Perception & Psychophysics
1988, 43, 94-98

Perceived timing is produced timing: A reply to Howell

CAROL A. FOWLER, D. H. WHALEN, and ANDRÉ M. COOPER
Haskins Laboratories
New Haven, Connecticut

In his commentary on our paper (Cooper, Whalen, & Fowler, 1986), Peter Howell (1988) attempts to show that the center of gravity of an acoustic signal is sufficient to predict perceived timing of the signal. In our reply, we confine ourselves to four main points:

1. Howell offers a *description* of acoustic correlates of perceived timing. As such, it need not conflict with our *explanatory* account that listeners perceive timing as produced. However, Howell does not have the description right. It conflicts with already published data.

2. Howell declined to apply his formulas to the stimuli of Cooper et al. and to model their data quantitatively. Apparently, he neglected to try it out on his own stimuli and findings too. It makes the wrong predictions. The problem is with his geometry, however; *centers of gravity* of his own stimuli and those of Cooper et al. do pattern in the same way as listener judgments.

3. In a replication and extension of the work by Tuller and Fowler (1981), we show that the phonetic composition of a syllable, and not its center of gravity, predicts perceived timing when phonetic composition and center of gravity are opposed in a set of stimuli.

4. Our account of perceived and produced timing meets Howell's proposed criteria for adequacy better than does his own.

These points, we acknowledge, go beyond our charge to address Howell's specific complaint that Cooper et al. wrongly discount his view on the basis of their findings. We take the longer route for several reasons. First, although we stand behind our conclusions based on Howell's (1984) remarks, he has elaborated his earlier ideas here and they require their own assessment. Second, it is easy enough to contrast certain predictions of his account and ours experimentally. In our view, Howell should have run such a test to ascertain that his account was worth trying to resuscitate before going to press with his complaint. He did not, and so we did it for him. Finally, our view and Howell's differ markedly in respect to their contexts of theoretical and experimental support. We wanted to point that out.

Preparation of the manuscript was supported by NICHD Grant HD-01994 and NINCDS Grant NS-13617 to Haskins Laboratories. We thank George Wolford for commenting on an earlier version of the manuscript. C. A. Fowler is also affiliated with Dartmouth College, and A. M. Cooper is also affiliated with Yale University. The zip code for Haskins Laboratories in New Haven, CT, is 06511.

The Nature and Relationship of the Claims

The general classes of account that we (Cooper et al., 1986; Fowler, 1979; Tuller & Fowler, 1980) and Howell (1984, 1988) offer to handle findings of systematic variation in perceived timing of syllables and of acoustically specified events are different and, indeed, need not conflict in principle. In their particular instantiations, however, they do conflict.

Howell's account is descriptive. That is, it is meant to characterize the timing judgments of listeners in terms of a higher order description of the amplitude envelope of acoustic signals. Although one can imagine an explanation standing behind the description, couched perhaps in terms of the response characteristics of the auditory system, or in terms of the information that the higher order description provides the listener about the acoustic signal's distal source (see below), as yet none has been offered.

By contrast, one of us has proposed an explanation for listeners' judgments within the context of a more general (direct realist) theory of speech perception (Fowler, 1986). The general theory is that, in speech perception, as in perception generally (Gibson, 1966, 1979), proximal stimulation (acoustic signals, reflected light, etc.) serves not as something perceived, but rather as a carrier of information about the distal source of its structure—about linguistically organized activity of the vocal tract in speech perception and about other sound-producing or visible environmental events in auditory and visual perception, respectively. It is the distal event that is perceived, then, not its information carrier.

Our specific account of listeners' perception of the timing of monosyllables derives from several observations of talkers' attempting to produce isochronous sequences of monosyllables. They produce precisely the measured acoustic anisochronies that listeners require to hear the sequences as isochronous (Fowler, 1979); in the CVCs we have examined, muscle activity for initial-consonant, vowel, and even final-consonant production is isochronous, whereas measurements of acoustic onset-onset times exhibit the usual anisochronies (Tuller & Fowler, 1980). When talkers produce perceptually isochronous sV, stV, and strV syllables, articulation of initial consonants is not isochronous, but articulation of vowels probably is (Fowler & Tassinary, 1981).

Our account of the talker's behavior, then, is quite simple, and our account of the listener follows directly from it. Asked to produce a sequence of isochronous CVCs, talkers do so; listeners extract information from the acoustic signal that specifies articulatory timing.

Were the center of gravity to provide reliable information for articulatory vowel timing in speech, we would have a good idea about how the listener recovers vowel timing from the acoustic speech signal. Indeed, in natural,

undoctored speech, there must be a close relationship between a syllable's center of gravity and articulatory vowel timing. Vowels are associated with an overall higher amplitude signal than consonants, and so the peak jaw opening for the vowel is associated with the peak amplitude of the syllable, itself an important determinant of the center of gravity for the syllable.[1] There is other, spectral, information for vowel timing too, however, and, as we will show, this information must be more salient to listeners than information provided by the amplitude envelope.

That demonstration aside, Howell's account of the center of gravity of the amplitude envelope already conflicts with published findings: (1) Fowler (1979) had two talkers alternate prevoiced and voiced stops under isochrony instructions. Because prevoiced stops have a voiced closure and voiced stops a silent one, the center of gravity of the prevoiced stops should be earlier than that of the voiced stops. However, both of the talkers produced isochronous sequences measured from stop release. Nonsignificant numerical departures from isochrony exhibited by both talkers were in the wrong direction for the center-of-gravity hypothesis. Howell apparently was unaware of this finding. (2) Marcus (1981) found no effect on perceived timing of an increase in amplitude of a word-final stop burst. Howell acknowledges that this finding cannot be handled by the center-of-gravity proposal. (3) Tuller and Fowler (1981) effected radical changes in the amplitude contours of syllables and found no change in perceived timing. Howell raises legitimate objections to this research, which we address under "A Direct Test of the Center-of-Gravity Account" below.

The Center of Gravity and the Findings of Cooper et al. (1986)

As a review of Howell's (1988) submission, one of us urged him to eliminate his formula in both its original and simplified forms, to compute centers of gravity directly from the waveforms, to substitute talk of waveforms for talk of geometric figures, and to *show* quantitatively, using our raw data if he wished, that his account indeed handles the findings of Cooper et al. Howell chose not to follow any of this advice. In response, we offer this brief tutorial.

Center of gravity is an intuitively clear concept. Accordingly, the formulas are not needed to make the concept more accessible to intuition. Also, it is easy to compute a center of gravity from a digitized waveform, so the (crude, see our Figures 1 and 2) geometric approximations are not needed for that either. Finally, it is peculiar, to say the least, to offer a formula in *two* versions, original and simplified, that is never used to derive and test quantitative predictions.

Clearly, Howell did not check out the adequacy of the formulas with respect to his own stimuli either. We did, and we found that they make the wrong predictions even of relative P-center locations for the stimuli of Howell's (1984) Experiment 1,[2] even though Howell is certainly right about the effects his experimental manipulations had on his stimuli's centers of gravity.

The center of gravity of a waveform can be found by summing absolute values of digitized voltages and dividing the sum by 2 to obtain half the waveform's area. The point in time from onset in which half the area is reached is the center of gravity.

We computed the centers of gravity of the continuum endpoints of Experiments 3 and 4 from Cooper et al. Howell's guesses concerning their centers of gravity are quite correct. In Experiment 3, we traded duration of a silent interval interposed between frication and vowel with frication along a continuum, and we found no difference in perceived timing among continuum stimuli. The endpoints of the continuum both have centers of gravity 331 msec from syllable onset. In Experiment 4, we traded silence in the syllable onset for vowel duration in the rhyme and found a later P center for syllables with longer silent intervals. The endpoints of that continuum have centers of gravity 331 and 377 msec from onset. Howell's center-of-gravity account, stripped of its geometrization, can generate the findings of those experiments and probably of Experiments 1 and 2 as well.[3]

Where does that leave the center-of-gravity account? It is still inconsistent with at least the first two of the three findings on produced and perceived timing that we listed earlier. It still offers a more complicated account than ours of the talker's behavior (see "Speech and Nonspeech" below). And it is no *better* than our account of the findings of Cooper et al. If, given this state of affairs, Howell considered his account still viable, he should then have gone on to contrast predictions of his view from those of ours. After all, a theoretical account is unlikely to be superseded by an account that does almost, but not quite, as well. Had he tried an experimental comparison, as we show below, he would have learned that his account was wrong.

A Direct Test of the Center-of-Gravity Account: Tuller and Fowler (1981) Revisited

Tuller and Fowler (1981) peak-clipped naturally produced CVC syllables and made a test tape in which pairs of syllables were presented in sequences that had one of two timing patterns. The sequences had either acoustically isochronous onset-onset times or they had the onset-onset times produced by the original talker under isochrony instructions. Pairs of sequences, matched in their component syllables but different in their timing pattern, were presented to listeners who were asked to judge which sequence had the more isochronous pattern. Listeners consistently chose the sequence with the pattern used by the original talker. Tuller and Fowler concluded that the amplitude contour of a syllable does not affect its perceived timing.

Howell guessed correctly that the syllables might not, in fact, be peak-clipped. We neglected the effects that deemphasis of higher frequencies might have on our syllables as they were output from the PCM system at

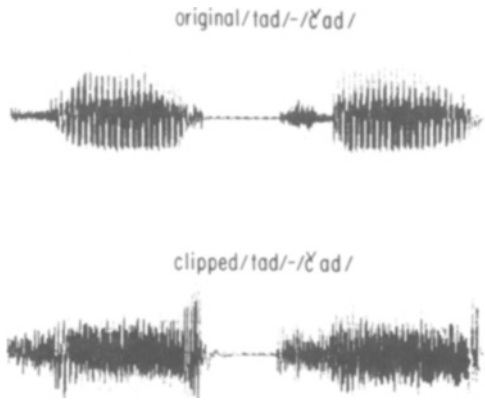

Figure 1. A pair of syllables from Tuller and Fowler (1981) in their original form and after peak-clipping.

Haskins Laboratories. Figure 1 shows two original syllables from that experiment and their "peak-clipped" counterparts. Obviously, we did not succeed in giving the syllables a rectangular amplitude envelope as we had intended. Nonetheless, we did radically change the contours. Howell speculates, however, that any changes to the centers of gravity effected by changes in amplitude contour might be undetectable by the insensitive experimental paradigm we used.

We could, of course, speculate differently, but someone has to get up out of the armchair and find out. And so, in the tradition of those who actually go out into the stable and *count* the horse's teeth, we offer the following experimental test of the center-of-gravity account.

Figure 2 displays the stimuli we used. In the top row, left and middle cells display waveforms of a /ba/ (duration: 351 msec) and a /sa/ (duration: 533 msec) syllable, produced by a male, native speaker of English. These syllables were filtered at 10 kHz and sampled at 20 kHz by a New England Digital computer. This computer neither preemphasizes high frequencies of speech input nor deemphasizes them on output. The center of gravity for /ba/ was 125 msec from onset; for /sa/, it was 291 msec following onset. These points are indicated by arrows in the figure.[4]

We then digitally peak-clipped the syllables to obtain the waveforms shown in Figure 2 directly below those of normal /ba/ and /sa/. Peak-clipped /ba/ (henceforth "ba_p") had a center of gravity 161 msec from onset, 36 msec later than that for normal /ba/ ("ba"). The rightward shift occurred because the major change in the amplitude contour was at the end of the syllable, where, in the original /ba/, the vowel tailed off. Peak-clipped /sa/ ("sa_p") had a center of gravity nearly identical to that of normal /sa/ ("sa") at 287 msec. The center of gravity shifted little because major changes to the envelope occurred at both ends of the syllable.

We made two other versions of "sa." In one ("sa_s"), the amplitude of the frication was clipped but the vowel was unchanged. In the other ("sa_e"), the amplitude of the vowel was clipped, but that of the fricative was unchanged. The center of gravity of "sa_s" was 143 msec after syllable onset, fortuitously exactly midway between those for "ba" and "ba_p." The center of gravity for "sa_e" was 320 msec from syllable onset, 30 msec to the right of the center of gravity for "sa"and 34 msec to the right of that for "sa_p." Perceptually, these syllables maintained their identifiability as speech syllables and, specifically, as /ba/ and /sa/. They sounded as if the speech were being passed through a very poor loudspeaker.

Three subjects (2 naive subjects from an introductory psychology class and C.A.F.) were run in an experiment using the "method of adjustment" of Cooper et al. (1986; see also Marcus, 1981). In this procedure, subjects hear a pair of syllables and use labeled keys on the calculator pad of a computer terminal to shift the syllables' relative timing until they sound isochronously timed. Details of the procedure are given in Cooper et al.

In all trials, the subjects heard (normal) "ba" (henceforth, the reference syllable) paired with itself or with one of the other five syllables (the target syllable). The total cycle from onset of one reference syllable to onset of the next was 1,400 msec. Listeners made 36 adjustments in each of two sessions lasting about 1 h, so that each subject made 12 adjustments to each syllable pair.

Figure 3 displays the outcome of the experiment separately for the two sessions. The figure plots differ-

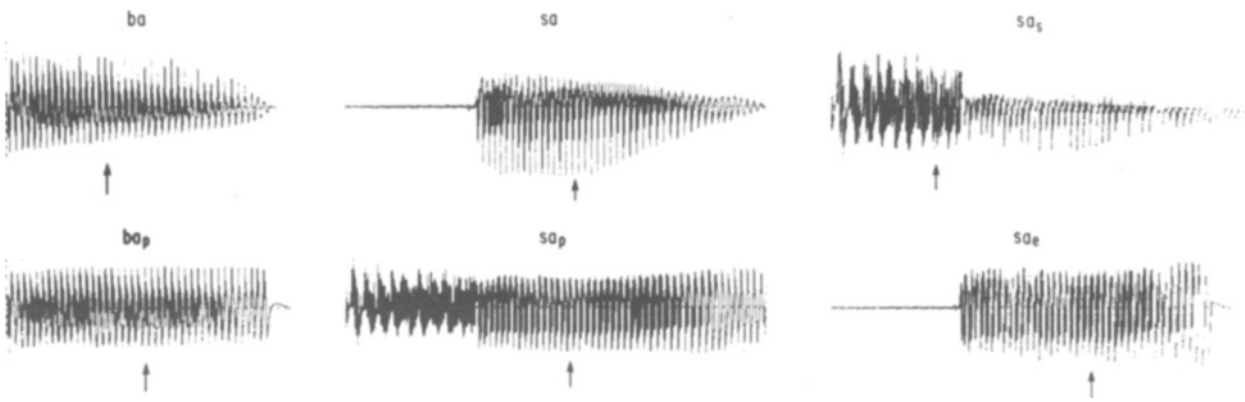

Figure 2. The stimuli used in the present experiment. Along the top, normal /ba/, normal /sa/, /sa/ with frication peak-clipped. Along the bottom, peak-clipped /ba/, peak-clipped /sa/, /sa/ with the vowel peak-clipped. Arrows mark the syllables' centers of gravity.

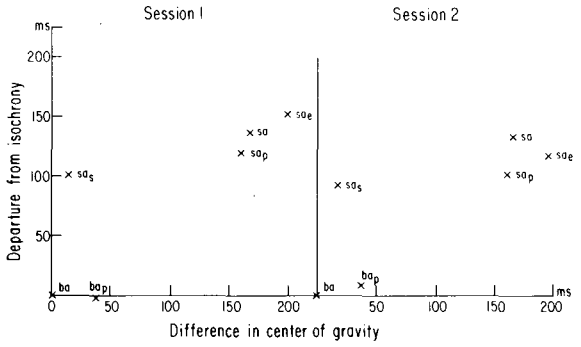

Figure 3. Departures from isochrony of listeners' adjustments. Data from 3 subjects are averaged; data are presented separately for each of two experimental sessions.

ences between reference and target syllable in center of gravity on the abscissa and departure from acoustic isochrony on the ordinate. If perceived timing covaries with center of gravity, then the points should increase monotonically to the right. The value should be zero for "ba" paired with itself and should increase as center of gravity for the target syllable moves farther into the syllable. If, instead, perceived timing covaries with articulatory timing, then departures from isochrony for pairs of /ba/ syllables should be zero; those for pairs including any of the /sa/ syllables should be similar to each other and well above zero. (This particular offset from isochrony occurs because /b/, a stop, has a silent closure interval, whereas /s/'s closure is noisy. Talkers producing articulatorily isochronous syllables, then, will produce closures at temporally equal intervals, but one of the closures will be silent.)

The results provide a clear disconfirmation of the center-of-gravity account. The major effect in the data is a difference in departures from isochrony for pairs of /ba/ syllables as compared with pairs consisting of /ba/ and /sa/ in alternation. There are numerical differences among the various /ba/-/sa/ conditions, and, in Session 1, these differences do pattern ordinally just as Howell predicts. That is, departures from isochrony increase monotonically as the centers of gravity move farther into the /sa/ syllables. However, they do not pattern properly with respect to /ba/-/ba/ pairs. Nor is their ordinal relationship preserved in Session 2.

An analysis of variance with session, syllable pair, listener, and token as factors and listener and token as random factors revealed just one significant effect, a main effect of syllable pair [$F(5,10) = 58.93, p < .001$]. The center-of-gravity and articulatory timing accounts were each evaluated in separate planned comparisons. In one, weights were applied to departures from isochrony that reflected center-of-gravity differences among the syllables of an adjusted pair. The F for this comparison was highly significant [$F(1,10) = 188.41, p < .001$], and the comparison accounted for 64% of the variance in the main effect of syllable pair. (A prediction that may better respect Howell's wish to keep things "qualitative" groups the three pairs with target-syllable centers of gravity 161 msec or less from onset and the three with target-syllable centers of gravity 287 msec or more from onset. This comparison explains 68% of the variance.) To test the articulatory-timing predictions, we contrasted sequences involving only /ba/ syllables with those involving /ba/ and /sa/. This F also was highly significant [$F(1,10) = 278.93, p < .001$], and the comparison captured nearly all (95%) of the variance in the main effect of syllable pair. Clearly, the articulatory-timing hypothesis is superior to the center-of-gravity account, and it is nearly as good as it could be.

We acknowledge that the nonsignificant, numerical differences among the /ba/-/sa/ conditions do tend to pattern according to center of gravity, /ba/-/ba/ pairs excluded from consideration. Statistically, this is random variation, but possibly, with more data, it would emerge as significant. We point out, however, that the effects are very weak, as compared with those of the phonetic composition of the syllable, and that they are as consistent with a view that listeners extract information about vowel timing as they are with a view that they track center of gravity.

Despite his presentation of formulas, Howell prefers to keep his speculations qualitative, on grounds that the center-of-gravity idea as outlined may not quite work without elaboration in terms of the spectral composition of stimuli. He is probably correct that an elaborated model of this sort might work better than the present version. We predict, however, that when Howell discovers how to weight acoustic energy by its spectral composition, he will have discovered the acoustic consequences of vowel articulations.

Speech and "Nonspeech"

As Howell points out, an adequate account of the P center will have to handle listeners' judgments of speech and "nonspeech" events and talkers' productions of perceptually isochronous sequences of syllables.

Our account of the talker's behavior is superior to Howell's. We propose that when talkers are asked to produce isochronous monosyllables, they follow instructions. Howell's account (see Howell, 1984) is that talkers must estimate where in time the centers of gravity of their to-be-produced syllables will fall relative to syllable onset, and they must adjust their articulatory timing so as to make centers of gravity isochronous. Coincidentally, in the syllables studied by Tuller and Fowler (1980), this led to isochronous muscle activity for successive syllables. If only the talkers could have anticipated that, they could have saved themselves computation.

Our account of listeners to speech is straightforward, too, though incomplete. We suppose that listeners extract information about articulatory timing from the acoustic speech signal just as they extract information about environmental events from informational media generally. The account is incomplete because we do not yet know precisely how vowel timing is specified.

As for listeners' perception of the timing of "nonspeech," our theory requires that a distinction be made.

"Nonspeech" is not a natural category of sound-producing event in the way that "speech" is. In addition to all manner of sound-producing events in the natural environment other than speech, it includes signals that researchers may concoct in the laboratory that may have no readily identifiable distal source.

We do have a prediction to make about perceived timing of natural sound-producing events. It is that listeners will use their acoustic consequences as information about the timing of the events themselves. As for laboratory creations, we know what to predict only for those that can be identified with an apparent distal source. For those that can, their perceived timing should be that of the apparent distal event. (The "nonspeech" stimuli of Howell's, 1984, first experiment may fall into that category; that is, they were made to mimic fricative- or affricative-vowel syllables; possibly, they sounded like degraded "sha"s and "chas.") As for others that have no identifiable distal source, we do not know what to predict, but we are willing, for the present, to leave the account of those stimuli to Howell.

By way of summary, we contrast Howell's approach to the theoretical and experimental problem of the P center to our own. He presents a quantitative model, but uses it to make only qualitative predictions. The model fails to make the right predictions for his own stimuli, and the idea of center of gravity's ostensibly standing behind the model fails to account for previously published data. He neglects to test his model before offering it to the public, and it fails a test when one is devised. Our theoretical and experimental work has produced qualitative predictions of relative P-center location that are correct for our own stimuli and those of other researchers. As importantly, we have offered a rationale, based on the direct perception of the source of an acoustic signal, that allows us to understand what we have found. Until Howell can falsify this account, he has made no significant contribution. Until he can describe the array of published experimental findings, he has made no contribution at all.

REFERENCES

COOPER, A. M., WHALEN, D. H., & FOWLER, C. A. (1986). P-centers are unaffected by phonetic categorization. *Perception & Psychophysics*, **39**, 187-196.

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

FOWLER, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, **14**, 3-28.

FOWLER, C. A., MUNHALL, K., SALTZMAN, E., & HAWKINS, S. (1986, December). *Acoustic and articulatory evidence for consonant-vowel interaction*. Paper presented at the 112th meeting of the Acoustical Society of America, Anaheim, California.

FOWLER, C. A., & TASSINARY, L. (1981). Natural measurement criteria for speech: The anisochrony illusion. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (Vol. 9, pp. 521-535). Hillsdale, NJ: Erlbaum.

GIBSON, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.

GIBSON, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

HOWELL, P. (1984). An acoustic determinant of perceived and produced anisochrony. In M. P. R. Van den Broecke & A. Cohen (Eds.), *Proceedings of the 10th International Congress of Phonetic Sciences* (pp. 429-433). Dordrecht, Holland: Foris.

HOWELL, P. (1987). Prediction of P-center location from the distribution of energy in the amplitude envelope: I. *Perception & Psychophysics*, **43**, 90-93.

MARCUS, S. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics*, **30**, 247-256.

TULLER, B., & FOWLER, C. A. (1980). Some articulatory correlates of perceptual isochrony. *Perception & Psychophysics*, **27**, 277-283.

TULLER, B., & FOWLER, C. A. (1981). The contribution of amplitude to the perception of isochrony. *Haskins Laboratories Status Report on Speech Research*, **SR-65**, 245-250.

NOTES

1. We explain the findings of Howell's (1984) second experiment, in which long and short vowels had different P centers, in precisely this way. The peak jaw opening for an inherently long vowel is later, relative to articulatory onset, than it is for a short vowel (e.g. Fowler, Munhall, Saltzman, & Hawkins, 1986). Thus, not only is the center of gravity shifted later in the long vowel, so is the peak opening of the jaw.

2. Howell's stimuli consisted of a pair of speech syllables and a pair of nonspeech analogs. The following description is of the speech stimuli, but it characterizes both stimulus pairs. Both members of the pair had vocalic segments (312 msec in duration) preceded by frication noises (148.8 msec in duration). Members of a pair differed in the amplitude rise time of the frication noises. One ramp extended over 120 msec and the other over just 40 ms. If we try to partition these stimuli into geometric forms, we have two choices as to how to treat the frication noise. If we impose a triangle on the whole of the frication noise (as Howell instructs: "The amplitude envelope of the frication in a syllable ... can be represented as a right triangle"), from onset to end, then the two members of a pair will have identical centers of gravity as computed by the formulas, because the frication noises start and end at the same amplitude and have the same duration. If we impose a triangle just over the amplitude ramp and include the remaining 20.8 msec of the frication of the one stimulus and the remaining 108.8 msec of frication of the other in the vowel "rectangle," then the computed centers of gravity are different for the two stimuli, but the difference is opposite to what it should be. The syllable with the longer ramp has an earlier center of gravity than the syllable with the shorter ramp. (Our calculations gave a "center of gravity" of 138.9 msec for the stimulus with a long ramp and one of 200.2 msec for the stimulus with a short ramp.)

3. As Howell complains, Cooper et al. do conclude that Howell's (1984) proposals are incompatible with their findings. Howell (1984) does not refer to the center of gravity, but only to the distribution of energy in the amplitude envelope of stimuli. In our experiments in Cooper et al., we changed that distribution and either saw P-center shifts or did not, depending on whether the syllable onset did not or did maintain an invariant duration. We consider our conclusions justified based on Howell's (1984) account.

4. Waveforms in Figure 2 are stimuli as they appeared after we output them through the listener's port interfaced to the computer, recorded them on audio tape and re-input them to the computer. We did that to satisfy ourselves that the stimuli in Figure 2 did accurately depict the syllables presented to listeners. On re-input, we adjusted overall amplitude to make it roughly similar across stimuli. Centers of gravity were computed on these re-input stimuli, and these were the stimuli subjects heard. Fricatives are low in amplitude relative to those in the figures of Cooper et al. because the latter show effects of preemphasis.

(Manuscript received March 5, 1987;
accepted for publication May 25, 1987.)