

Prediction of P-center location from the distribution of energy in the amplitude envelope: I

PETER HOWELL
University College London
London, England

Morton, Marcus, and Frankish (1976) reported that digits presented with the same duration between onsets were not perceived as isochronous (equally timed). If subjects are allowed to temporally position a pair of digits so that they appear to be regular, they are not adjusted to have the same distance between their onsets. That is, the stimuli have to be physically anisochronous in particular ways to appear perceptually isochronous. Morton et al. accounted for these phenomena by arguing that the psychological moment of occurrence of a sound (its P center) did not coincide with its onset.

A number of studies have investigated P centers, and these fall into three principal groups: First, the acoustic waveform has been altered in a variety of ways to see what manipulations cause the location of the P center to shift (e.g., Cooper, Whalen, & Fowler, 1986; Marcus, 1981). Second, speakers have been asked to produce isochronous lists. When this is done, the timing corresponds to what would be needed for the items to be perceived as isochronous (e.g., Fowler, 1979). Finally, analogous nonspeech tasks have been examined (Howell, 1984; Vos & Rasch, 1981). The requirements of an adequate account of the P-center phenomenon are that it should account for the way in which P centers of speech are judged by listeners to vary (whether the waveforms are naturally spoken syllables or ones that have had their acoustic forms manipulated) and account for why the phenomenon occurs both with nonspeech sounds and in speech production.

Howell (1984) proposed that the temporal distribution of energy in a syllable affects a subject's judgments of P center location. Thus, if the energy tends to occur early in a syllable, its P center will be nearer the start of the syllable. Conversely, if the energy tends to occur late in a syllable, its P center will be nearer the end of the syllable. This explanation would apply directly to nonspeech and could account for why the phenomenon occurs in speech production too; when speakers try to produce isochronous lists, their judgments about their productions may be affected by the temporal distribution of energy.

Two experiments to test this proposal were reported by Howell (1984). In the first experiment, the P centers of speech and nonspeech sounds were compared when the temporal distribution of energy was altered by ramping

the onsets at different rates; when the onset of the sound is gradual, there is comparatively little energy in the amplitude envelope at the beginning of the sound, and so the P center should be judged to occur late. This prediction was borne out by the data; with both speech and nonspeech, sounds with gradual onsets had the P center occur late in the syllable relative to sounds that had an abrupt onset (overall duration was the same in each case).

In the second experiment, the P-center phenomenon in speech production was examined. To keep the articulatory requirements as constant as possible while at the same time varying the temporal distribution of energy within syllables, speakers were asked to produce alternating sequences of long and short vowels. The prediction was that in each case, the energy would be distributed about the center of the syllable but that, because of the length differences, the judged P center would co-vary with syllable duration; the P center of a long syllable should occur later than that of a short syllable. This prediction too was upheld. Thus, the proposal appeared to account for all three groups of findings concerning P centers—its occurrence in the perception of speech and nonspeech sounds and in speech production.

Cooper et al. (1986) reported four experiments that supported the view that duration is a critical acoustic parameter in determining the location of the P center and that segment durations are differentially effective depending on the segment's location in the onset or rhyme of a syllable. Basically, their experiments involved manipulations on speech waveforms and assessments of P center locations after these manipulations. Such an approach (also employed by Marcus, 1981) is logically sound. Although not specifically addressed to an assessment of Howell (1984), three of their experiments (Experiments 2, 3, and 4) were interpreted as contrary to that theory.

The manipulations made in Cooper et al.'s (1986) experiments were to alter the rate of onset and duration of an initial frication noise, the duration of the vocalic portion, and the duration of silence interposed between the frication noise and vowel. Qualitative predictions about how these manipulations affect the distribution of energy can be derived and it can be seen whether the shifts in P center location are related to these. I contend that the distribution of energy within the syllable envelope will bias P center judgments. An estimate of the different distributions of energy within syllables can be obtained by comparing the location of the centers of gravity on the time axis. This essentially adds a metric to show how the temporal distribution of energy within the amplitude envelope affects the location of the P center.

The amplitude envelope of the frication in a syllable consisting of frication followed by a vowel (i.e., with no silence between) can be represented as a right triangle. Similarly, the envelope of the vowel can be represented by a rectangle (Figure 1). For an object with such a shape,

The author's mailing address is Department of Psychology, University College London, Gower Street, London WC1E 6BT, England.

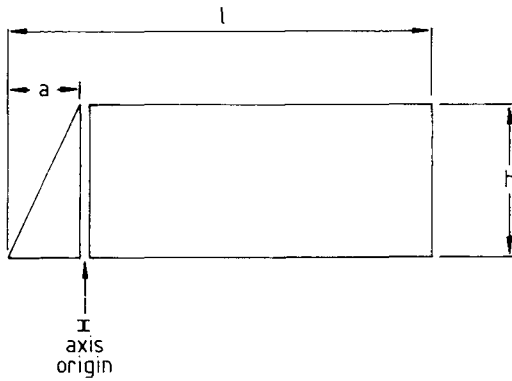


Figure 1. Stylized figure of the amplitude envelope of a fricative-vowel syllable. The fricative is represented by the rising portion and the vowel by the steady part.

it is straightforward to derive an expression for the center of gravity along the abscissa (time axis): The center of gravity (C of G) of the triangle and rectangle may be obtained separately, with the point on the abscissa where they meet defined as the origin. With these coordinates,

$$C \text{ of G} = \frac{(-\frac{1}{2}ah)(a/3) + ((l-a)/2)(l-a)h}{\frac{1}{2}ah + (l-a)h}, \quad (1)$$

where a = length of triangle along abscissa, h = height of triangle and rectangle along ordinate, and l = length of the polyhedron along abscissa (each of these is shown on the diagram in Figure 1). Because the height of the rectangle and triangle are the same in this situation, this simplifies to:

$$C \text{ of G} = \frac{2a^2 + 3l^2 - 6al}{6l - 3a}, \quad (2)$$

After the center of gravity of the object has been obtained in this coordinate system, it can be translated into values measured from the left-hand side of the shape (onset of the syllable envelope). By selecting suitable parameters, shapes with envelopes equivalent to those used by Cooper et al. (1986) can be obtained and the center of gravities obtained and compared. When this is done, this leads to the following intuitively obvious predictions: (1) when the triangle rises quickly to the same height as the rectangle and its overall duration is short, the center of gravity occurs nearer the onset of the sound relative to when the envelope rises gradually over a longer period of time (no alteration to rectangular portion). (2) Increasing the length of the rectangle with no alterations to the triangular part (increasing l) shifts the center of gravity right (away from the start of the syllable).

The effects of introducing a portion of silence between the frication and vocalic portion (between the triangle and rectangle) was assessed in a similar way (Figure 2). Using Equation 1 with the appropriate alterations for lengths along the abscissa by adding the extra parameter b for the duration of silence, it can be shown that (3) if there

is a part with no shape between the triangle and rectangle, the center of gravity shifts away from the start of the syllable.

If the triangle rises at the same rate but for a shorter time (as shown in Figure 3) using Equation 1, allowing the triangle to reach only some proportion of the height of the rectangle by introducing an extra parameter, again (4) the center of gravity occurs nearer the stimulus onset relative to the situation in which the triangle continues to rise until it reaches the same height as the rectangle.

The next goal is to see whether the qualitative features of Cooper et al.'s (1986) experiments can be predicted from the influence their experimental manipulations had on the temporal distribution of energy. To do this, the above predictions can be used to show how the center of gravity shifts as a result of the factors detailed above. Below, the properties of the fricative are modeled as alterations to the parameters of the triangular portion, the properties of the vowel as the rectangular portion, and silence between frication and vowel as a portion missing from the shape (as in Figure 2). (Inspection of Cooper et al.'s [1986] oscillograms shows that this is a reasonable way to model the stimuli they constructed.)

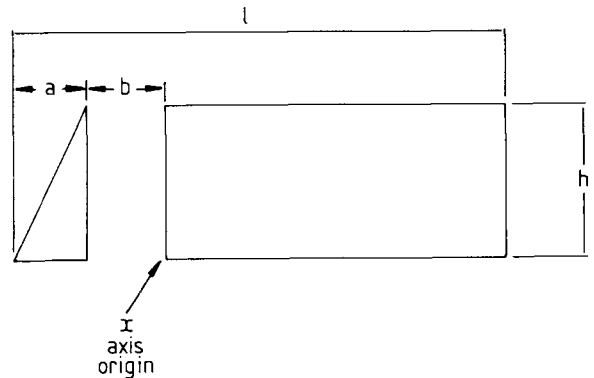


Figure 2. Amplitude envelope of fricative-vowel stimulus (as Figure 1) with a period of silence interposed between the fricative portion (triangle) and vowel (rectangle).

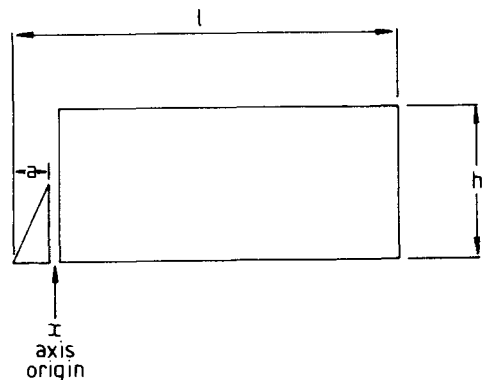


Figure 3. Stylized amplitude envelope of fricative-vowel syllable with the rise-time rate the same as in Figure 1. The duration of the portion that is rising has been reduced.

Experiment 1: Decreased friction duration and increased rise time rate. Cooper et al. (1986) found that as friction duration decreased and rise time rate increased, the P center occurred nearer the onset of the stimulus. This follows directly from Prediction 1 above.

Experiment 2: Introducing a portion of silence between friction offset and vowel onset. In this experiment, the P center shifted away from stimulus onset (tending to occur more toward the center of the syllable) as the duration of the interposed silence increased. This finding follows directly from Prediction 3.

Experiment 3: Part of the friction near the vowel exchanged for silence. With this manipulation, Cooper et al. (1986) found no shift in P-center location. This manipulation consists of two alterations to the stimulus: First, friction duration is decreased and the friction does not reach its full amplitude. As shown in Prediction 4, this causes the center of gravity to occur nearer stimulus onset than when the triangle goes to its full height and there is no silence between the triangle and rectangle. Second, the rising portion is shifted to the left and a period of silence is introduced. As stated in Prediction 3, this tends to make the P center occur away from the start of the syllable. These two effects (reduced duration of friction rise and insertion of silence) have opposite effects on the movement of the P center. Since these factors covary in Cooper et al. (1986), it is possible that they cancel each other out.

Experiment 4: Removal of medial portion of the vowel and insertion of a section of silence between the end of the friction and beginning of the vowel. The effect of shortening the vowel is to shift the center of gravity toward the start of the syllable (Prediction 2). The effect of interposing the silence is to shift the center of gravity away from stimulus onset (Prediction 3). Thus, the manipulations in this experiment have competing effects on the location of the P center. One possible result is that the two cancel each other out (Experiment 3). However, this did not happen. Instead, as silence was introduced, there was some shift in P-center location, though not an equal shift for every millisecond of vowel exchanged by silence, in contrast to Experiments 1 and 2 in which such a direct relationship was found. This may indicate that the reduction in length of the vowel (tends to cause the center of gravity to shift towards stimulus onset) is less effective than silence at shifting P centers.

Instead of quantitative predictions, only an ordering of P centers has been made. This was done for the following reasons: (a) The tenet of the theory is that perceptual judgments are affected by the temporal distribution of energy throughout a syllable. This does not mean that the center of gravity of a syllable is a physical measure of the "center" of the syllable. All that is implied is that the energy in all parts of a syllable affect alignments. Now it is necessary to derive the psychometric scale relating physical properties of the stimuli to judgments of P-center location. (b) Further to the preceding point, equal weight-

ing has been assigned to energy in all parts of the envelope (assuming equal amplitude). In the syllables that Cooper et al. (1986) employed, the energy at different points in a syllable has different spectral content (i.e., fricatives have high-frequency energy, vowels have comparatively low). It is likely that energy in different frequency regions determines the weight that should be given to the parts of the signal in assessing P-center location.

Part of Cooper et al.'s (1986) argument rests on comparison of the slopes of functions relating experimental manipulations and P-center shifts. So, for example, the slope between fricative duration and P-center location in Experiment 1 is unity, whereas the slope between gap duration and P-center location in Experiment 3 is less than unity. Cooper et al. (1986) use this to argue that P-center location is less affected by the vowel than by the initial consonant. In particular, they base this on a comparison of the effects of removing a millisecond of friction (Experiment 1) against a millisecond of vowel (Experiment 4).

This finding requires some comment because it seems to imply that their view can make more detailed quantitative predictions than the P-center orderings derived from the current proposal. The removal of a portion of friction in their Experiment 1 is accompanied by alterations of the unremoved friction that follows: thus, after so many milliseconds of friction have been removed, the following friction is ramped. The ramping method they employed alters in form depending on whether a long period of friction with a gradual rise or a short period with a steep rise is contoured (Howell & Rosen, 1983). Essentially, the form of the rise changes from approximately linear to positively accelerating (this can be seen in Cooper et al.'s [1986] own oscillograms, their Figure 1). Thus, the contouring technique they employed differentially affects the surrounding speech, depending on the number of milliseconds of speech excised. The abscissa in their Figure 3 (Experiment 1) is number of milliseconds removed at short durations (since the alteration to friction is negligible in this situation) with milliseconds removed plus a large alteration to the following friction when large portions of the friction are removed. What is puzzling is, then, the almost direct relationship between amount of "duration of fricative noise" removed and P-center shift. This suggests that if alterations due to rise per se were partialled out (which is known to have an effect even when overall duration is kept constant; Howell, 1984), the form of the curve would be asymptotic.

Corresponding alterations to the contour of the vowel were not made in Experiment 4. Thus, in that experiment, the form of the curve may be linear with a slope less than unity. However, since the contouring was applied only in Experiment 1, a comparison of the effects of amounts excised across the two experiments is dangerous.

The view outlined here differs from that proposed by Vos and Rasch (1981). They argued that the perceived onset of musical tones is determined by the time taken

for the envelope to exceed a threshold. This predicts that sounds with gradual onset will appear to start later than sounds with abrupt onsets. The present account also makes this prediction. However, the two views are not mimics of each other; for example, the views make different predictions about a situation in which the shape of the part of the musical waveform that is above Vos and Rasch's threshold is altered. The present account predicts that the P center should vary (see Prediction 4 above), whereas Vos and Rasch predict no shift when variations in amplitude occur in the portions of the sound above their threshold.

Although Cooper et al.'s (1986) own experiments do not appear to undermine the view that the distribution of energy in the amplitude envelope affects P-center judgments, they add two other arguments against the present view: First, they state that Tuller and Fowler (1981)¹ changed the amplitude envelope by infinite peak-clipping and found no shift in the P center. Second, Marcus (1981) noted that increasing the amplitude of the release burst of the final plosive in /t/ in the word "eight" does not affect the location of the P center although it alters the amplitude envelope.

In Cooper, Whalen, and Fowler's reply to this commentary, oscillograms of one pair of syllables that Tuller and Fowler employed are presented (made off of the tape used in the study). Clearly, these stimuli are not infinitely peak-clipped, and the processing performed has done curious things to different portions of the signal: there are still differences in rise time at syllable onset; at the end of the syllables, where the nonclipped sounds are falling in amplitude, the clipped ones appear to rise. Inspection of these oscillograms shows that there are durational differences between this pair of syllables. Indeed, the durational difference is more marked in the oscillograms of the pair of syllables shown in the article of Tuller and Fowler (1981). These factors would cause the P center of the infinitely peak-clipped set of stimuli to tend to be similar to the normal sounds.

Although Cooper et al. (1986) state that no shift in the P center was found after "infinite peak clipping" in the Tuller and Fowler study, it must be stated that Tuller and Fowler did not look very hard for a shift: "infinitely peak clipped" and not peak-clipped sequences were presented with intervals between syllables that were acoustically equal or corresponded with those that sounded rhythmically regular when spoken naturally. When forced to choose which way of presenting a pair of syllables made them sound more rhythmic, subjects chose the sounds separated by the natural interval. At present, it is not

known whether subjects would have adjusted infinitely peak-clipped and natural syllable pairs to have the same distance between onsets.

One finding that is difficult to deal with is Marcus's (1981) failure to find shifts in P center when the amplitude of the word-final burst was increased. However, an inspection of the oscillograms in his Figure 3 shows that the vowel portions of his stimuli were clipped. It is safest to ignore these data until the experiments are run with stimuli that are recorded properly.

All in all, Cooper and his colleagues' arguments against the view that distribution of energy within the amplitude envelope affects P-center location are not compelling. Since Cooper et al.'s (1986) experiments seem to have been the principal argument against an acoustic basis underlying Fowler, Smith, and Tassinary's (1986) experiments on syllable perception by prebabbling infants, the conclusions of the latter need reappraising.

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., SMITH, M. R., & TASSINARY, L. G. (1986). Perception of syllable timing by prebabbling infants. *Journal of the Acoustical Society of America*, **79**, 814-825.
- 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., & ROSEN, S. (1983). Production and perception of rise time in the voiceless affricate/fricative distinction. *Journal of the Acoustical Society of America*, **73**, 976-984.
- 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.
- 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.
- VOS, J., & RASCH, R. (1981). The perceptual onset of musical tones. *Perception & Psychophysics*, **29**, 323-335.

NOTE

1. When discussing infinite peak-clipping, Cooper et al. (1986) referred to, and cited, the wrong Tuller and Fowler paper. The correct reference to the infinite peak-clipping experiment performed by Tuller and Fowler is Tuller and Fowler (1981).

(Manuscript received August 4, 1986;
revision accepted for publication May 22, 1987.)