

SEGMENTS, SYLLABLES, AND THE PERCEPTION OF SPEECH RATE AND RHYTHM

B. Pompino-Marschall*

ABSTRACT

The mean distance of measured syllabic P-centers from one another in complex syllable sequences is shown to be a good indicator of perceived rate of speech. The variation in P-center position in simple synthetic CV-syllables in turn is shown to be dependent on the segmental composition of these syllables in a very complex fashion, relevant factors being the duration of the consonant and the vowel, the amplitude envelope, as well as the spectral composition of the syllable.

INTRODUCTION

Alternating sequences of monosyllables, when presented with equal intervals between successive acoustical onsets are not perceived as having a subjectively uniform rhythm, because the perceived onset (P-center) of a syllable typically does not correspond to its acoustical onset. Similarly, perception of 'momentary tempo' in syllable sequences seems to be dependent rather on the interval between vowel onsets than on the interval between syllable onsets (ref 11). Generally, it is assumed that the location of the P-center is dependent on the duration of the initial consonant(s) and that of the syllable rhyme as represented by a linear equation proposed by Marcus (ref 5).

In the following experiments we wanted to test this latter hypothesis and whether the perception of 'momentary tempo' can be explained on the basis of the P-center phenomenon.

STIMULI AND METHOD

The stimuli for the P-center experiments were on the one hand simple synthetic /ma/- and /fi/-syllables with systematically varying consonant and vowel durations: two sets of 25 CV-syllables with consonants varying in duration from 40 to 200 ms and vowels varying from 100 to 260 ms, both in steps of 40 ms. Furthermore, five /mam/-syllables were synthesized with the following segment durations: 40 ms /m/, 260 ms /a/, 80 ms /m/; 80, 220, 120; 120, 180, 160; 160, 140, 200; and 200, 100, 240. Vowel duration always included two symmetrical CV-(and VC-)transitions of 40 ms each. Fundamental frequency was set at 100 Hz for the entire periodical part of the stimuli, and the amplitude was held constant over the steady-state parts. These syllables were synthesized on a PDP 11/50 with a program based on the Klatt software synthesizer (ref 3). On the other hand, the /ma/- and /mam/-syllables were paralleled with respect to dB-envelope by 100-Hz rectangular signals, while the original /fi/-syllables were paralleled by /fi/-syllables with rectangular sound pressure envelopes.

For the last pilot experiments two seven step /sta/-/spa/-continua were synthesized: The /s/ was of 230 ms duration (including a 45 ms implosive transition), the /a/ had a duration of 220 ms (including the variable 45 ms explosive transition); in the first set, the silent stop interval also varied from 45 to 75 ms in steps of 5ms, in set two it was held constant at 60 ms. To determine the position of the P-center the subjects had to adjust the timing of these syllables alternating with clicks (5-ms 1-kHz-tone bursts) in sequences of five signals with an overall rate of 120 signals per minute to perceived isochrony by turning a potentiometer knob. We decided to use the time instant bisecting the duration between two successive clicks to determine the P-center of the test syllable.

*Institut für Phonetik und Sprachliche Kommunikation der Ludwig-Maximilians-Universität München, Schellingstr. 3/II/VG, D-8000 Munich 40, Federal Republic of Germany.

All stimuli were adjusted in alternating sequences beginning with a click signal, as well as in sequences beginning with the syllable itself six times in each session. The extreme adjustments were omitted from the analysis. There were five (two in the pilot experiments) sessions for every stimulus, resulting in 40 (16 in the pilots) adjustments for every subject (2 sequences * 4 adjustments * 5 (2) sessions).

For the 'running tempo' experiment the /mam/-syllables and their nonspeech counterparts were concatenated to a five-item sequence paralleling those of Ventsov (ref 11) and ours (ref 6, 7), yielding a sequence of open syllables of 300 ms and closed syllables of 340 ms duration (round vs square brackets):

(40m+[260a])+(80m)+[220a]+(120m)+[180a]+(160m)+[140a]+(200m)+[100a]+240m]

To determine the perceived 'running tempo' of this sequence it was paired with click sequences of varying click onset intervals ranging from 280 to 360 ms in steps of 20 ms. 14 subjects had to judge 10 randomized presentation of these five pairs as being same or different with respect to rate.

RESULTS AND DISCUSSION

Mean P-center distances and 'running tempo'

In the first part of this experiment with two subjects the P-center location of the /mam/-syllables and their nonspeech analogues was determined. The pooled results (80 adjustments) are shown in Table I.

Table I:
Mean P-center locations (and sds)
in ms

segment duration /m, a, m/	speech	nonspeech
40, 260, 80	90.78 (16.78)	84.44 (21.31)
80, 220, 120	135.75 (16.98)	133.29 (14.02)
120, 180, 160	164.85 (13.97)	164.96 (15.76)
160, 140, 200	196.01 (16.33)	192.03 (17.84)
200, 100, 240	227.81 (13.5)	224.23 (22.12)

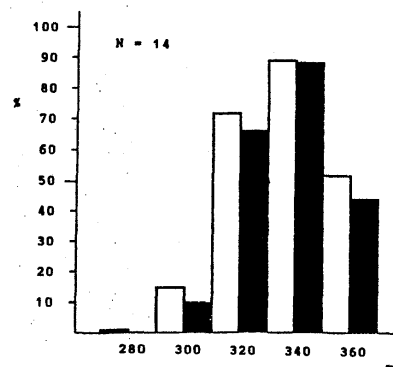


Fig. 1: Percent 'same'-responses to pairs of the /mam.../-sequence (open columns) or their dB-paralleled counterpart (filled columns) and click-sequences with different onset times (abscissa).

The results of the 'running tempo' judgements are shown in Figure 1. For both sets of stimuli the pooled median of the 'same'-response distribution lies near the duration of the closed syllable, but significantly differing from it ($p < .001$; speech: 336. ms, sd = 6.62; nonspeech: 336.3, sd = 6.02). These values are approximated best by the mean P-center distances of the component stimuli (speech: 334.26, sd = 7.24; nonspeech: 334.95, sd = 9.55).

The influence of segment durations on P-center location

The results of the adjustment experiments pooled over three subjects (120 adjustments for each stimulus) for the 25 /ma/-syllables and their nonspeech counterparts are shown in Figure 2 left, those for the original

/fi/-syllables and their loudness-equalized counterparts (ref 2) in Figure 2 right.

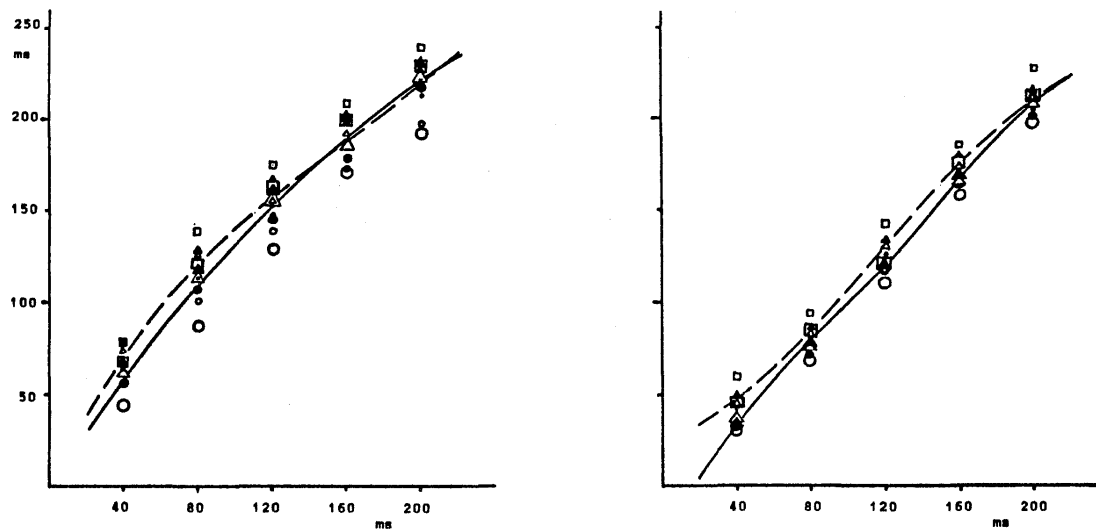


Fig. 2: Variation in P-center location (ordinate) due to the duration of the initial segment in /ma/-syllables (left, small symbols, dashed curve of best fit), their dB-paralleled counterparts (left, large symbols, unbroken curve) and in /fi/-syllables (right; small symbols, dashed curve: original; large symbols, unbroken curve: loudness equalized) with different durations of the final segment (open circles: 100 ms, filled circles: 140 ms, open triangles: 180 ms, filled triangles: 220 ms, open rectangles: 260 ms).

For all sets of stimuli analysis of variance revealed a significant ($p < .01$) effect of initial consonant duration: the longer this segment the larger is the displacement from acoustic isochrony in sequences perceived as rhythmically uniform. These effects are all significantly nonlinear. The effects are clearly different for the different sets of stimuli: The anisochrony being strongest with the /ma/-syllables, lesser for their nonspeech analogues, lesser for the original /fi/-syllables, and least for their loudness-equalized counterparts. The same rank order can be seen for the significant effect of vowel duration: The effect of vowel variation being strongest for the /ma/-syllables. In contrast to the /fi/-materials the /ma/-materials also show a significant interaction between the consonant and vowel effect. There is always a significant interaction between the vowel effect and the parameter original stimuli vs derived material as well as between the consonant effect and this latter variation (for the /fi/-material only at $p < .05$).

The results clearly show that the location of the P-center cannot be accounted for by a simple linear combination of two linear effects of consonant and rhyme duration as proposed by Marcus (ref 5). A psychoacoustic model, which takes the rising auditorily filtered sound pressure envelope as the determining parameter for the location of 'perceptual moments' (ref 10, 4), would clearly better fit the general tendencies in the results, but it has to be extended by spectrally induced 'perceptual moments' to account for the observed differences between the different materials.

Independence of P-center location of phonetic categorization?

In the last pilot experiments we wanted to replicate an experiment of Cooper et al. (ref 1) with different material. They found that P-center location in categorically perceived continua is solely linearly dependent on initial consonance duration.

The results pooled over two subjects are shown in Figure 3. In the upper part identification (20 repetitions) and discrimination (10 repetitions of two-step AX-pairs) functions are displayed. The lower part shows the results of 32 adjustments to uniform rhythm.

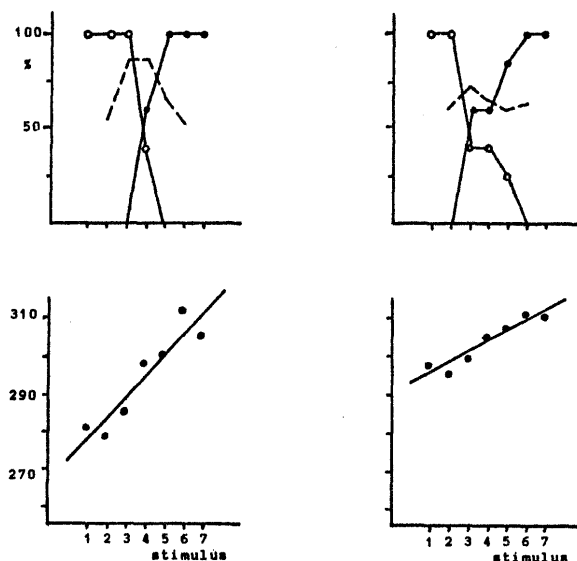


Fig. 3: Categorization (upper part; open circles: /sta/, closed circles: /spa/, dashed: discrimination) and P-center location (lower part) of two different /sta/-/spa/-series (left: with variation in explosive transitions and duration of the silent interval; right: without durational variation).

The results for the series varying in explosive transition and the duration of the silent interval are consistent with the findings of Cooper et al. (ref 1): The significant ($p < .01$) effect of initial consonance duration on P-center displacement is clearly linear in the magnitude of 1:1 and without any discontinuities at the category boundary, but also for the series without any durational variation there is a significant effect of the stimulus on P-center location. The effect again is linear in spite of categorical perception, but here the slope of the regression line is only .53. Whether this latter result is an effect of the spectral composition of the test items or due to the normally given covariation of spectral and temporal cues in natural productions is an open question for the time being.

Generally, it can be stated that parallel to the effects seen in rhythmic productions (ref 8) the perceptual P-center is dependent on a number of variables in quite a complex fashion (see also ref 9).

REFERENCES

1. A M Cooper et al., *Percept.Psychophys.* 39, 187 (1986)
2. P Janker, *Experimentelle Untersuchungen zum P-center Effekt* (in prep.)
3. D H Klatt, *J.Acoust.Soc.Amer.* 67, 971 (1980)
4. M Köhlmann, *Rhythmische Segmentierung von Schallsignalen und ihre Anwendung auf die Analyse von Sprache und Musik* (Dr.-Ing.diss., München, 1984)
5. S M Marcus, *Percept.Psychophys.* 30, 247 (1981)
6. B Pompino-Marschall et al., *Phonetica* 39, 358 (1982)
7. B Pompino-Marschall et al., in: M P R v d Broecke & A Cohen (ed.), *Proc. Tenth Int.Congr.Phon.Sci.* (Foris, Dordrecht, 1984) p 537
8. B Pompino-Marschall, H G Tillmann, in: *Proc.Eleventh Int.Congr.Phon.Sci.* (Tallinn, 1987, in print)
9. B Pompino-Marschall, H G Tillmann, B Kühnert, in: *Proc.Eleventh Int.Congr. Phon.Sci.* (Tallinn, 1987, in print)
10. H Schütte, *Biol.Cybernet.* 29, 49 (1978)
11. A V Ventsov, *Phonetica* 38, 193 (1981).