# Abstractness in Speech-Metronome Synchronisation: P-Centres as Cyclic Attractors

*Plínio A. Barbosa, Pablo Arantes, Alexsandro R. Meireles, and Jussara M. Vieira*

Speech Prosody Studies Group, Department of Linguistics
State University of Campinas, Brazil
plinio@iel.unicamp.br

## Abstract

The p-centre phenomenon is revisited here under a dynamical paradigm. It is our contention that p-centres can be seen as the projection onto the time axis of an underlying phrase stress oscillator inferred during the realisation of the task of producing (and perceiving) a sequence of syllables in time with a metronome. In this kind of inference, vowel onsets act as point attractors for the task of synchronisation. The p-centre, defined as the metronome pulse position relative to the syllable, is considered as the attractor predicted by the subject. The innovative aspects in the experiments carried out to sustain these hypotheses refer to: taking into account as many as 21 Brazilian Portuguese CV syllables; exploring the dynamical aspects of the task, such as variation of the subject's performance due to rate changes; and explaining the subject's behaviour based on period and/or phase locking between syllable production and inferred-from-metronome phrase stress oscillator. The phase differences between the p-centres and the vowel onsets are treated as on-line perturbations of energy rise detected on account of the consonants' spectral composition.

## 1. Introduction

This paper aims at revisiting the so-called p-centre phenomenon as a consequence of the perturbation of the subject's ability to anticipate vowel onset position from the sequence of syllable s/he produces and to use these temporally predicted anchoring points in order to utter syllables in time with a metronome. It will be sustained that this prediction is achieved by the simulation of a cyclic attractor, more specifically a phrase stress oscillator, whose behaviour changes depending on the rate of the sequence produced. On the other hand, it will be shown that the deviations from vowel onset position can be explained by differences in the spectral tilt of (onset) consonants.

It has been repeatedly demonstrated in Experimental Psychology that p-centres reflect the so-called "moment of occurrence" of a syllable or monosyllabic word [1, 2, 3], and that they are the actual anchor points of the production and perception of regularly-timed syllable sequences, instead of syllable onsets. One of the methodological consequences of this discovery is reflected in the way scholars delimit stress groups along the utterances [4, 5]. (The fact that this practice can be traced back to early works such as Classe's [6], only reinforces the strong perceptual attraction of these anchor points.)

The experimental tasks carried out to compute the location of p-centres begun in the early nineteen seventies [5, 7] and continued in the late nineties to be a theme of heated debate [8, 9]. To our knowledge, the issue of the abstractness of time relations needed to produce/perceive synchronous speech has not been properly raised, despite this span of time. The consideration of the level of abstractness involved in the tasks reviewed here shows that p-centres are in fact a surface phenomenon of an underlying specification of vowel onsets.

### 1.1. Previous findings

The experimental tasks carried out to compute p-centres usually took into account speakers or listeners of Germanic languages and comprised either (1) the synchronisation of a sequence of syllables with a non-speech stimulus, such as a metronome [10, 11] or finger taps [5, 8]; or (2) the production, or the adjustment by a listener, of the relative timing of a previously spoken sequence of alternating syllables in order to perceive them as regularly-timed [3, 9, 12]. The synchronisation itself is reported to be achieved by means of the perception of simultaneity [5, 8, 11] or alternation of two regular sequences [2, 3, 9, 10, 12]. The issue of the link between production and perception in such tasks is almost uncontroversial, despite differences in the way perception is treated, either as a relatively autonomous system [3, 10, 13] or as a system intimately entwined with production [11, 12].

The interpretation of the results outlined in the literature usually follows three directions: (1) the global vs local nature of the time events in the syllable enabling successful synchronisation, (2) the articulatory vs (psycho)acoustic nature of the events the subject uses to succeed in performing the task, and (3) the role of vowel onsets (either articulatory or acoustic) as p-centres, given the variability of the results.

The first issue seems to bend to a local explanation of the subjects' performance, even though the models proposed by scholars such as Marcus [14], Pompino-Marschall [10] and Howell [13] require the measurement of the energy from the entire syllable to compute p-centres. Contesting this need, further experiments have shown that local events are sufficient to explain the subjects' behaviour: the use of natural speech in the tasks [8, 9, 12] strongly suggested that the anchor point is located at the vicinity of the vowel onset. The extension of the battery of tests with a few synthetic syllables [3] suggests that a local event corresponding to a local rise within the $F_1$ band is the actual candidate. Furthermore, careful statistical analyses showed that the contribution of offset events in Marcus' or Pompino-Marschall's models is not significant in the regressions [8].

The second issue can only be untangled if, as research advances, we manage to get a deeper understanding of the relation among acoustic, gestural, and articulatory specifications for producing and perceiving speech.

As to the third issue, Janker [8] and de Jong [12] suggest that, despite the variability of the results, that is, despite the significant errors between actual acoustic or articulatory vowel onsets and p-centres, for some syllable types and/or for some subjects, the errors can partly be considered as an articulatory anticipation of the execution of the cognitive plan for the mutual task (finger tapping and perception of a syllable sequence in [8]) or constitute a set of singularities to be investigated further: "What can be said is simply that two of the subjects, and the average taken across the subjects point toward [the tongue dorsum velocity peak] as the best predictor of p-center location" [12]. On the other hand, some authors suggested that the minor discrepancies between vowel onsets and p-centres cannot be discarded and postpone the understanding of the whole picture [9, 11]. The analysis outlined in this article favours both views by considering two levels of time analysis. Furthermore, our results consider a proactive link between speech perception and production [15].

## 2. Speech-metronome synchronisation in Brazilian Portuguese

In the present study the task of producing a sequence of identical syllables in time with a metronome is proposed to a Brazilian Portuguese (henceforth BP) subject. A Matrix metronome, model MR-500, with limit rates of 40 and 208 bpm was used for the tasks. Both the subject's production and the metronome signal were simultaneously captured in a sound-treated room into the two channels of a sound card and digitised at 22.05 kHz. After hearing a few pulses, the subject started producing a sequence composed only by a syllable previously read from a card. One of the four authors kept track of the number of syllables and signalled the subject to stop, after counting a few more than ten repetitions of the syllable. The cards were randomised before each session (one session per metronome rate).

The study aims at deepening the understanding of the variability of the asynchrony between p-centres and acoustic vowel onsets by introducing new, orthogonal experimental conditions: (1) a greater number of syllables which combine seven consonants whose energy extend from a minimum to a maximum: /p f s r ʃ l m/ with three distinct degrees of vowel aperture: /i ɛ a/. This choice has the advantage of using syllables which contrast the total energy of the consonant/vowel pair from a minimum, with [mi], to a maximum, with [pa]; and (2) the variation of the metronome rate including one computed from the spontaneous repetition of a syllable by the subject. CV syllables are chosen not only because they are canonical in BP (see additional comments on this in [16]) but, more crucially, because this choice minimises possible sources of variability introduced by the use of coda consonants, whose role would be secondary in the general pattern examined here. No compensation of anticipation of articulatory activity is done here (as found in [8]): p-centres are computed by projecting the onset of each metronome pulse onto the chain of syllables. The first p-centre is associated with the closest syllable in the chain, the second with the second syllable, and so on.

The 21 different sequences of identical syllables produced by the subject allowed us to test two possible hypotheses for being the task of speech-metronome synchronisation well succeeded: (1) the steeper the rise of total energy at the CV transition, the closer the p-centre to the vowel onset, i. e. p-centres for [pa] would be closer to the CV transition than for [mi], and (2) the higher the energy in the high-frequency band for the consonant, the earlier the p-centre, i. e. p-centres for syllables beginning with [s] or [ʃ] are placed earlier in the syllable. The latter hypothesis was raised after examining the results of the work by Fowler [11], with the same task as the one presented here. In the Figure 2 of her paper (p. 390), one sees that the metronome pulse is placed earlier with respect to the vowel onset for syllables beginning with [s] or containing a consonant with a high frequency component ([t]), in comparison with the other syllables.

The results exhibited here by the patterns of response of the speech-metronome synchronisation in BP points to an explanation in similar grounds. To shed some light on the matter, the difference of energy in bands 1500 Hz-Nyquist frequency (11,025 Hz in our case) and 0-1500 Hz was computed in dB and used as an indication of the consonant spectral tilt. The frequency of 1500 Hz corresponds to the $F_2$ of [a] for this subject and splits the spectral region between high- and low-frequency components.

Four rate values were set for the metronome: the first one was obtained after asking the subject to spontaneously produce a sequence with the syllable [pa], without any external reference. His production paced 108 bpm in average, as measured from consecutive vowel onsets. This corresponds to a cycle of 576 ms. The closeness of this value to the average foot duration in Eriksson's [17] (578 ms), Fant and Kruckenberg's [18] (548 ms), and Fraisse's [19] (600 ms) experiments, is not a coincidence and will be explored later. The following metronome rates used were 30 % lower (80 bpm), and 30 % higher (138 bpm) than the spontaneous one. The last rate corresponds to the metronome upper limit: 208 bpm. The four rates yielded four different behaviours of syllable sequence production, which can be better understood by using the notions of frequency and phase synchronisation.

Frequency synchronisation is achieved when the difference between the metronome rate and the syllable production rate is non-significant. Phase synchronisation is achieved when the p-centre location is stable relative to the vowel onset along the sequence. P-centre and vowel onset are said to be in phase when the phase is statistically null. For all metronome rates and syllables, the subject succeeded in implementing frequency synchronisation, even though the variability indicates two distinct processes being used. As for phase synchronisation the situation is completely distinct. The time difference $d_{ms}$ (in ms) between p-centre – defined as the position of the metronome pulse onset relative to the syllable – and vowel onset, computed by formula (1) was transformed into a phase ($\Phi$) value, relative to the period M of the metronome, by applying formula (2). Note that $\Phi$ can be either positive or negative, depending whether the p-centre respectively precedes or follows the vowel onset.

$$d_{ms} = t_{vowel\ onset} - t_{p\text{-}centre} \quad (1)$$

$$\Phi = d_{ms}.360^{o}/M \quad (2)$$

All repetitions of the syllables produced were retained for each metronome rate. A single-mean t-test carried out for each syllable type was applied in order to test the null hypothesis of phase zero between p-centre and vowel onset (with $\alpha = 5\%$).

In the following, the effects of syllable composition are discussed in the light of the differences in phase synchronisation, whereas the effect of metronome rate changes on the subject's behaviour is analysed by examining how this change affects the phase synchronisation stability.

### 2.1. Effects of syllable composition and of change in metronome rate

Results for the spontaneous metronome rate are shown in Figure 1 (small squares signal mean, box edges indicate standard-errors, and whiskers, standard deviations). Null hypothesis is maintained for the syllables [pɛ], [pi], [fa], [ɣa], [la], [lɛ], and [mi] (/r/ was always realised as [ɣ].) The coincidence between p-centre and vowel onset for syllables such as [mi] and [lɛ] contradicts the first hypothesis above, since the discrepancy of total energy at vowel onset is minimum for these syllables. For syllables beginning with the fricative [ʃ], p-centres precede the vowel onset more than for those with other consonants, which conforms to hypothesis 2 above (the Scheffé homogeneous group test separates three groups: the first formed by [p], the second by [ʃ], and the third by the other five consonants, all vowels confounded. [s] is the consonant with the highest nominal phase value in the third group). The consonant [p] is the one with the least phase difference. The average phase for the vowel [a] is significantly smaller ($F_{2,239} = 16.9$) than that for [ɛ] ($p < 10^{-5}$, Scheffé test), and that for [i] ($p < 10^{-6}$) all consonants considered. Even though the p-centre for [pa] is not statistically synchronous with the vowel onset, the results for [p] and [a], separately, conform to the prediction of the first hypothesis: in synchronisation tasks, the vowel onset is a proper attractor when the syllable has a large amount of total energy rise at that position.

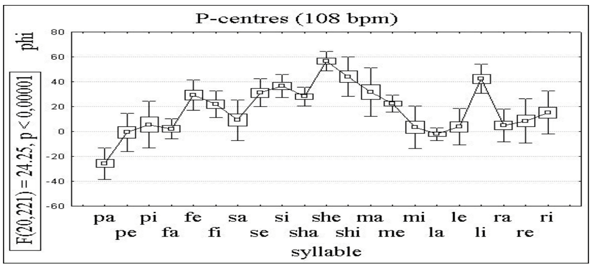


*Figure 1: Phase differences (phi) for p-centre and vowel onset for each syllable with metronome at 108 bpm.*

The correlation of spectral tilt for the consonants in all sequences produced at this rate and phases is -61% ($p < 10^{-2}$), whereas the correlation of the discrepancy of total energy at vowel onset and phase is statistically the same: -63% ($p < 10^{-2}$). The multiple correlation using both acoustic measures of energy discrepancy/spectral tilt and phase is -75% ($p < 10^{-3}$), indicating that both cues explain the success of the subject in predicting the next metronome pulse and synchronising it with a point in the syllable, at least at this rate.

Figure 2 shows the results of the task when the metronome rate decreased to 80 bpm. In this case, the null hypothesis is maintained for the syllables [pɛ], [pi], [fa], [ɣɛ], [ɣi], [sa], and [sɛ]. The homogeneous group Scheffé test separates two groups: that formed by [m] and [p] from that formed by the others. Syllables with [ʃ] are the most asynchronous, which corroborates hypothesis 2. Furthermore, the correlation between the consonant spectral tilt with the phase of the respective syllable, is –57 % ($p < 0.007$). No significant correlation was found between discrepancy of total energy at vowel onset and phase.

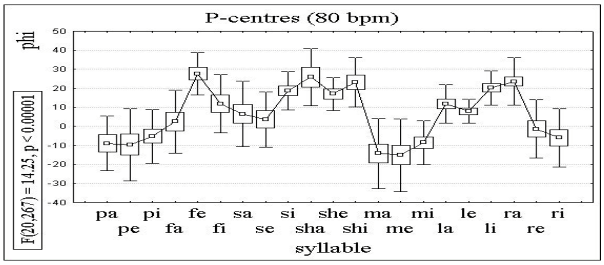


*Figure 2: Phase differences (phi) for p-centre and vowel onset for each syllable with metronome at 80 bpm.*

The results with the metronome at 138 bpm are shown in Figure 3. Only the syllable [ɣi] maintains the null hypothesis. In this case, as well as for the metronome at 208 bpm (for which the [mi] and [lɛ] vowel onsets are synchronous with the metronome pulse), the phase synchronisation is completely lost for almost all syllables. In fact, for these rates, the metronome pulse has no stable point along the syllable: it slides along the syllable envelope as the task progresses. The reason is probably because the subject is unable to evaluate the gap between syllable and metronome pulse in this kind of task (at 138 bpm the maximum gap between vowel onset and pulse is of 217 ms; at 208 bpm, of 144 ms).

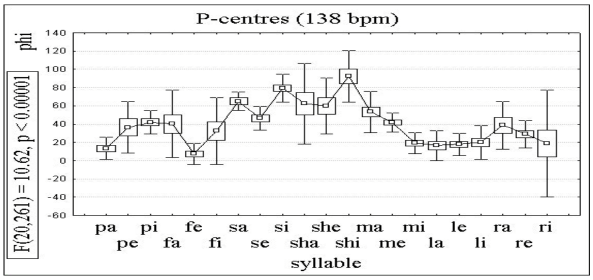


*Figure 3: Phase differences (phi) for p-centre and vowel onset for each syllable with metronome at 138 bpm.*

This change in behaviour when the metronome rate crosses a boundary somewhere between 108 bpm and 138 bpm, which respectively corresponds to inter-beat intervals of 556 ms and 435 ms is called a bifurcation [20] in dynamic systems theory. Another indication of this changing is the fact that the correlation between consecutive syllable cycles are negative for the two slower rates and positive for the two faster ones. This result is the same obtained by Kohno [21], who suggests two different cognitive processes underlying the subjects' behaviour: analytic for slow tempos, and holistic for fast tempos. The average standard-deviation of phases for the four rates also confirms the change in regime: $13^o$ (80 bpm), $13^o$ (108 bpm), $22^o$ (138 bpm), $63^o$ (208 bpm).

In order to do the task, this subject modifies mainly the vowel duration, which significantly shortens when the rate increases (Scheffé: $p < 0.12$, between 80 and 108 bpm; $p<10^{-4}$ between 108 and 208 bpm). The consonant duration is preserved (Scheffé: smallest $p < 0.37$), with the exception of that of [m] and [l] (Scheffé: largest $p < 10^{-3}$), which are smaller at 80 bpm!. At 208 bpm these two consonants have statistically the same duration of the vowel, i. e. the metronome pulse is precisely at the middle of the syllables [mi] and [lɛ].

## 3. General Discussion

The results shown here suggest that the vowel onset acts as an attractor for the task of speech-metronome synchronisation, at least when the subject is asked to repeat a syllable simultaneously with the pulses of a metronome. Consonants with lesser energy such as /p/ and /r/ (and even /s/ in some cases), as well as a vowel with greater energy such as /a/ (with the exception of the task at 80 bpm, /i/ produced the highest phase values) favour the task of phase synchronisation. These results corroborate hypothesis 1. On the other hand, a consonant such as /ʃ/, with a great amount of energy in a high-frequency band (the average central frequency is around 4000 Hz for this subject), perturbs the phase synchronisation task. This result corroborates hypothesis 2, which is confirmed by the significant correlation values above: the higher the amount of energy in the high-frequency band, the earlier the p-centre relative to the vowel onset. Both results can be understood if p-centres are viewed as a surface manifestation of an underlying task of predicting the vowel onsets of a sequence of syllables. The task would be perturbed by noise in some frequency bands: the higher the amplitude of this noise, the more perturbed is the prediction. The task of predicting the next crucial acoustic events conforms to a production-guided, general view of perception [15]. The strategy used by the subject to synchronise two syllables at the fastest rate suggests that he is possibly aware of the difficult in synchronising the other syllables at 208 bpm. This strategy can be viewed as related to habitual patterns in humming BP songs.

The fact that the inter-beat interval at the spontaneous rate is similar to the average foot duration mentioned in the literature, and the shortening of consonants as /m/ and /l/ (which favours the perception of the rhyme duration) at the slowest rate suggest that the subject is aligning the pulses of a phrase stress oscillator with stressed vowels and trying to synchronise them with the metronome pulses. Each stressed syllable he produces delimits the edge of a stress group, in the same way he does when communicating in natural conditions. This corroborates the coupled-oscillator model of speech rhythm production we proposed some years ago [22].

## 4. Acknowledgements

This work was part of the FAPESP project nº 01/00136-2. The first author acknowledges the research grant 352367/2002-0, from the CNPq. The other three thank the post-graduate grants 03/11619-0, 03/09199-2, and 03/05821-0 from FAPESP. We also thank our subject, and Sandra Madureira for suggestions. All authors equally contributed to this paper.

## 5. References

[1] Marcus, S. M., *Perceptual-centres*. PhD Thesis, Cambridge University, 1976.

[2] Morton, J., Marcus, S. M., and Frankish, C., "Perceptual-centers (P-centers)", *Psychological Revue*, 83 (5): 405-408, 1976.

[3] Scott, S. K., *Perceptual centres in speech: an acoustic analysis*. PhD Thesis, University College London, 1993.

[4] Lehiste, I., "Isochrony reconsidered", *J. Phon.*, 5: 253-263, 1977.

[5] Allen, G. D. "The location of rhythmic stress beats in English I & II", *Language & Speech*, 15: 72-100, 179-195, 1972

[6] Classe, A., *The Rhythm of English Prose*. Oxford: Blackwell, 1939.

[7] Rapp, K., "A study of syllable-timing", *STL-QPSR*, Stockholm, 1: 14-19, 1971.

[8] Janker, P. M., "Evidence for the P-Center Syllable-Nucleus-Onset Correspondence Hypothesis", *ZAS Papers in Linguistics*, 7: 94-124, 1996.

[9] Patel, A. D., Löfqvist, A., and Naito, W., "The Acoustics and Kinematics of Regularly Timed Speech: a Database and Method for the Study of the P-Center Problem", *Proceedings of the XIV$^{th}$ ICPhS*, 2: 1075-1078, 1999.

[10] Pompino-Marschall, B. "On the psychoacoustic nature of the P-center phenomenon", *J. Phon.*, 17: 175-192, 1989.

[11] Fowler, C. A., "Converging Sources of Evidence on Spoken and Perceived Rhythms of Speech: Cyclic Production of Vowels in Monosyllabic Stress Feet", *Journal of Experimental Psychology: General*, 112 (3): 386-412, 1983.

[12] De Jong, K., "Acoustic and Articulatory Correlates of P-Center Perception", *UCLA Working Papers in Linguistics*, 81: 66-75, 1992.

[13] Howell, P., "Prediction of p-center location from the distribution of energy in the amplitude envelope I & II", *Perception and Psychophysics*, 43: 90-93, 99, 1988.

[14] Marcus, S. M., "Acoustic determinants of Perceptual-center (p-center) location", *Perception and Psychophysics*, 30 (3): 247-256, 1981.

[15] Berthoz, A., *Le Sens du Mouvement*. Paris: Odile Jacob, 1997.

[16] Barbosa, P. A. and Albano, E. C., "Brazilian Portuguese. Illustrations of the IPA", *Journal of the International Phonetic Association*, 34 (2): 227-232, 2004.

[17] Eriksson, A., *Aspects of Swedish speech rhythm*. PhD Thesis, University of Göteborg, 1991.

[18] Fant, G. and Kruckenberg, A., "Preliminaries to the study of Swedish prose reading and reading style". *STL-QPSR*, 2: 1-80, 1989.

[19] Fraisse, P., "Rhythm and tempo", in *The Psychology of Music*, Deutsch, D., Ed. New York: Academic Press, 1982, pp. 149-180.

[20] Norton, A., "Dynamics: an introduction", in *Mind as motion: explorations in the dynamics of cognition*, Port, R. F. and van Gelder, T., Eds. Cambridge, MA: MIT Press, 1995, pp. 45-68.

[21] Kohno, M., "Two different systems for rhythm processing and their hierarchical relation". *Proceedings of the XIII$^{th}$ ICPhS*, 1: 94-97, 1995.

[22] Barbosa, P. A., "Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production", *Speech Prosody 2002 Conference*, 163-166, 2002.