P-CENTERS AND THE PERCEPTION OF 'MOMENTARY TEMPO'

BERND POMPINO-MARSCHALL HANS G. TILLMANN BARBARA KÜHNERT

Institut für Phonetik und
Sprachliche Kommunikation
Universität München, FRG

ABSTRACT

Variation in p-center location is shown to be dependent on the segmental composition of a syllable in a far more complex way than assumed by the model of Marcus [2] that represents a linear combination of two different linear effects: of the duration of the syllable-initial consonance and of the syllable rhyme. The mean distance of measured P-centers from one another in syllable sequences is shown to be a good indicator of perceived rate of speech.

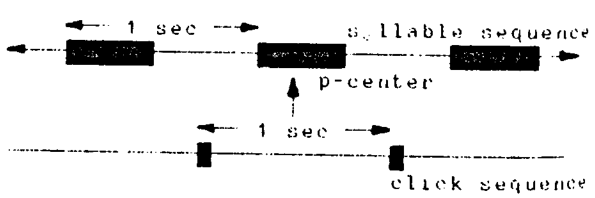
INTRODUCTION

Alternating sequences of monosyllables, when presented with equal intervals between successive acoustical syllable onsets (so-called isochronous sequences) are not perceived as having a subjectively uniform rhythm because the perceived onset (P-center) of a syllable typically does not correspond to its acoustic onset. Generally, it is assumed that the location of the P-center of monosyllables is solely dependent on the duration of the initial consonant(s) and that of the syllable rhyme as represented by a linear equation proposed by Marcus [2].
In the following experiments we wanted to test this hypothesis with systematically varied material.

EXPERIMENT I

Method

In simple synthetic syllables composed of the consonant /m/ and the vowel /a/ segment durations were varied systematically:
25 /ma/-syllables with /m/ varying in duration from 40 to 200 msec in steps of 40 msec and /a/ varying from 100 to 260 msec also in steps of 40 msec and 25 /am/-syllables with /a/ varying as before and /m/ from 80 to 240 msec, again in steps of 40 msec. Finally 5 /mam/-syllables were synthesized with the following segment durations: 40 msec /m/, 260 msec /a/, 80 msec /m/: 80, 220, 120; 120, 180, 160; 160, 140, 200; and 180, 100, 240. Vowel duration included two symmetrical transitions (from and to /m/) of 40 msec each. Fundamental frequency was set at 100 Hz for the entire duration of the stimuli and amplitude was held constant over the steady-state parts. The syllables were synthesized on a PDP 11/50 with a program based on the Klatt software synthesizer [1].
In perception experiments the subjects had to adjust the timing of these syllables alternating with clicks (5-msec 1 kHz-tone bursts) in sequences of five signals with an overall tempo 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 was used to determine the location of the p-center of the test syllable:



All stimuli were adjusted by two subjects 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 sessions for every stimulus, resulting in 40 adjustments (2 sequences * 4 adjustments * 5 sessions).

Results

The results of the adjudstments for the /ma/-syllables pooled over both subjects are seen in Figure 1. Two-factorial analyses of variance revealed highly significant effects of both duration of the initial /m/ and of the vowel /a/ as well as a significant interaction of the effects on P-center location for both subjects:

B.P.M. (male): $F(4,975) = 1432.36$; $p < .001$; $F(4,975) = 70.67$; $p < .001$; $F(16,975) = 4.19$; $p < .001$;
B.K. (female): $F(4,975) = 7453.57$; $p < .001$; $F(4,975) = 769.23$; $p < .001$; $F(16,975) = 19.64$; $p < .001$.

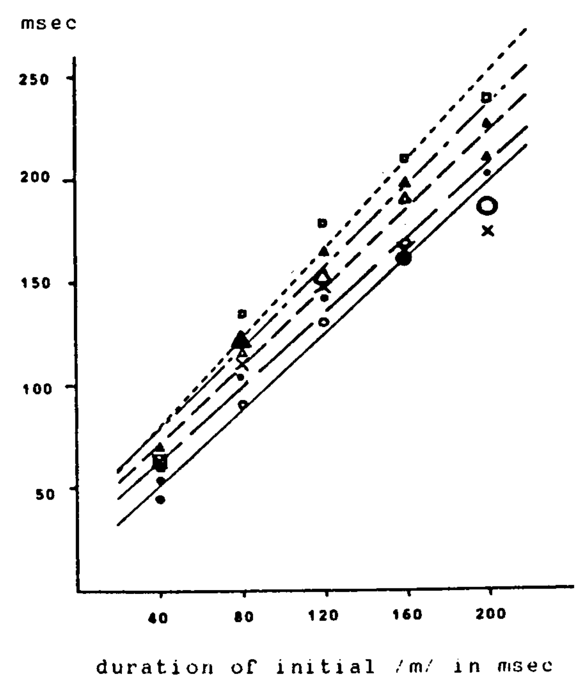


Fig. 1: Variation in p-center position (ordinate) due to the duration of the initial /m/ (abscissa) with different /a/-durations: open circles, unbroken regression line: 100 msec /a/; filled circles, long dashed: 140 msec; open triangles, short dashed: 180 msec; filled triangles, dash-dotted: 220 msec; open rectangles, dotted: 260 msec; enlarged symbols represent those stimuli dB-paralleled in the psychoacoustic experiment which are represented as crosses.

The simple main effects of consonant and vowel duration are shown in Table I and II for both subjects individually along with the parameters of the regression lines and the levels of significance for linear and nonlinear components of trend. As can be seen, quite generally there are significant nonlinear components speaking against the hypothesis of Marcus [2].
For the /am/-syllables, the results are shown in Figure 2 and Tables III and IV in the same way. Here too, we get significant effects of vowel duration, of final consonant duration and an interaction of both effects for both subjects:
B.P.M.: $F(4,975) = 71.15$; $p < .001$; $F(4,975) = 24.24$; $p < .001$; $F(16,975) = 6.08$; $p < .001$;

Table I:
Simple main effects of initial /m/-duration, parameters of the regression lines and significance of linear and nonlinear trend components

duration of /a/ in msec	level of significance	r	a	b	lin.	non lin
100	***	.92	19.59	.97	***	***
	***	.97	9.62	.85	***	***
140	***	.89	32.92	.94	***	*
	***	.97	20.49	.84	***	***
180	***	.91	36.6	.97	***	***
	***	.97	30.6	.9	***	***
220	***	.91	39.74	1.04	***	**
	***	.97	38.27	.91	***	***
260	***	.93	39.62	1.09	***	***
	***	.97	31.57	1.05	***	***

(r, a, b, lin., non lin: analysis of trend)

here and in the following tables: first line subject B.P.M., second line subject B.K.;
***: $p < .001$; **: $p < .01$; *: $p < .05$;
- : n.s.

Table II:
Simple main effects of /a/-duration, parameters of the regression lines and significance of linear and nonlinear trend components

duration of /m/ in msec	level of significance	r	a	b	lin.	non lin
40	***	.23	44.14	.11	***	*
	***	.65	25.34	.15	***	***
80	***	.47	83.28	.22	***	*
	***	.87	48.47	.31	***	***
120	***	.58	107.69	.29	***	-
	***	.87	94.07	.29	***	*
160	***	.48	157.69	.23	***	***
	***	.86	109.08	.35	***	***
200	***	.53	175.27	.28	***	-
	***	.92	135.15	.36	***	-

(r, a, b, lin., non lin: analysis of trend)

B.K.: $F(4,975) = 459.47$; $p < .001$; $F(4,975) = 362.79$; $p < .001$; $F(16,975) = 2.22$; $p < .01$.
As before we have quite a number of nonlinear effects. Furthermore the syllable rhyme seems not to be an integral part

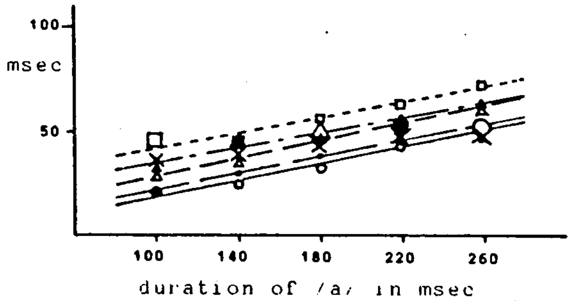


Fig. 2: Variation in p-center position due to the duration of /a/ (abscissa) with different final /m/-durations: open circles, unbroken regression line: 80 msec /m/; filled circles, long dashed: 120 msec; open triangles, short dashed: 160 msec; filled triangles, dash-dotted: 200 msec; open rectangles, dotted: 240 msec; enlarged symbols: dB-paralleled (crosses).

with respect to the determination of p-center location. As in open syllables, there are different influences of vowel and consonant durations interacting with one another in a complex fashion.
The results for the /mam/-syllables pooled over both subjects is shown in Figure 3.

EXPERIMENT II

Method

Some of the stimuli (those marked in Figure 1 and 2 by enlarged symbols and the /mam/-syllables) were paralleled with respect to dB-envelope by 100-Hz rectangular signals to test the influence of the envelope-parameter. Adjustments were done by the two subjects of Experiment I in the same fashion.

Results

The results for these nonspeech analogues are marked by crosses in Figure 1 - 3. Analyses of variance for both subjects individually show a speech-nonspeech effect for /ma/-syllables only for B.K. (p < .001: the nonspeech analogues always showing smaller p-center delays), for /am/-syllables for both subjects (p < .01 and < .001 respectively; again smaller delays for the nonspeech stimuli) and for the /mam/-syllables also for both subjects (p < .001) but here with different orientation (B.K. as before, B.P.H. with longer p-center delays for the nonspeech material). These latter results deserve further testing for a final interpretation.

Table III:
Simple main effects of /a/-duration, parameters of the regression lines and significance of linear and nonlinear trend components

duration of final /m/	level of significance	r	a	b	analysis of trend: lin.	non lin
80	***	.6	-4.32	.24	***	*
	***	.78	.01	.17	***	-
120	***	.52	3.58	.2	***	***
	***	.78	3.83	.18	***	*
160	***	.49	13.03	.19	***	**
	***	.85	2.92	.22	***	-
200	***	.33	21.33	.13	***	*
	***	.81	13.18	.22	***	-
240	***	.33	29.76	.14	***	**
	***	.81	17.77	.22	***	-

Table IV:
Simple main effects of final /m/-duration, parameters of the regression lines and significance of linear and nonlinear trend components

duration of /a/	level of significance	r	a	b	analysis of trend: lin.	non lin
100	***	.42	5.9	.16	***	***
	***	.78	3.51	.15	***	**
140	***	.34	18.29	.11	***	*
	***	.79	9.27	.16	***	-
180	***	.31	26.82	.11	***	***
	***	.77	16.2	.18	***	-
220	***	.09	48.23	.03	-	**
	***	.78	21.25	.19	***	-
260	***	.16	47.67	.06	**	***
	***	.75	26.03	.21	***	-

EXPERIMENT III

In a last experiment the /mam/-syllables and their nonspeech counterparts were concatenated to a five-item sequence paralleling those of Ventsov [6] and ours [3, 4] in the following form, yielding a sequence of open syllables of 300 msec and closed syllables of 340 msec duration:

40m260a80m220a120m180a160m140a200m100a240m

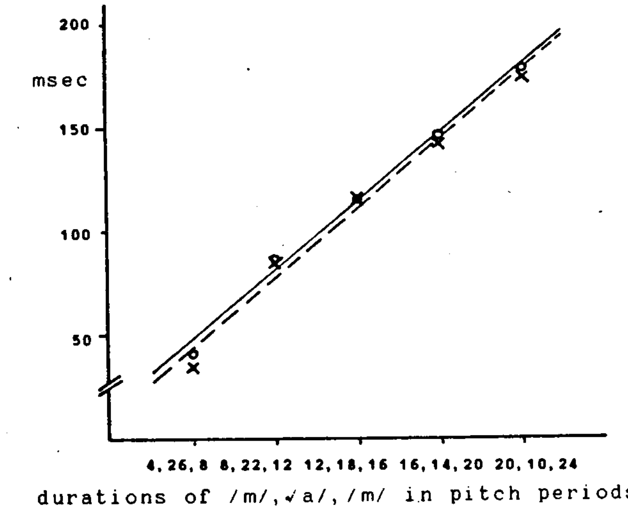


Fig. 3: P-center position (ordinate) due to the duration of syllable segments (abscissa): initial /m/-duration, /a/-duration, final /m/-duration (circles); dB-paralleled stimuli: crosses.

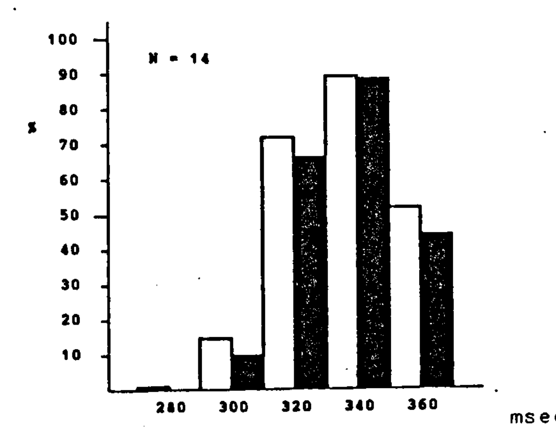


Fig. 4: Percent 'same'-responses to the /mam .../-sequence (see text; open columns) and dB-paralleled sequence (filled columns) in combination with click-sequences of variable onset intervals

This sequence was combined in pairs with a click sequence of varying click onset intervals (from 280 to 360 msec in steps of 20 msec). 14 subjects judged 10 randomized presentations of these five pairs as being same or different with respect to rate. The results are shown in Figure 4. For both stimulus sets 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. msec, sd = 6.62; nonspeech: 336.3, sd = 6.02). These values resemble the mean p-center distances of the used stimuli (speech: 334.26, sd = 7.24; nonspeech: 334.95, sd = 9.55).

DISCUSSION

Our results clearly show that the location of the p-center is not simply dependent on segmental durations of the tested syllable. The psychoacoustic model of Schütte [5] which takes the rising auditorily filtered sound pressure envelope as the p-center determining parameter gives a better prediction of the p-center location for our /ma/-syllables. But it does not explain why there are different p-center locations in the set of /am/-syllables (because they all have the same rising envelope). And also the speech-nonspeech differences in our results lead us to the conclusion that the p-center location in speech material cannot be accounted for by a pure psychoacoustic modelling.

REFERENCES

[1] Klatt, D.H. 1980, Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America 67, 971-995.

[2] Marcus, S.M. 1981, Acoustic determinants of perceptual center (P-center) location. Perception & Psychophysics 30, 247-256.

[3] Pompino-Marschall B; Piroth, H.G.; Hoole, P.; Tilk, K; Tillmann, H.G. 1982, Does the closed syllable determine the perception of 'momentary tempo'? Phonetica 39, 358-367.

[4] Pompino-Marschall, B.; Piroth, H.G.; Hoole, P.; Tillmann, H.G. 1984, 'Koartikulation' and 'Steuerung' as factors influencing the perception of 'momentary tempo'. In: Van den Broecke, M.P.R.; Cohen, A. (ed.), Proceedings of the Tenth International Congress of Phonetic Sciences (Dordrecht, Cinnaminson - Foris), 537-540.

[5] Schütte, H. 1978, Ein Funktionsschema für die Wahrnehmung eines gleichmäßigen Rhythmus in Schallimpulsfolgen. Biological Cybernetics 29, 49-55.

[6] Ventsov, A.V. 1981, Temporal information processing in speech perception. Phonetica 38, 193-203.