact that we

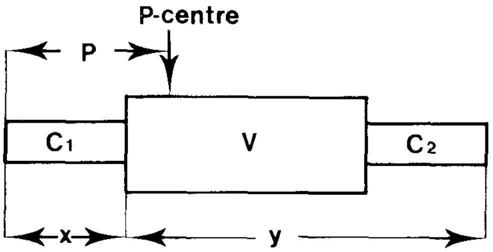

Figure 5. Schematically illustrating the parameters of the P-center model for a monosyllable C_1VC_2.

Table 4
Results of Experiment 4

Run	Stimulus									Residual	F(24,32)	Order
	1	2	3	4	5	9	8	8_1	8_2			
					(a) Effects of Changes of Closure Duration							
1	16	– 8	1	46	11	11	**–24**	–37	–16	173	.97	–2.0
2	12	–14	0	47	15	10	**–28**	–35	– 8	217	1.17	1.1
3	19	–13	2	45	7	8	**–22**	–30	–15	193	1.13	–1.2
4	16	–12	–1	49	12	11	**–30**	–34	–11	186	1.29	4.2
5	15	–16	0	46	17	7	**–23**	–33	–12	98	.88	.9
Mean	15	–13	0	46	12	9	–25	–34	–12			
					(b) Effects of Changes in /t/ Burst Amplitude							
1	21	–18	4	49	7	11	–25	–25	–25	178	1.39	2.2
2	24	–15	0	47	11	6	–28	–28	–16	229	1.04	2.8
3	15	–17	–2	53	16	11	–22	–26	–27	78	.63	1.9
4	17	–18	–3	54	9	13	–28	–20	–24	112	.92	– .7
5	13	–16	–2	55	15	6	–22	–28	–20	102	1.14	2.4
Mean	18	–17	–1	52	12	10	–25	–26	–23			

Note–8_1 and 8_2 = 8– and 8+, respectively, for closure duration and 8 and 8**, respectively, for burst amplitude.*

are only determining *relative* P-center location of stimuli to one another.

Rapp's model is a special case of Equation 5 with $\beta = 0$, that is, no effect of vowel duration. She found $\alpha = .7$, and her stimuli showed only small variations in vowel and final consonant duration, these being negatively correlated with initial consonant duration (r = .98, measured from Rapp-Holmgren, Note 2, Figure I-B-6).

Table 5 gives values of x and y and relative P-center location for the digits used in Experiment 1. Table 6 gives values for a second set of digits spoken by the author. Onsets and offsets were determined as the earliest and latest points at which stimulus energy exceeded a fixed criterion, and vowel onsets from the most rapid increase in energy in the region of the first two formants (500-1,500 Hz). Values were computed automatically by a computer algorithm, and in the case of vowel onsets corresponded to better than 15 msec to the point of vowel onset determined by prior subjective visual and auditory inspection of the stored acoustic waveform. P-center values are of course relative and involve an arbitrary contrast.

Optimized values of α and β were determined for both sets, together with optimized values of α for Rapp's single parameter model. These, together with the percentage of variance in P-center location accounted for by the model (Equation 5), are given in Table 7. The model is valuable, however, only if a single pair of values of (α,β) can be used for all stimuli. If (α,β) need to be individually determined for each set of stimuli, the model remains descriptive rather than predictive. Table 7 therefore also gives compromise values of (α,β), (.65, .25), which can be seen to account for almost as much of the variance as the specific solutions for each set of stimuli. Eling, Marshall, and van Galen (1980) have independently shown a correlation of r = .88 with the same model for their stimuli, a set of Dutch digits. Marcus (1976) has also demonstrated the applicability of the model to a range of other speech stimuli.

Table 5
Initial Consonant Duration (x), Vowel and Final Consonant Duration (y), and P-Center Location Relative to Stimulus Onset (P) for the Stimuli Used in Experiment 1 and in Runs on APU Research Workers

Digit	x	y	P
1	50	260	−20
2	80	200	−16
3	100	230	4
4	50	400	2
5	10	350	− 9
6	140	300	50
7	130	230	39
8	20	260	−30
9	70	280	− 1

Note—Values are in milliseconds.

Table 6
Initial Consonant Duration (x), Vowel and Final Consonant Duration (y), and P-Center Location Relative to Stimulus Onset (P) for Stimuli Used by Marcus (1975)

Digit	x	y	P
1	50	370	−32
2	70	330	−29
3	110	400	33
4	30	420	−28
5	20	460	−18
6	160	300	30
7	130	310	7
8	20	350	−47
9	90	490	25

Note–Values are in milliseconds.

Table 7
Fit of Rapp's One-Parameter Model and the Two-Parameter Model to P-Center Values for Two Sets of Stimuli

	Digit Set 1 (from Table 5)			Digit Set 2 (from Table 6)		
	α	β	V	α	β	V
Rapp's Model	.50	.00	72	.50	.00	69
Two-Parameter Model	.65	.20	89	.67	.27	96
Fixed Two-Parameter Model	.65	.25	88	.65	.25	96

Note– V = percent explained variance. (See text.)

CONTINUOUS SPEECH

Although the paradigm described by Morton et al. (1976) can be used only for isolated speech stimuli, the simple acoustic model of P-center location that results is of far greater generality. In order to visualize the process it models, it is perhaps most instructively expressed as a differential form of Equation 5:

$$dP = \alpha dC + \beta dV. \qquad (6)$$

Measuring change in P-center location relative to vowel onset rather than stimulus onset, we obtain:

$$dP_V = -(1-\alpha)dC + \beta dV. \qquad (6a)$$

The model may be seen to incorporate two forces working relative to vowel onset, their resultant determining P-center location. One, proportional to initial consonant duration, tends to pull the P-center toward the onset of the stimulus; the other moves the P-center toward stimulus offset and is proportional to vowel and final consonant duration.

Huggins (1972b) described a set of experiments in which he manipulated segment durations in continuous speech. He took a naturally spoken sentence and produced two versions in which he, respectively, shortened and lengthened a particular vowel or consonant. Further experimental versions were produced in which a second adjacent segment was also modi-

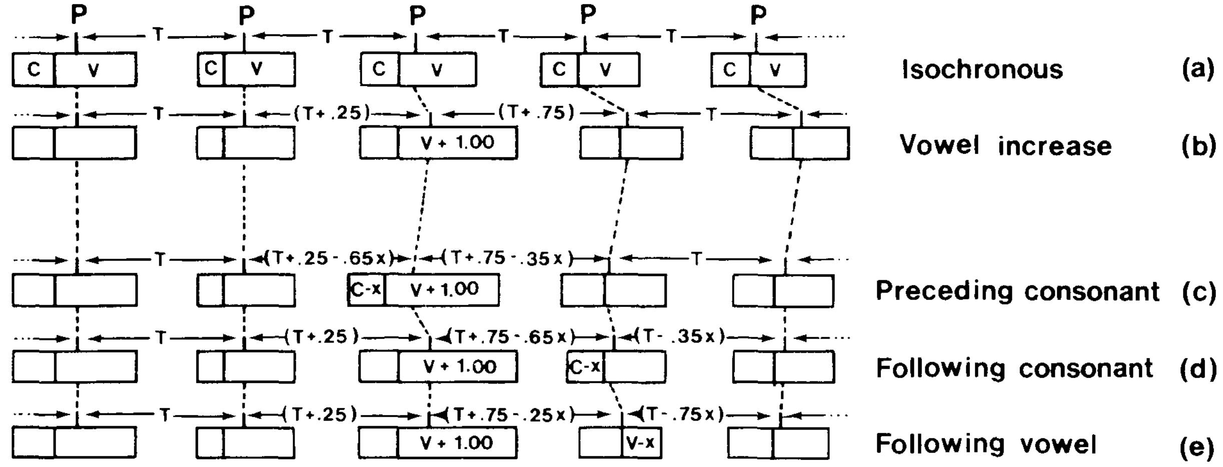

Figure 6. The effect on relative P-center timing of a regular sequence produced by an increase of 1 time unit in duration of the indicated vowel, and attempts to "compensate" for this increase in vowel duration by changes in adjacent segments.

fied and given a range of durations. The subjects' task was to judge these versions as having "natural" or "unnatural" timing. Huggins expected that, in some cases, he should be able to observe *compensation* for the shortening or lengthening of the first segment by preference for a longer or shorter "most natural" duration of the second segment.

Huggins found such compensation only when the two modified segments lay between an adjacent pair of vowel onsets, and he proposed a "vowel onset hypothesis": The crucial factor for the listener is the maintenance of timing of vowel onsets, and the modification of a particular segment can thus only be compensated for by changing another segment lying between the same vowel onsets.

Since, by our definition, perceived speech timing corresponds to P-centers, and we know that, for isolated speech stimuli at least, P-centers do not correspond to vowel onsets, Huggins' "vowel onset hypothesis" appears to conflict with our definition. We have seen that P-center location can be reasonably predicted as an interaction of events both before and after vowel onset, so why should Huggins have found compensation only for segments lying between vowel onsets? A closer analysis will show that, rather than contradicting the acoustic model of P-center location developed above, Huggins' data admirably demonstrate its applicability and extension to continuous speech.

Let us schematically represent Huggin's sentence by an isochronous sequence of equal CV syllables (Figure 6a). Figure 6b illustrates the effect of an increase of one time unit in duration of the indicated vowel on relative P-center timing, as predicted by the two-parameter acoustic model. In each subsequent section of Figure 6, an attempt has been made to compensate for the disturbance in relative timing produced by this increase in vowel duration by a decrease of x time units in the duration of the preceding consonant, the following consonant, or the following vowel, respectively (the preceding vowel is not considered, as the situation is symmetrical to that for the following vowel). Since the original timing is, in this case, isochronous, we may assume that compensation would attempt to restore isochrony to the sequence. One simple measure of isochrony, or lack of it, that may be chosen is the difference in inter-P-center timing between adjacent intervals, $D_j = T_{j+1} - T_j$. These are each zero in the original sequence. For each case shown in Figures 6c, 6d, and 6e, x was chosen to minimize the sum of squares of these differences, ΣD_j^2, and thus to give optimum temporal compensation. These optimized values of x, together with the minimized sum of squares of differences, are given in Table 8. It can be seen that only for the case shown in Figure 6d can a change in an adjacent segment substantially compensate (by 66%) for the original timing disturbance. This is precisely the case

Table 8
Optimum Compensation for an Increase in Vowel Duration by Changes in Preceding and Following Consonants and Vowels

Condition	Optimum x	ΣD_j^2	Reduction in ΣD_j^2
(b) Vowel Increase of 1 Time Unit		.9375	
(c) Preceding Consonant Decrease of x Time Units	.32	.8032	14%
(d) Following Consonant Decrease of x Time Units	.89	.3209	66%
(e) Following Vowel Decrease of x Time Units	−.29	.8036	14%

in which Huggins found compensation, when both segments lay between the same vowel onsets. The amount of compensation is much larger than the maximum found by Huggins, but we are here ignoring the fact that, in Huggins' experiment, subjects' judgments about the "naturalness" of the utterance will almost invariably be affected, not only by the naturalness of relative timing, but also by the naturalness of the duration of the modified consonant. Huggins (1972a) has also demonstrated subjects' sensitivity in making such judgments. Thus, although a value of $x = .89$ may optimally compensate for relative P-center timing, it may result in an abnormally short consonant, and, in Huggins' experiments, subjects thus chose a compromise consonant duration.

P-CENTERS AND ARTICULATION

Morton et al. (1976) began with the observation that a speaker can produce sequences of words—in their case, number names—which a listener will perceive as regular. This implies that the speaker's articulatory program obeys the same rules as his (or another's) perception. It is in fact difficult to imagine how it could be otherwise, and speech timing could have little or no communicative relevance if it were a phenomenon different for speaker and listener.

Fowler (1979) has demonstrated that speakers' speech timing is indeed affected by parameters similar to those found to be of importance in speech perception. She suggests that P-centers correspond to the timing of articulatory gestures, which are then complexly coded in the speech signal. She suggests that the listener may be judging perceptual regularity on the basis of the timing of *articulatory onsets* coded in the speech signal.

There may naturally be some articulatory parameter or parameters that correlate highly with P-center location. Indeed, there must be some basis, which may, for example, correspond to a subjective feeling of articulatory effort on which the speaker judges and controls the regularity of his own productions. However, experiments involving modification of relatively peripheral aspects of the speech signal, such as the duration of the /t/ closure in "eight," and the model consequently derived above, involve the *whole* speech signal in the determination of P-center location, not just some overt or covert onset locations. Thus, even given that it is possible to determine the moment of articulatory onset from the acoustic waveform, the whole speech signal is involved in perceived timing and must, therefore, be involved in this determination. Of course, these acoustic events have their origins in the timing of the speaker's articulations, and planning of speech production must take such perceptual consequences into account. Thus, whether or not some single acoustic *or* articulatory event in normal speech correlates with P-center location, the timing of this event forms a part in a complex of articulatory activity, and the timing of this whole complex needs to be controlled equally carefully. There thus seems little point in advocating either a simply acoustic or simply articulatory account. We must, instead, see speech as a medium of communication between speaker and listener. The properties of this medium will reflect characteristics of both production and perception.

P-CENTERS: ABSOLUTE OR RELATIVE?

Attempts to specify P-center location as corresponding to the location of some articulatory parameter may be seen as particular cases of attempts to specify absolute P-center location. The acoustic model derived in this paper shows P-center location to be the result of a computation based on events throughout the whole speech signal and suggests that it is not fruitful to continue to search for any single acoustic or articulatory event as the sole determinant of P-center location. However, I would like to approach this from a different viewpoint and state, quite simply, that it is never possible to do more than determine P-center location of a given stimulus relative to the timing of other events, whether within the same or different modalities. Let us take Rapp-Holmgren's (Note 2) experiments as an example. She asked subjects to speak nonsense words in time with an audible click; yet, although a click is undoubtedly a simpler stimulus than is speech, we know no more about the time course of its processing than we know about that of the more complex speech stimulus—indeed, we do not even know if it is subject to the same processing. Similarly, in Allen's (1972) experiments, the temporal location of a finger tap is an equally ambiguous landmark: Is the subjective moment, the P-center (*Production* Center, in this case), of a tap the moment of physical contact (as Allen assumes), the moment of transmission of motor neuron commands, of proprioceptive feedback, or some moment entirely different from any of these? Such paradigms may well give us an approximate idea of subjective P-center location if we allow ourselves to make some reasonable assumptions about the P-center location of such clicks and taps, but we should not hope for more than that—and certainly not for the determination of some meaningful and reliable point whose location is known with a precision of a few milliseconds. In the paradigm described here, we are instead measuring the timing of speech relative to speech, and, although this does not allow us to make any estimates whatsoever of absolute P-center location, we obtain accurate alignments replicable within and between subjects to as high temporal precision. Indeed, the finding of substantial individual differences in Allen's and Rapp's paradigms, as against their absence in our own, suggests that synchronization of speech with nonspeech, far from revealing anything about absolute or universal timing patterns, shows individual idiosyncrasies that are only of peripheral importance to the central issue of speech timing.

CONCLUSIONS

These results begin to demonstrate some of the complex interactions involved in the perception of speech timing. They show that before considering such questions as isochrony and "syllable-" or "stress-timing" in continuous speech, we need to be very clear as to *what* we are measuring the timing of. We must be wary of assuming that simple instrumental measurements, such as consonant and vowel onsets and durations, are related in other than a complex way to our perception. Adequate treatment of *this* complexity may save us from additional complexity in dealing with the temporal structure of continuous speech. We should also be aware that many of the data that have been used to demonstrate either isochrony or lack of isochrony now need to be carefully reexamined.

REFERENCE NOTES

1. Marcus, S. M. *Perceptual centres.* Unpublished fellowship dissertation, King's College, Cambridge, 1975.

2. Rapp-Holmgren, K. *A study of syllable timing* (Quarterly Status and Progress Report 1/1971). Stockholm, Sweden: Speech Transmission Laboratory, 1971.

REFERENCES

Allen, G. D. The location of rhythmic stress beats in English, Parts I & II. *Language and Speech,* 1972, **15**, 72-100; 179-195.

Eling, P. A., Marshall, J. C., & van Galen, G. P. Perceptual centres for Dutch digits. *Acta Psychologica,* 1980, **46**, 95-102.

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

Huggins, A. W. F. Just noticeable differences for segment duration in natural speech. *Journal of the Acoustical Society of America,* 1972, **51**, 1270-1278. (a)

Huggins, A. W. F. On the perception of temporal phenomena in speech. *Journal of the Acoustical Society of America,* 1972, **51**, 1279-1290. (b)

Marcus, S. M. *Perceptual centres.* Unpublished doctoral thesis, Cambridge University, 1976.

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

Morton, J., Marcus, S. M., & Ottley, E. A. The acoustic correlates of "speech-like": Some notes on the suffix. *Journal of Experimental Psychology: General,* in press.

Warren, R. M., & Gregory, R. L. An auditory analogue of the visual reversible figure. *American Journal of Psychology,* 1958, **72**, 612-613.

Winer, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

(Received for publication June 30, 1980;
revision accepted for publication June 5, 1981.)