An Acoustic Determinant of Perceived and Produced Anisochrony

P. Howell
London, United Kingdom

1. Introduction

Morton, Marcus and Frankish (1976) reported that digits presented with the same distance between onsets are not perceived as isochronous (equally timed). If subjects are allowed to position them so that they appear to be regular, the adjustments when measured between stimulus onsets, vowel onsets or peak amplitude value in the syllable are not equal. That is, the stimuli have to be physically anisochronous to appear perceptually isochronous. Marcus (1981) has examined what acoustic factors determine p-center location by editing speech to see what factors cause the p-center to vary. He found that varying the duration of the initial consonant of a syllable, lengthening its vowel and extending the period of closure before release of a syllable-final stop affected judgments about perceptual isochrony but that altering the amplitude of a final burst did not affect such judgments. Marcus interpreted his findings as showing that several acoustic factors determine p-center location.

A finding that may be related to those concerning perceptual isochrony is that when speakers are asked to produce isochronous lists, the timing corresponds to that which would be needed in order for the items to be perceived as isochronous (Fowler, 1979). Tuller and Fowler (1980) showed that certain syllables appear to be aligned with respect to orbicularis oris muscle activity. Fowler and her associates (Fowler, 1979; Tuller and Fowler, 1980) consider that the regularity in timed speech activity occurs in production with respect to activity in certain muscle groups and departures from physical isochrony in perception occur because perception is referred to production. Thus, perceptual adjustments do not align with respect to any acoustic referent because of the complex relationship between articulation and the sound produced. These results demonstrate that the acoustic onset of a syllable is not the same as the onset of the sound during production or perception. But before we accept that there are complex acoustic or productive determinants of p-center location, simple acoustic determinants should be ruled out. The criteria for a satisfactory factor that determine the location of the p-centers are, first, that it should vary in alignment across stimuli in the same way that the perceptual judgments do. Second, it should vary in location relative to stimulus onset in the same way that perceptual alignments vary when the acoustic properties of test stimuli are altered. Third, the factor should account for why the phenomenon occurs in perception and production.

To date, the acoustic factors that have been examined as candidates for determining p-center location have been acoustic reference points within a syllable, not acoustic factors associated with the syllable itself (e.g., reference points associated with the vowel). The principal acoustic factor that is associated with the syllable is the amplitude envelope (Mermelstein 1975). It is not directly related to any of the acoustic factors examined. So, for example, the parameters of the amplitude envelope are not fixed relative to acoustic factors associated with the vowel.

2. Experiments

2.1. Speech

The intention of the first experiment is to see whether variation in the amplitude envelope is a sufficient cue to cause variation in p-center location in perception with speech and non-speech sounds. The speech stimuli employed varied in the amplitude envelope alone. Thus, if p-center location varies it can only be attributable to this factor. Since variation in amplitude envelope can occur with non-speech sounds, there may be differences in the p-center of non-speech too.

To test this, different envelopes were introduced onto speech sounds by contouring the onset of a naturally spoken /ʃa/ to produce /tʃa/ (short rise) or /ʃa/ (long rise). The procedure only affects the envelope and, if the present account is correct, should be sufficient to cause variation in p-centers. To construct the stimuli, a recording was made of a /ʃa/ spoken by a male adult. Both the fricative and vowel were sustained so that they remained at the same amplitude for some time. The next step was to truncate the frication to 148.8 ms (measured back from vowel onset). It was then contoured by multiplying by a linear ramp of 40 ms and left at its original amplitude for the remainder of the frication. The rise of the stimulus with a value of 120 ms was constructed in the same way except that it was ramped over the first 120 ms. In each case the vowel was tapered by a linear ramp over 312 ms.

2.2. Nonspeech

The non-speech stimuli were constructed from a portion of white noise followed by a portion of sawtooth waveform. The noise and sawtooth were approximately the same peak-to-peak amplitude as the aperiodic and periodic portions had been in the recording of the original speech syllable. The contours of the speech stimuli with 40 and 120 ms rises were calculated from the rectified digital waveform low-pass filtered at 25 Hz (Fant, 1959).

These were used to contour the non-speech stimuli by multiplying the calculated contour by the non-speech stimulus just described. Subjects had to adjust the timing between one member of a pair of stimuli which were played repeatedly until they appeared to be perceptually regular. One of the intervals between the members of a pair was fixed at 750 ms and the duration of the other could be varied by means of a potentiometer. When the subject was satisfied that the items were regularly timed, he pressed a key and the duration of the adjusted interval was stored in the computer. The next pair was then selected and tested in like manner. On each trial the null position of the potentiometer was varied randomly and the subject was told this and informed that he would need to alter the position of the potentiometer. The stimulus pairings could both be the same (40/40 and 120/120) or differ (40/120 and 120/40) for both types of material. This gave four sorts of trial and there were three repetitions of each in a block of twelve adjustments. Altogether twelve subjects performed the experiment - six with the speech sounds and six with the non-speech sounds. Mean adjustments across subjects are presented in Table I separately for each type of adjusted interval. Inspection of the table shows that there is little difference in adjustments when stimulus pairs consisted of the same stimulus but with stimulus pairs with different envelopes at onset, longer adjustments were made when the interval between the stimulus with the slow rise at onset and that with the quick rise at onset was varied for both speech and non-speech sounds. The different adjustments were not significant by analysis of variance for stimulus pairs that were the same but were for both the speech and non-speech stimuli when the pairs had different envelopes at onset. ($F(1,5) = 39.1$, $p < 0.005$ and $F(1,5) = 15.5$, $p < 0.05$ respectively).

3. Discussion

Thus, altering the distribution of energy in the envelope is sufficient to cause variation in p-center location for both speech and non-speech stimuli. It is possible that amplitude envelope is just one other factor that determines p-center location (cf. Marcus, 1981). However, all Marcus's manipulations which were effective in altering p-center location are effective in altering the distribution of energy in the amplitude envelope. Another finding of note is

Table I. Mean duration of adjusted intervals (in ms) in Experiment 1

Speech			
All /tʃa/	All /ʃa/	Mixed lists /tʃa/-/ʃa/ adjusted	/ʃa/-/tʃa/ adjusted
784	811	733	831
Non-speech			
All 40 ms	All 120 ms	Mixed lists 40 and 120 ms	120 and 40 ms
744	776	722	798

that variation in p-center location occurs for both speech and non-speech despite earlier claims to the contrary (Lehiste, 1973, Morton et al. 1976, though see Vos and Rasch, 1981 for another report of variation in p-center location of non-speech stimuli differing in amplitude envelope).

The correlation between the anisochronies in production and those in perception might occur because subjects judge the timing of their productions from the distribution of energy in the amplitude envelope of their own speech. If so, alteration to the distribution of energy in the amplitude envelope should affect their ability to produce isochronous sequences. Variations in the envelope during production should cause subjects to position syllables anisochronously.

This can be tested by requiring subjects to speak vowels that vary in length. A sustained vowel will have the energy in its envelope late relative to that of a short vowel. Thus speakers should advance the onset of the long vowel when it is spoken in alternation with a short vowel to make their onsets appear regular if subjects use the amplitude envelope to make this judgment. The ratio of the interval from the onset of a short vowel to the onset of a long vowel to the interval from the onset of a long vowel to the onset of the next short vowel in repeated vowel pairs should, then, be less than one, greater than one when the order of long and short vowels is reversed and equal to one when the vowels are of the same length. This prediction was tested in the same experiment. The vowels /i/ and /u/ were paired together or with themselves. With alternating vowel pairs the order of vowels was counterbalanced giving four sets of vowels in all. Each of the vowels could be of short, medium or long duration giving nine durations per vowel set (3x3). Eight subjects were told to repeat each of the vowel sets so that their onsets appeared regular. The mean duration for each set was calculated (omitting the initial and final vowel pair). The predicted ratios were obtained for all stimulus sets and there was no statistical difference between vowel sets.

4. Conclusion

It appears, then, that the amplitude envelope of speech and non-speech is an important factor in determining p-center location. Variations in this factor alone give differences in the location of p-centers in speech and non-speech. In addition, Marcus's (1981) manipulations affect the amplitude envelope and data have been presented showing that subjects may use the amplitude envelope to judge the p-centers of the productions.

References

Fant, G.M.C. (1959). Acoustic analysis and synthesis of speech with applications to Swedish. *Ericsson Technics,* **1**, 1-105.

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

Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America,* **51**, 2018-2024.

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

Mermelstein, P. (1975). Automatic segmentation of speech into syllabic units. *Journal of the Acoustical Society of America,* **58**, 880-883.

Morton, J., Marcus, S., and Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Rview,* **83**, 405-408.

Tuller, B., and Fowler, C.A. (1980). Some articulatory correlates of perceptual isochrony. *Perception and Psychophysics,* **27**, 227-283.

Vos, J., and Rasch R. (1981). The perceptual onset of musical tones. *Perception and Psychophysics,* **29**, 323-335.