The correlation of P-center adjustments with articulatory and acoustic events

KENNETH J. DE JONG

University of California, Los Angeles, California

To evaluate articulatory models of perceptual center (P-center) location, listeners performed perceptual adjustments on stimuli which were extracted from a corpus of articulatory data. To avoid streaming effects, the stimuli were not edited to obtain temporal variation; instead, they varied in stress and segmental content. Adjustments were evaluated as to their simultaneity with acoustic and articulatory events. The first experiment yielded various articulatory and acoustic correlates of P-center location; the second yielded different articulatory predictors and no acoustic effective predictors. Multiple correlation analyses showed a variation from P-center locations predicted by the articulatory events that were associated with other predictors. Thus, P-center locations do not correspond to any particular kinematic articulatory event, but rather to a complex of events taken from throughout the stimuli. These results are discussed in terms of their relevance to a model of P-centers as indices of underlying gestural timing.

The term *P-center* has been applied to two different types of data. On the one hand, its coiners (Morton, Marcus, & Frankish, 1976) spoke of a syllable's *perceptual* moment of occurrence. That is, the P-center is the point in time that subjects use to align syllables in a regular rhythm. To determine relative P-center locations, Marcus (1981) had listeners adjust the timing of a target syllable to appear halfway between regular repetitions of a base token. As has often been demonstrated, listeners do not align the onset of the target to occur halfway between the bases' onsets, but rather seem to use some point within the syllable. Further, this point did not correspond in a straightforward manner to any of the acoustic events Marcus studied.

Morton et al. (1976) noted that the P-center can also be thought of as a *production* center, and referred to earlier work on "stress beats" (Allen, 1972; Rapp, 1971) in which speakers produced various syllables rhythmically (with or without a metronome to help them). In those experiments, speakers did not produce isochronous onsets,

but rather produced tokens with isochrony of some point within the syllable. Thus, the production and perception data seem to be indexing (broadly speaking) the same aspect of speech events (Fowler, 1979).

Correspondingly, there are two different types of theoretical accounts for P-centers. Pompino-Marschall (1989) has developed a detailed account of some of his perceptual P-center data, which is stated entirely in terms of the timing of energy onsets and offsets in a set of auditory filters. In his modeling, the temporal location of the thresholds in each of the filters is subjected to a weighted averaging to determine a P-center location. Rhythmic productions, then, would be productions designed to create the pattern of auditory stimulation that will appear isochronous, according to this perceptual model.

Fowler (1979, 1983), by contrast, proposed that P-centers are associated directly with gestural events in speech production. Specifically, she suggested that listeners in P-center experiments are locating the temporal onset of the vowel gesture. Because of the complexity of the articulatory-to-acoustic mapping, these gestural events are not straightforwardly reflected in the acoustic signal. For example, the onset of movements associated with the production of a vowel precedes the acoustic release of a previous obstruent. An even more likely measure of articulatory vowel onset, such as the beginning of formant transition toward the vowel, may be difficult to obtain if the onset of vocalic movement occurs during the acoustic implosion of previous obstruents. Thus, the P-center studies and other speech isochrony studies in which fairly simple acoustic criteria were used failed to locate isochronous events in rhythmic speech. In Fowler's account, speakers simply execute rhythmic productions by producing isochronous vowel gesture onsets.

I would like to thank Keith Johnson for extensive aid in developing the experiment-running software, as well as for extensive discussion of the research reported here. I acknowledge also the important contributions of Doug Whalen, Peter Ladefoged, and two anonymous reviewers, who provided extensive comments on earlier versions of this paper. Also thanks go to Victoria Anderson, Barbara Blankenship, Aaron Shryock, and Richard Wright for allowing themselves to be subjected to the P-center paradigm. Portions of the data presented here have also appeared in *UCLA Working Papers in Phonetics*, 81, 66-75, and 85, 54-70. This work was supported by the NIDCD under Grant T32DC00029. It was carried out in the Phonetics Lab of the Department of Linguistics at UCLA. The author is currently affiliated with the Department of Linguistics at Indiana University, Bloomington, IN 47405 (e-mail: kdejong@indiana.edu).

—Accepted by previous editor, Charles W. Eriksen

Although the perceptual and production aspects of the P-center are closely bound together, thus making the separation of these two types of accounts extremely difficult, the accounts can and must be evaluated in terms of how well they integrate P-center data into either a general theory of production on the one hand, or a general theory of perception on the other. In the present paper, I will focus on evaluating a production account in which a P-center is an index of the timing of one of the gestures that has produced the stimulus.

Most of the P-center data loosely suit the sort of gestural account proposed by Fowler, except for one class of results—the effect of consonant codas on P-center location. Marcus (1981), for example, found that inserting an additional period of silence during the stop gap for [t] in the word *eight* tended to shift the P-center to later in the syllable. Cooper, Whalen, and Fowler (1988) achieved similar results by increasing the duration of later portions of the vowel. The effect of acoustic events following the vowel on P-center location was not consistently replicated by Fox and Lehiste (1987). Acoustic events in the later portion of the syllable cannot be directly indicative of the gesture going into the vowel. However, it is unclear whether or not acoustic differences in the later half of the syllable might act as secondary cues to the timing of the vowel gesture. The relative timing of movements associated with consonants and vowels are affected by differences in the constituency of the syllable coda (de Jong, 1991a; Munhall, Fowler, Hawkins, & Saltzman, 1992) and stress (de Jong, 1991b; Summers, 1987). Thus, the vowel gesture account cannot simply be ruled out because changing later portions of the signal affects P-center location.

Marcus (1981) accounted for coda effects with a regression model, which treated onset and nucleus durations as separable factors. The onset factor carried a stronger weight. In addition to being weaker, the coda factor was also found by Cooper et al. (1988) to be variable between listeners. Cooper et al. do not offer any explanation for this variability; however, if coda effects on P-center location indicate a connection between coda durations and intergestural timing in production, the variability in P-center perception may indicate a difference in the perceiver's motor organization that affects their interpretation of the durational differences in the stimuli. The effects of coda structure on vowel production vary from speaker to speaker (de Jong, 1991a; Munhall et al. 1992).

There are two barriers to interpreting these previous perceptual P-center results in terms of a gestural account. First, the articulatory behavior associated with a given acoustic stimulus is a matter of conjecture. Second, temporal aspects of the P-center stimuli are typically controlled either by using synthetic speech or by digitally editing a naturally spoken base. These techniques also create problems, because it is difficult to determine what articulatory behavior could have produced the edited tokens.

In addition, digital editing may cause problems because it weakens the psychological link between the auditory stimuli and the articulations that produced them. The results of a pilot experiment for the ones reported below illustrate the problem. The pilot was performed to set base-to-target repetition rates and determine the range of durational values that would be necessary in order to achieve rhythmic regularity with the present stimuli. The stimuli in this experiment consisted of one rendition of the word *toast* with digitally stretched and shrunk vowels. Vowel duration was manipulated by repeating or excising pitch periods in the center of the vowel at zero crossings; the same technique was used by Marcus (1981) for increasing vowel durations and by Cooper et al. (1988) for decreasing vowel durations. The results of this pilot experiment with 3 subjects indicated that perceptual adjustments occurred with consistency at 300 and 500 msec between base offset and target onset. The subjects noted that the perceptual adjustments were more difficult at the longer durations; and longer durations also introduced more fatigue into the task, owing to the longer amount of time absorbed by each trial.

At a shorter base-to-token interval of 100 msec, the stimuli tended to lose cohesion for all the listeners. That is, the high-frequency and low-frequency components separated into two percepts. This effect, called *streaming*, has been investigated in some detail by Bregman (see Bregman, 1990, for a review of this research). Streaming was especially troublesome in the present pilot experiment, because 1 subject gave different responses, depending on which percept he attended to. To avoid interference from streaming, Cooper, Whalen, and Fowler (1986, 1988) had their subjects take breaks whenever the effect occurred. In the present pilot experiment, subjects were self-paced between trials, and rest periods occurred between every block of six trials. But, even so, 1 subject especially reported that the streaming effects occurred almost immediately within the first few trials after a rest period. By contrast, unedited tokens in the experiments presented below did not tend to stream in a very similar task with the same interstimulus intervals. Listeners who had not participated in the pilot reported no streaming when listening to the unedited stimuli.

In the two experiments reported here, a perceptual adjustment paradigm, similar to that in Marcus (1981), was used. To determine accurately the articulatory behavior associated with the stimuli, and thus simplify the interpretation of P-center locations with respect to the underlying gestures associated with the stimuli, the stimuli were extracted from a corpora of X-ray microbeam data (de Jong, 1991a, 1991b; for a description of the data-gathering facility, see Nadler, Abbs, & Fujimura, 1987). Thus, the stimuli were recorded along with time-aligned articulatory records. In addition, to avoid the problems of stimuli losing cohesion, the tokens were not digitally edited; rather, temporal variation between stimuli was due to segmental and suprasegmental variation.

EXPERIMENT 1

The stimuli in Experiment 1 differed in level of stress. The speaker uttered the stimuli in different frames, which were designed to induce stress placement in various parts of the stimuli's sentences. Earlier analyses of this data have shown that varying stress in this manner causes changes not only in the overall acoustic duration of the stimuli, but also in the relative timing of the gestures that make up the stimuli (de Jong, 1991b). Thus, examining P-center adjustments of these stimuli should pinpoint which gesture, if any, is aligned with P-center location.

Method

In a perceptual adjustment paradigm, listeners were presented with a base token alternating with a target token (Figure 1). During each trial, the subjects were presented with six syllables, three alternations of a base and a target token. The subjects could then move the targets with respect to the base and replay the six syllables until they were satisfied that the syllables occurred in a regular rhythm ("like soldiers marching"). The subjects used a mouse and a horizontal series of boxes to select from a range of rhythmic possibilities that differed from one another in 3-msec increments. Initially, the base and the token were separated by 100 msec of silence. By clicking on boxes closer to the edges of the screen, the subjects chose presentations with a greater interval added or subtracted between the base and target. The same amount of time was either subtracted from or added to the duration of time between the

target and the base, so that the interval of time between presentations of the base was fixed within each trial. The final intervals between base and target were recorded as the subjects' responses.

The target stimuli consisted of 12 renditions of the word *toast* and 12 renditions of the word *totes* by a Northern Midwestern male speaker who was an undergraduate at the University of Wisconsin. The tokens occurred in the frame sentence, "I said, 'Put the toast/totes on the table.'" The speaker had been given a miniature discourse in which the utterance was to appear. In these discourses, the speaker responded to an imaginary hearer who misheard a portion of the utterance, by placing a nuclear (sentence) accent on the item that the imaginary hearer misheard. The nuclear accent occurred on the target word, on the following preposition *on* (cases in which the target bore a prenuclear accent), and on *put* (cases in which accent was precluded from being placed on the target word). In this way, the stimuli differed in how accented they were and, thus, in their temporal characteristics. Figure 2 shows the range of temporal variation in the tokens. The overall duration of the stimuli ranged over approximately 200 msec—somewhat more than for previous studies, such as those of Marcus (1981) and Cooper et al. (1988), whose stimuli continua varied from 60 to 150 msec.

From each set of 12 tokens, one of the tokens of nearly average overall duration was selected as a base. Note that the points within each stimulus that occurred isochronously in the initial presentation were the temporal midpoints of each token. Thus, under the hypothesis that listeners would align the acoustic onsets of the syllables to occur isochronously, listeners would have to actively introduce a greater period of silence before tokens of longer duration. A token of roughly average overall duration was chosen, so that under this same hypothesis, subjects were expected to add and

Figure 1. An illustration of the perceptual adjustment method. Subjects are presented with an alternating base and target. They then adjust the relative timing of the target with respect to the repeating base to achieve perceptual isochrony by using a mouse to select positions on a bar presented on a CRT.

Figure 2. Durations of target stimuli. The base token for each word is marked with a vertical line. Token numbers are as in Table 1.

subtract durations in roughly equal numbers of trials. In addition, there was also one set of trials in which the base and the target were identical.

The listeners were the author (S1) and three graduate students in linguistics (S2–S4) who were naive to the precise purposes of the experiment. For each of the listeners, each of the 12 target stimuli (including the one selected for the base) was paired with the base four times for a total of 48 trials for the two words, making a total of 96 trials.

To have a direct measure of the articulatory activity that produced each stimulus, the stimuli were extracted from an articulatory database recorded with the Wisconsin X-ray microbeam system (Nadler et al., 1987). Included in the database are movement trajectories for three points on the tongue (dorsal, mid, and tip), the upper and lower lip, and two points on the jaw. The articulatory data and elicitation procedures are described more fully in de Jong (1991a). Tokens were excised from their carrier sentences by means of digital editing. The onset and offset of each acoustic record occurred in periods of silence just before the release of the [t], and in a period of silence or near silence that occurred consistently before the onset of the following syllable. Each stimulus was checked for unnatural-sounding discontinuities at the offset of the coda consonants. Finally, the gain was set in each token so that the peak amplitudes were roughly equal in all of the tokens.

Assuming that P-centers are temporally aligned with some articulatory kinematic or acoustic event, the P-center adjustments should correspond to the anisochrony of the appropriate event in the initial presentation. To calculate the anisochronies, several different kinematic events for the jaw, the tongue tip, and the tongue dorsum were located. The jaw and the tongue tip motion should

both index any specification of openness for the vowel. Five articulatory events were extracted from jaw movement trajectories and tongue tip movement trajectories—the maximum position during the initial [t], the minimum during [o], the maximum during the consonant coda, and points of peak velocity in transition between these points. These points are illustrated in Figure 3 for a schematized jaw-movement trace. In addition, to get an articulatory measure that was more directly sensitive to the articulation of the back vowel in the tokens, four events were taken from tongue dorsum trajectories: the point of extreme retraction during the [o], speed maxima into and out of this location, and the point in time near the onset of dorsal retraction for the vowel at which the pellet crossed a threshold speed of 0.05 m/sec—a value obtained by inspecting the pellet trajectories for the moment at which the pellet was clearly under way toward its retracted position. (A similar criterion has been used in other articulatory studies, such as that of Saltzman, Löfqvist, Kinsella-Shaw, Rubin, & Kay, 1992.)

Four acoustic events were also included for comparison with the articulatory predictors: the release of the initial [t] (in this case, the onset of the stimuli), the onset of voicing for the vowel, the point of peak amplitude during the vowel (determined through visual inspection of wave-form envelopes displaying rms amplitude values calculated with a window size of 20 msec), and the offset of voicing for the vowel.

Anisochronies were then calculated as the time following an (acoustic or articulatory) event in the base and the time preceding the event in the target minus the time following the event in the target and the time preceding the event in the base. Subtracting the anisochrony of the appropriate event divided by two from the subject response should ideally yield the original intertoken duration

Figure 3. An illustrative trace of jaw movement segmented into five intervals. To calculate interevent anisochronies, the duration of the time following the event in the target word plus the duration preceding the event in the base is subtracted from the duration of time following the event in the base plus the duration of time preceding the event in the target.

of 100 msec. The anisochrony should be halved, since subject adjustments are both added to the base-to-target interval and subtracted from the target-to-base interval.

In addition, correlations were calculated between the anisochronies and the subject responses in order to compare the present results with earlier P-center models, such as those presented in Marcus (1981). First, simple regressions were performed. Then, multiple regressions were performed on the subject responses with some of the anisochronies subtracted from them, in order to determine whether there were multiple factors influencing the P-center adjustments.

Results

To determine whether the listeners responded in a consistent way to the different tokens, their responses for each word were subjected to a two-way analysis of variance (ANOVA) with subject and token as factors. In this analysis, subject was treated as a random factor, token as a fixed factor. For *toast*, there was a strong main effect of token [$F(11,33) = 11.49, p < .01$] and of subject [$F(3,144) = 21.61, p < .01$]. The interaction was weaker but significant [$F(33,144) = 2.77, p < .01$]. Patterns were similar for *totes* [subject, $F(3,144) = 12.16$; token, $F(11,33) = 8.95$; subject $\times$ token, $F(33,144) = 3.36$; all $ps < .01$]. Means and standard deviations for each token by each listener are given in Table 1. Thus, each stimulus elicited different responses, although the responses were not entirely consistent across listeners.

Table 2 summarizes the results of regressing subject responses against event anisochronies. In general, the anisochrony of events that occurred before the middle of the syllable predicted over half the variation in the subjects' responses. This was true regardless of which articulatory measures were used—jaw, tongue tip, or dorsum—or whether acoustic measures were used. These were events likely to be indices of the vowel gesture, as opposed to consonant gestures. Acoustic predictors, as a

whole, performed well for each speaker and for both words, though quite often one or more articulatory predictors outperformed the best acoustic predictor. The ranking of events differed between speakers and between words.

The standard deviation of the responses with the event anisochronies, however, presents a somewhat different picture. Figure 4 illustrates the standard deviations of subject responses with each of the event anisochronies subtracted. The standard deviation in the unsubtracted responses is marked here with a horizontal line. Fewer predictors performed well in this analysis, but the pattern of results is similar for the two words. The articulatory event that corresponds closest to the subjects' responses is the minimum position of the tongue tip (and jaw for *toast*) in the vowel. Acoustically, the most appropriate event is the onset of voicing, although the timing of peak amplitude and offset of voicing also reduce the variance in subjects' responses.

Table 3 shows the results of subtracting event anisochronies for each of the listeners. The table reveals a difference in the responses between S1 and the other subjects. Speed maxima corresponding to the vowel gesture (maximum retraction speed of the tongue dorsum, and minimum downward velocity of tongue tip, as well as dorsal events in general) correspond more with S1's responses than with those of the other subjects.

Since earlier studies of P-centers (Cooper et al., 1988; Marcus, 1981) suggested that multiple factors determine P-center locations, an analysis was performed to detect additional factors in the listeners' responses. The responses with the most effective articulatory and acoustic event anisochronies (tip minimum and voice onset time) subtracted were additionally subjected to a stepwise regression against all the event anisochronies. The results are summarized in Table 4. The analyses yielded different results for the two different words. The analyses for *totes* showed that the relative timing of the tip velocity minimum had an additional effect on the P-center adjustments. Also, voice onset anisochrony was correlated with the subjects' responses with the voice onset anisochrony subtracted from them, indicating that the voice onset anisochrony overcompensated for the subject adjustment. That is, voice onset anisochrony predicted larger subject adjustments than actually occurred. This was not true of the tip minimum event. The analyses for *toast* were somewhat more complicated for the tip minimum than for the voice onset. Both analyses showed that the timing of the tip maximum in the coda had an additional effect on P-center adjustments. As with the *totes* analyses, the voice onset event predicted larger adjustments than actually occurred.

Taken as a whole, the results indicate that the timing of the tongue tip minimum corresponds well to P-center responses, as does the timing of voice onset. Other events may correlate with P-center responses, but they either underpredict or overpredict the magnitude of the adjustment.

Table 1
Experiment 1: Mean Adjustments Performed by Each Listener for
Each Target Word (in Milliseconds)

Each Target Word (in milliseconds)										
Token Number	Listener								Average	
	1		2		3		4			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Toast</i>										
1	21.8	8.6	3.8	17.0	-16.3	11.3	-4.5	27.4	0.9	21.5
2	10.5	25.5	2.3	9.9	-16.5	7.1	-32.3	6.7	-9.0	21.5
3	16.5	5.7	-3.0	20.6	-16.5	9.0	-11.5	16.0	-3.4	18.0
4	21.8	6.2	2.3	7.5	-17.3	13.5	0.0	2.4	1.7	16.1
Base	0.0	0.0	-1.5	3.0	-0.8	1.5	0.0	0.0	-0.6	1.6
6	-9.3	6.2	-4.5	15.0	-10.3	7.9	-21.5	17.4	-11.4	13.2
7	-26.3	2.9	-22.5	8.7	-27.8	13.0	-33.0	8.8	-27.4	9.0
8	-30.0	5.5	-17.3	12.7	-34.0	5.5	-41.3	6.2	-28.1	11.8
9	-30.0	9.5	-23.3	6.7	-28.5	9.3	-51.0	4.2	-32.2	12.9
10	-34.5	5.7	-33.0	13.2	-24.7	11.3	-30.7	22.2	-30.7	13.4
11	-32.8	6.7	-22.5	9.9	-28.5	6.2	41.3	8.6	-31.5	10.1
12	-40.5	9.9	-27.0	9.5	-53.5	9.3	-51.8	8.6	-42.9	13.6
Average	-11.1	24.8	-12.2	16.7	-22.1	14.7	-26.6	21.7		
<i>Totes</i>										
1	35.3	11.3	2.3	6.7	-3.8	9.0	20.3	9.9	13.5	17.9
2	27.0	3.5	11.3	7.9	1.5	8.7	14.3	16.3	13.5	13.1
3	33.0	7.7	18.8	15.0	20.3	6.7	19.5	17.1	22.8	12.7
4	27.8	2.9	1.5	3.0	-1.5	1.7	6.8	7.9	8.6	12.5
5	11.3	10.2	6.0	11.0	3.0	6.0	4.5	10.2	6.2	9.1
6	8.3	6.7	-0.8	1.5	2.3	6.7	4.5	9.0	3.6	6.8
Base	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	3.0	7.3	-0.8	11.1	3.8	5.1	-6.8	8.6	-0.2	8.6
9	-1.5	9.0	-6.0	10.1	-4.5	13.1	-11.3	9.9	-5.8	10.2
10	-12.0	5.5	2.3	9.0	-3.7	5.7	-24.8	14.4	-9.6	13.4
11	-14.3	5.7	-9.0	10.7	-5.3	10.8	-32.3	10.2	-15.2	13.7
12	-11.8	11.8	-11.3	13.5	-20.3	6.7	-25.5	14.2	-19.7	11.9
Average	8.0	19.8	1.2	11.5	-0.7	11.0	-2.6	19.9		

Note—Adjustments = subjects' responses - original interval.

EXPERIMENT 2

To distinguish the two predictors that performed well in Experiment 1, the tip minimum and the voice onset, Experiment 1 was repeated using matched pairs of stimuli with aspirated and unaspirated initial consonants. Comparing syllables with aspirated and unaspirated onsets is especially useful for evaluating a vowel gesture model of P-center location, vis-à-vis more acoustic accounts of P-center location. Aspirated and unaspirated onsets differ dramatically in terms of the distance of the voice onset from the release of the consonant, whereas the time course of lingual and mandibular articulation is much more the same. This relationship is illustrated in Figure 5, in which an acoustic tracing and a jaw movement trace are plotted for the *tot/dot* stimuli used in this experiment. As is typical of all of the aspiration minimal pairs, there will be a large anisochrony in voice onset in the initial presentation, whereas the anisochrony of other articulatory events will be either small (such as the jaw maxima in the onset in Figure 5) or nonexistent (such as the jaw minima in Figure 5).

Method

The stimuli consisted of minimal pairs contrasting initial aspiration, extracted from a different microbeam database from that

used in Experiment 1. This database consisted of consonant-vowel-consonant (CVC) combinations uttered in the frame sentence, "Say a ___ of a ___ again." Nuclear (sentence) accent had been placed consistently on the CVC combination that occurred on the first blank. In order to avoid problems associated with streaming, the CVC combinations chosen were real lexical items; listeners reported no problems with streaming. In order to facilitate articulatory analyses, the stimuli had lingual consonants and low vowels. This restriction entailed that the tongue had to execute a clear movement into and out of the vowel. Word pairs were *gap/cap*, *gob/cob*, *gab/cab*, *dot/tot*, and *dab/tab*. The speaker spoke a St. Louis dialect of American English. The present database contains trajectories for four tongue pellets—tip, blade, mid, and dorsum—and only one jaw pellet. For more detailed information concerning this database, see de Jong (1991b).

The stimuli were extracted from their original context and presented to listeners with the technique used in Experiment 1, except that words were presented with the other member of their pair instead of with a consistent base. Each word pair was presented in both orders. For each listener, there were five word pairs by two orders by five repetitions for a total of 50 trials. Listeners consisted of 3 subjects from Experiment 1 (S1, S3, and S4) and 1 more graduate student in linguistics (S5), who was naive to the purposes of the experiment.

Articulatory and acoustic measures were taken from the stimuli as in Experiment 1. One additional acoustic event was marked from the amplitude envelope, a point at which the amplitude curve ceased increasing sharply at the beginning of the vowel. As in Experiment 1, jaw movement events were included as an index of the opening of the vocal tract. Since not all of the initial consonants

Table 2
Experiment 1: Pearson R^2 Values for Regressions Between
Intervent Anisochronies and Perceptual Adjustments

Events	Listener				Average
	1	2	3	4	
Toast					
Jaw maximum in [t]	.811	.730	.471	.279	.693
Jaw velocity minimum	.841	.753	.479	.405	.754
Jaw minimum in [o]	.757	.784	.664	.601	.830
Jaw velocity maximum	.409	.444	.059	.691	.591
Jaw maximum in coda	.109	.069	.000	.068	.019
Tip maximum in [t]	.837	.738	.658	.341	.782
Tip velocity minimum	.907	.876	.634	.476	.870
Tip minimum in [o]	.660	.701	.704	.517	.764
Tip velocity maximum	.347	.432	.398	.356	.445
Tip maximum in coda	.093	.109	.097	.086	.113
Dorsal speed threshold	.913	.808	.540	.496	.800
Dorsal speed into [o]	.968	.891	.727	.507	.932
Dorsal extremum	.686	.665	.582	.345	.354
Dorsal speed out of [o]	.193	.213	.387	.418	.324
Burst of [t]	.911	.864	.597	.486	.858
Voice onset	.832	.771	.779	.636	.896
Peak amplitude	.792	.765	.822	.650	.898
Voice offset	.458	.520	.545	.563	.598
Totes					
Jaw maximum in [t]	.805	.329	.715	.295	.702
Jaw velocity minimum	.928	.588	.821	.287	.839
Jaw minimum in [o]	.907	.626	.816	.348	.854
Jaw velocity maximum	.500	.312	.531	.270	.514
Jaw maximum in coda	.199	.024	.132	.052	.135
Tip maximum in [t]	.860	.493	.697	.474	.799
Tip velocity minimum	.843	.474	.679	.245	.721
Tip minimum in [o]	.898	.779	.884	.569	.962
Tip velocity maximum	.226	.099	.234	.045	.198
Tip maximum in coda	.032	.098	.049	.079	.063
Dorsal speed threshold	.765	.195	.598	.314	.548
Dorsal speed into [o]	.796	.268	.671	.126	.614
Dorsal extremum	.687	.243	.492	.147	.516
Dorsal speed out of [o]	.277	.144	.300	.118	.272
Burst of [t]	.951	.592	.851	.325	.870
Voice onset	.566	.858	.555	.520	.702
Peak amplitude	.720	.583	.711	.316	.729
Voice offset	.350	.041	.464	.042	.292

were coronals, the tip trajectory could not be relied upon for indexing the size of the oral opening as in Experiment 1. Three events in the consonant articulator movement (where consonant articulator = tongue tip in [d] and [t], but consonant articulator = tongue midsection in [g] and [k]) were also marked: maximum position for the initial consonant, minimum position in the vowel, and the speed maximum between these two points. Three events were extracted from the movement of the tongue midsection and also for the tongue dorsum for the articulation of the vowel: the point of maximum lowering during the vowel, and the maximum speed points into and out of this position. The statistical procedures were as in Experiment 1.

Results

To determine whether listeners performed different adjustments for different tokens, the listener responses were subjected to a three-way ANOVA with listener, word order (aspirated or unaspirated first), and word pair as factors. As in Experiment 1, subject was treated as a random factor. There was a strong main effect of

word order [$F(1,3) = 25.75, p < .0001$] and an interaction between word order and listener [$F(3,160) = 27.22, p < .0001$]. There was also an interaction between word pair and word order [$F(4,160) = 4.82, p < .01$], as well as a three-way interaction [$F(12,160) = 2.54, p < .01$]. Other effects were insignificant. To ensure that all the listeners gave different responses to the different tokens, each listener's data was subjected to a two-way ANOVA with word pair and word order as factors. Each listener had main effects of word order ($p < .0001$). In addition, S1 and S4 had significant interactions (at the .05 level) between word pair and word order. Thus, although there were some differences between subjects, they were consistent in making different adjustments for aspirated and unaspirated tokens. The means for each listener are given in Table 5.

Regressing average subject responses against the various articulatory and acoustic predictors yielded the results summarized in Table 6. As in Experiment 1, responses were strongly correlated with the onset of

Table 3
Standard Deviations of Listeners' Responses Relative to
Articulatory and Acoustic Events (in Milliseconds)

Event	Listener			
	1	2	3	4
Jaw maximum	23.88	32.78	36.88	33.40
Minimum velocity	17.81	26.01	28.99	27.64
Jaw minimum	14.78	17.20	18.39	18.21
Maximum velocity	21.34	20.22	16.47	18.67
Jaw maximum	34.30	25.95	20.06	27.68
Tip maximum	20.61	30.21	32.97	27.71
Minimum velocity	14.13	22.67	25.55	22.96
Tip minimum	16.06	15.83	16.16	16.00
Maximum velocity	20.87	17.03	15.76	19.05
Tip maximum	28.83	24.28	24.42	27.42
Dorsal speed threshold	17.88	12.12	12.06	17.09
Dorsal speed maximum	15.15	25.39	28.64	24.40
Dorsal extremum	18.06	22.36	24.80	22.01
Dorsal speed maximum	23.45	19.17	15.07	19.17
Burst of [t]	13.83	22.34	25.15	23.31
Voice onset	13.13	14.56	13.98	14.68
Peak amplitude	19.20	14.44	12.78	17.03
Voice offset	14.07	18.33	18.03	15.93
Jaw maximum	13.47	20.40	22.53	17.20
Minimum velocity	10.58	17.80	21.47	14.88
Jaw minimum	9.88	16.60	19.66	14.21
Maximum velocity	16.03	17.68	17.86	16.03
Jaw maximum	22.20	23.73	22.17	24.21
Tip maximum	14.65	22.41	24.05	19.66
Minimum velocity	10.36	16.58	18.62	15.18
Tip minimum	8.94	12.89	13.82	11.71
Maximum velocity	17.75	16.10	13.54	18.18
Tip maximum	22.08	19.19	15.49	21.90
Dorsal speed threshold	18.59	15.31	8.27	18.96
Dorsal speed maximum	14.31	21.79	25.28	18.53
Dorsal extremum	15.70	22.15	24.12	20.55
Dorsal speed maximum	17.70	17.00	15.23	18.19
Burst of [t]	9.01	16.76	20.12	13.82
Voice onset	14.20	14.35	13.54	15.71
Peak amplitude	16.45	15.68	14.62	16.11
Voice offset	12.53	15.84	17.75	14.64

Figure 4. Standard deviations of subjects' responses from Experiment 1 after the anisochronies of predictors have been subtracted.

voicing. Most articulatory predictors performed rather poorly, especially those taken from the earlier half of the syllable. An exception was the maximum position for the jaw during the onset consonant. Curiously, the next best articulatory predictor of P-center location was the timing of the velocity peak of the tongue body moving *out of* the vowel in the second half of the syllable.

The standard deviations for responses with event anisochronies subtracted are illustrated in Figure 6. The

patterns of results for the articulatory events match the correlation results. The most effective events were the maximum position of the jaw in the onset consonant, and the maximum tongue movement into the coda consonants. There is a major difference between the two analyses, however, for onset of voicing. Subtracting voice onset anisochronies from subject responses considerably increases the variance in responses. This was not true of the new acoustic measure—the end of amplitude rise—which performed well in both analyses.

The difference between the two analyses with respect to these events is illuminated by an inspection of Figure 7, in which the subjects' responses are plotted against the anisochrony of the relevant events, which represents the predicted responses. If P-centers correspond in time to the plotted event, all tokens should cluster around the $x + 100 = y$ diagonal, where 100 indicates the original interstimulus interval. The lower left panel shows the clear separation in the vertical dimension (subject responses) that is predicted by voice onset; however, the difference in voice onset predicts a greater P-center adjustment than actually occurs. This was true to a lesser extent of the acoustic amplitude rise event. This is not true of either the jaw maximum predictor or the dorsal speed maximum. Thus,

although an acoustic predictor that has a large variance, such as voice onset in the present experiment, may correlate well with P-center adjustments, it might align very poorly with P-centers. A reasonable interpretation of the present results is that voice onset correlates with smaller differences in the timing of oral gestures. Since the variance is larger for the voice onset predictor than for the articulatory predictors, the error in the measurement of the predictor is smaller relative to the P-center effect than it is in the articulatory predictors; thus, r values are higher for the voice onset predictor.

Since the stimuli in the present experiment differed in the point of articulation of the initial consonant, it is possible that some of the articulatory predictors performed

Figure 5. An aspirated and unaspirated pair of stimuli from Experiment 2 aligned at the onset of the stimulus file. Below each acoustic trace is a tracing of the vertical movement of the jaw during production of the stimulus. The anisochrony of voicing onset is indicated by the misalignment of the vertical marks in the upper and lower panels. By contrast, jaw minimum position for these two stimuli are almost isochronous in the initial presentation.

Table 4
Summary of Stepwise Regression:
(Subjects' Responses – Event Anisochrony) Versus Event Anisochronies

Tongue Tip Minimum		Voice Onset Time	
Variables in Rank Order ($F > 4$)	R^2 of Entire Model	Variables in Rank Order ($F > 4$)	R^2 of Entire Model
<i>Toast</i>			
Tip maximum in coda	.181	Tip maximum in coda	.070
Burst of onset	.246		
Dorsal speed into vowel	.292		
Tip minimum	.380	Voice onset time	.245
Tip velocity minimum	.370	Dorsal speed into vowel	.240
Dorsal speed into coda	.366	Tip velocity minimum	.237
<i>Totes</i>			
Tip velocity minimum	.181	Voice onset time	.190
		Maximum up velocity – maximum in coda	.370
Voice offset	.380	Tip minimum	.245
Burst of onset	.370	Tip velocity minimum	.240
Voice onset time	.366		

so poorly because they were inconsistent in their relationship with the underlying gestures for the vowels. For example, the midsection pellet movement would be more strongly affected by the initial consonant in the velar cases than in the coronal cases. Thus, the velocity maxima would likely occur later with respect to the initial consonant than it would in the coronal cases. To avoid the problem of such inconsistencies due to consonant place of articulation, the coronals and velars were separated and reanalyzed.

The standard deviations for the two groups are illustrated in the top two panels of Figure 6. To some extent,

the results confirm the expectation that how effective a measure is depends on how directly related it is to the articulation of the speech segments. The jaw minimum in the coronals performed quite well, as one would expect, since the jaw is involved both in the production of the low vowels and the coronals. Similarly, tongue midsection measures all performed better for the velar tokens. In other respects, these expectations were not met. For example, the tip motion events actually performed worse for coronals than velars.

To test for additional factors in P-center location, responses with the two best predictors (jaw maximum in

Table 5
Experiment 2: Mean Adjustments Performed by Each Listener for Each Target Word Pair in Each Order

Table 1: Word Error Rates (WER) in %										
First Token	Listener								Average	
	1		3		4		5			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Gab/Cab</i>										
Unaspirated	-27.6	4.6	-14.4	3.6	-24.0	7.5	-16.0	6.8	-20.5	7.8
Aspirated	28.0	7.9	15.6	3.8	15.2	4.8	7.6	17.2	16.6	11.8
<i>Gob/Cob</i>										
Unaspirated	-19.2	4.8	-14.4	2.2	-22.8	7.3	-5.2	9.2	-17.9	11.0
Aspirated	25.2	8.0	12.4	4.6	8.8	6.6	1.6	12.4	12.0	11.7
<i>Gap/Cap</i>										
Unaspirated	-14.0	5.8	-13.6	9.1	-8.4	9.8	-4.8	6.6	-10.2	8.3
Aspirated	19.6	7.0	10.4	9.7	10.0	13.7	5.2	12.9	11.3	11.6
<i>Dot/Tot</i>										
Unaspirated	-17.6	4.6	-12.4	5.9	-32.0	4.9	-17.8	5.0	-20.7	8.2
Aspirated	14.8	4.4	3.6	4.8	15.6	12.8	6.0	7.2	10.0	9.1
<i>Dab/Tab</i>										
Unaspirated	-20.0	8.8	-15.6	4.3	-18.0	9.8	-6.0	7.3	-14.9	9.1
Aspirated	19.6	3.8	13.2	1.8	22.4	7.9	9.6	12.5	16.2	8.8
<i>Average</i>										
Unaspirated	-21.7	8.1	-14.9	5.2	-21.1	10.8	-9.8	8.6	-16.9	9.6
Aspirated	21.4	7.6	11.0	6.6	14.4	10.2	6.0	12.0	13.2	10.8

Note—Adjustments = subjects' responses – initial interval.

Figure 6. Standard deviations of subjects' responses from Experiment 2 after the anisochronies of predictors have been subtracted.

the onset and tongue speed maximum going into the coda) subtracted were subjected to a stepwise regression against all of the event anisochronies. The results are summarized in Table 7. Both final regression equations include three factors, taken from throughout each stimulus. In the jaw maximum equation, the first factor is indicative of the production of the coda consonants. In the dorsal speed maximum equation, the first factor is an event indicative of onset production. Thus, there are several factors in the P-center responses which indicate a combination of information taken from throughout the tokens.

DISCUSSION

To summarize, then, the present experiments represent an attempt to reproduce earlier perceptual P-center studies with naturally varied acoustic stimuli with articulatory recordings. Although articulatory correlates were found to perform as well as or better than acoustic correlates, the present study failed to yield a single consistent kinematic correlate of P-center adjustments. In the first experiment, the tip minimum corresponded to P-center adjustments; in the second experiment, the best articulatory correlates were the maximum position of

Figure 7. Average adjustment for each subject to each token plotted against the adjustment predicted by the tongue dorsum peak speed going into the coda consonant (upper left), the jaw maximum in the onset consonant (upper right), voice onset time (lower left), and the end of the amplitude rise in the vowel (lower right).

the jaw in the onset consonant and the peak tongue body speed going into the coda. The results of the present set of experiments do not support a model in which P-center location corresponds to a particular articulatory kinematic event.

Evaluating the present results with respect to a model of P-center location corresponding to the timing of an abstract, underlying gesture, such as those proposed in articulatory phonology (Browman & Goldstein, 1986, 1990) is more difficult. In this model, observed articulatory movement can be the result of the forces exerted by more than one active underlying gesture at any given time. Thus, the kinematic events must be considered as

possible indices of the underlying vowel gesture, not as parts of the gestures themselves.

Velocity events should index underlying gestural timing best in conditions where that articulator must execute a large motion from one segment to its neighbor. Consider two examples from the present experiments. First, the tongue body must go from a high position in velar closures to a low position in the low vowels in the stimuli in Experiment 2. Second, the jaw is raised in coronal closures and lowered in the non-high vowels in stimuli used in both experiments. In both of these cases, the velocity events are aligned rather well with P-center adjustments. Position events, maxima, and minima

Table 6
Experiment 2: Pearson R^2 Values for Regressions Between
Interevent Anisochronies and Perceptual Adjustments

Event	Listener				Average
	1	3	4	5	
Jaw maximum in onset	.903	.907	.800	.776	.747
Jaw velocity minimum	.167	.080	.035	.002	.064
Jaw minimum in vowel	.100	.086	.039	.133	.072
Jaw velocity maximum	.065	.052	.063	.215	.070
Jaw maximum in coda	.035	.013	.015	.075	.025
C-articulator maximum in onset	.175	.153	.040	.060	.094
C-articulator minimum in vowel	.147	.063	.048	.047	.072
C-articulator maximum in coda	.140	.068	.138	.148	.111
Dorsal speed into vowel	.193	.131	.016	.002	.068
Dorsal extremum in vowel	.134	.169	.025	.055	.078
Dorsal speed out of vowel	.234	.302	.394	.574	.303
Mid speed into vowel	.132	.106	.020	.004	.055
Mid extrema in vowel	.552	.402	.257	.182	.330
Mid speed out of vowel	.370	.568	.684	.694	.592
Burst of onset C	.146	.221	.346	.239	.205
Voice onset	.961	.934	.892	.751	.816
End of sharp amplitude rise	.852	.871	.919	.894	.796
Peak amplitude	.829	.461	.348	.311	.384
Voice offset	.239	.199	.123	.244	.174

should also align well when the articulator is critical for the appropriate gesture. Position events that align well are the tip and jaw minima in Experiment 1, the jaw minima in coronals in Experiment 2, and the tongue body minima in all the stimuli in Experiment 2. All these are cases in which the articulator is critical in the production of the vowel.

Also promising for a gestural account is the performance of the dorsal velocity predictors in the latter half of the syllable in Experiment 2. Such results suggest that there might be an articulatory relationship between events in the latter half of a syllable and the timing of articulatory behavior in the first half of the syllable. Finding and describing such a relationship is necessary for any articulatory account to explain coda effects on P-center location.

However, the results of the stepwise regression analyses suggest that a complex of gestures is involved in determining P-center location. This is especially apparent in Experiment 2, in which the first factor included in the stepwise regression analyses represented the half of the syllable not already accounted for in the P-center responses. After subtraction of a factor indexing some aspect of the timing of the opening gesture (the jaw max-

imum in the onset consonant), the first factor included in the regression equation is an index of the timing of the closing gesture (the jaw velocity maximum going into the coda consonant). After subtraction of a factor indexing the closing gesture (the dorsal speed maximum out of the vowel), the first factor included in the equation is an index of the timing of the opening gesture (the jaw velocity minimum going into the vowel). Thus, P-center adjustments are not indicative of listeners locating a single gesture. The P-center, if indicating any articulatory event, is indicating an event that is more abstract than gestures, perhaps a rhythmic center, or stress beat, as suggested, for example, in Allen (1975).

The results of the present experiments are thus somewhat disappointing for the application of P-centers to articulatory research, since they have not revealed unequivocal evidence for an articulatory basis for P-centers. However, regardless of what the basis of P-center judgments may be, whether general perceptual mechanisms or higher level rhythmic organization of speech actions, they may still have application for articulatory research. Listeners' judgments about temporal location may uncover important principles in the organization of the various actions that constitute speech. One area of applica-

Table 7
Summary of Stepwise Regression:
(Subjects' Responses — Event Anisochrony) Versus Event Anisochronies

Jaw Maximum in Onset		Dorsal Speed out of Vowel	
Variables in Rank Order ($F > 4$)	R^2 of Entire Model	Variables in Rank Order ($F > 4$)	R^2 of Entire Model
Jaw velocity maximum	.169	Jaw velocity minimum	.485
Jaw maximum in onset	.223	Jaw maximum in onset	.600
Dorsal extremum in vowel	.246	Jaw velocity maximum	.628

tion, for example, is the coordination of intonational events—high and low tones—and the syllables with which they are associated, a problem discussed in Silverman and Pierrehumbert (1990). Correlating P-center adjustments with these kinds of poorly understood coordination phenomena remains a promising line of research.

REFERENCES

- ALLEN, G. (1972). The location of rhythmic stress beats in English: An experimental study. I. *Language & Speech*, **15**, 72-100.
- ALLEN, G. (1975). Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, **3**, 75-86.
- BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- BROWMAN, C. P., & GOLDSTEIN, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, **3**, 219-252.
- BROWMAN, C. P., & GOLDSTEIN, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, **18**, 299-320.
- COOPER, A. M., WHALEN, D. H., & FOWLER, C. A. (1986). P-centers are unaffected by phonetic categorization. *Perception & Psychophysics*, **39**, 187-196.
- COOPER, A. M., WHALEN, D. H., & FOWLER, C. A. (1988). The syllable's rhyme affects its P-center as a unit. *Journal of Phonetics*, **16**, 231-241.
- DE JONG, K. J. (1991a). The articulation of consonant-induced vowel duration changes in English. *Phonetica*, **48**, 1-17.
- DE JONG, K. J. (1991b). *The oral articulation of English stress accent*. Unpublished doctoral dissertation, Ohio State University.
- FOWLER, C. A. (1979). "Perceptual centers" in speech production and perception. *Perception & 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.
- FOX, R. A., & LEHISTE, I. (1987). Effect of unstressed affixes on stress beat location in speech production and perception. *Perceptual & Motor Skills*, **65**, 35-44.
- MARCUS, S. M. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics*, **30**, 247-256.
- MORTON, J., MARCUS, S. M., & FRANKISH, C. (1976). Perceptual centers (P-centers). *Psychological Review*, **83**, 405-408.
- MUNHALL, K., FOWLER, C., HAWKINS, S., & SALTZMAN, E. (1992). "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics*, **20**, 225-240.
- NADLER, R. D., ABBS, J. H., & FUJIMURA, O. (1987). Speech movement research using the new x-ray microbeam system. In *Proceedings of the 11th International Congress of Phonetic Sciences* (Vol. 1, pp. 221-224). Tallinn: Academy of Sciences of the Estonian S.S.R.
- POMPINO-MARSHALL, B. (1989). On the psychoacoustic nature of the P-center phenomenon. *Journal of Phonetics*, **17**, 175-192.
- RAPP, K. (1971). A study of syllable-timing. In *Quarterly Progress and Status Report*, pp. 14-19. Stockholm: University of Stockholm, Speech Transmission Laboratory.
- SALTZMAN, E., LÖFQVIST, A., KINSELLA-SHAW, J., RUBIN, P. E., & KAY, B. (1992). A perturbation study of lip-larynx coordination. In J. J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge, & G. E. Wiebe (Eds.), *ICSLP 92 Proceedings* (pp. 19-22). Edmonton, AB: University of Alberta, Department of Linguistics.
- SILVERMAN, K., & PIERREHUMBERT, J. B. (1990). The timing of pre-nuclear high accents in English. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 72-106). Cambridge: Cambridge University Press.
- SUMMERS, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America*, **82**, 847-863.

(Manuscript received October 7, 1992;
revision accepted for publication May 1, 1994.)