P-centre Position in Natural Two-Syllable Czech Words

Jan Volín, Eliška Churaňová, Pavel Šturm

Institute of Phonetics, Faculty of Arts, Charles University in Prague

jan.volin@ff.cuni.cz

Abstract

The ability to lock motor activity oscillator with external acoustic events is typical of various forms of human behaviour. Previous research showed that the beginning of an action is not necessarily the beginning of the rhythmic phase and led to the concept of p-centres. We present an experiment with 18 natural two-syllable Czech words spoken in synchrony with metronome beats by 18 subjects. Complexity of the consonantal onset and the type of coda together with distinctive phonological vowel length were carefully controlled to reveal a complex but comprehensible relationship between the word structure and phase locking.

Index Terms: Czech, p-centres, speech rhythm, syllable, synchronization

1. Introduction

Speech rhythm is a fascinating phenomenon attracting humans from their pre-linguistic babbling age through teenage years until their old age. Natural rhythmic flow apparently makes cerebral speech processing easier than uncommon or unpredictable patterns of prominence contrasts. Reaction times in monitoring experiments were by more than 150 milliseconds shorter in sentences with modal rhythm than in the same sentences with altered temporal structure [1], but cf. [2]. Certain types of rhythm in speech may contribute to speaker attractiveness or credibility of the speaker's propositions [3].

There is still considerable uncertainty, however, as to the descriptive principles that should be applied to speech rhythm [4]. One of the key missing pieces of the puzzle is the position of the "beat" – the moment of the perceptual emergence of an acoustic event in the mind of the listener. For syllables this moment was named a *p-centre* [5]. Several experiments of the early 1980s confirmed the intuitions of previous researchers about the importance of the vowel onset. It is not the beginning of the acoustic features pertinent to a syllable but the vicinity of its vocalic nucleus that starts the rhythmic phase.

The attempts to predict the position of the p-centre more accurately exploit mainly timing parameters [6,7], sometimes in combination with the information about energy of the signal [8,9,10]. Moreover, [11] suggested a link between the p-centre position and articulatory movements. Both acoustic descriptors and microbeam data explained substantial amount of variance also in [12], while [13] showed relevance of their kinematic data obtained with an articulometer.

However, Barbosa et al. [14] pointed out that mainly the Germanic languages had been investigated from this perspective. Cross-language comparison is therefore highly desirable. On the one hand, the fundamental principles of rhythmic structuring of speech are universal, since all languages are spoken in syllables, and, for instance, all healthy human ears detect the energy difference between /p/ and /a/ in the syllable /pa/. On the other hand, the rhythm type of an individual language predestines its users to specific strategies in speech perception, speech acquisition, and, naturally, speech production (e.g. a review in [15]). Not only the rhythm type (stress-based, syllable-based, mora-based), but also the phonotactic properties of individual phonemes in the language inventory might influence the exploitation of prominence alternations in speech. Our study is, therefore, aimed at mapping the behaviour of speakers of Czech – a west Slavonic language of central Europe.

Another issue addressed in our approach is the number of subjects. As, e.g., [10] noted, the rhythmic behaviour of human subjects is extremely disparate, and [4] rightly complains that the rhythmic proficiency of individual speakers who serve as subjects or respondents in experiments is often ignored. The pioneering study of [6] was based on four subjects, [7,9,10] on three speakers, [13] on six speakers (only three of whom provided kinematic data), while [14] studied data from one speaker only, and [8] hypothesized without empirical testing. Thus, one of our aims was to increase the number of subjects considerably and provide an estimate of the distribution of the rhythm aptitude in the population.

The third goal of our study concerned the speech material. Many of the previous experiments were carried out with meaningless syllables. This is advantageous in making the experiment neat and highly controlled, and desirable at initial stages of research. However, as [4] stresses, speech is not an object – it is a communicative behaviour. Also, de Jong noticed deviant reactions in his experiments to stimuli that departed from naturally sounding speech tokens [12]. On that account, we abandoned some of the comfort provided by nonsense syllables and used naturally occurring words.

In summary, our search for the position of the p-centre in a speech synchronization task was carried out with the following questions in mind:

- Will the results from a typologically underrepresented language in rhythm research – Czech – be consistent with the previous research? Specifically, do the structural and/or durational features of the stimuli affect the position of the p-centre? If yes, how?
- Given that p-centre position is dependent mainly on syllable onsets and less on codas [e.g., 9], will the rest of a two-syllable word matter if it is used instead of a monosyllable typical of p-centre research? (The rest of a word is actually beyond the stressed syllable coda.)
- Three pairs of our words do not differ structurally but in terms of phonetic identity of the onset consonants. A similar problem was studied by [9]. Do their results hold for a different language and testing conditions?
- Since [14] found dissimilar responses for different tempos, while other researchers seem to implicitly presume proportional changes in p-centre position, will two temporal modes natural to Czech speakers (4 and 6 syll/sec.) produce mutually coherent results?

We intend to answer these questions whilst expanding the number of participants beyond the numbers usual in the field.

10.21437/SpeechProsody.2014-172

2. Method

2.1. Target words

The design of the experiment required a reasonably restricted segmental structure of the targets, yet existing Czech words were favoured over nonsense words because of the reasons outlined above. The disyllabic targets contained the vowel pairs /e aː/ and /eː aː/. Other Czech vowels were not included due to their relative infrequency, occurrence only in foreign words or a salient qualitative difference. Moreover, the demand on using real words necessitated the selection of certain verb forms differing in their suffix, and many vowel combinations were therefore precluded. Each stem was used with the third-person singular zero suffix (open syllables), the first-person singular suffix *-m* (/m/), and the second-person singular suffix *-š* (/ʃ/). The intervocalic consonant was always a voiceless plosive (/t/ or /k/). The onset of the word contained either a single consonant (/l/ and /c/), a CC cluster (/sl/, /st/ and /ʃc/), or a CCC cluster (/spl/). Such a design yielded 18 combinations of word-initial syllable onsets and word-final syllable codas (6 different onsets × 3 different codas). The target words are recapitulated in Table 1.

onset	0-suffix	*m*-suffix	*š*-suffix
C	leːtaː	leːtaːm	leːtaːʃ
CC	sleːtaː	sleːtaːm	sleːtaːʃ
CCC	spleːtaː	spleːtaːm	spleːtaːʃ
CC	steːkaː	steːkaːm	steːkaːʃ
C	cekaː	cekaːm	cekaːʃ
CC	ʃcekaː	ʃcekaːm	ʃcekaːʃ

Table 1. *Summary of the 18 target words. Six lexical items (lines) and three grammatical forms (columns).*

2.2. Speakers and task procedure

18 native speakers of Czech participated in the experiment (14 female and 4 male, aged from 20 to 32). They were mostly students at Charles University in Prague, and reported no speech or hearing impediments. They received financial compensation for their involvement.

The speakers listened to a series of metronome beats (over headphones) and pronounced the target word presented on the computer screen several times. They were instructed to synchronize their repeated articulations with the isochronous sequence of beats so that each beat was aligned with the first syllable of the word spoken in isolation. A clear and "natural" pronunciation was demanded, as natural as it was possible in the experimental conditions (they were explicitly asked not to chant or recite the words). Each item contained 12 pulses followed by soft music and speakers began to articulate on the fifth – the initial four pulses served as a lead-in for steadying listeners' attention. The first and last realizations were omitted from analyses since they were considered prone to effects of uncertainty/accommodation at the beginning of the sequence or anticipation of the stimulus end.

As it is hypothesized that listeners use different processing modes depending on the pace [16], two tempos were tested: in the "normal" tempo the metronome pulse appeared at the rate of 70 bpm (i.e. every 857 ms), while in the "fast" tempo the pulse appeared at the rate of 90 bpm (i.e. every 667 ms). The two tempos were chosen to induce production rates of about 4 to 6 syllables per second (the pulse was associated with the disyllabic word, not with individual syllables).

The items were presented in two blocks separated with a one-minute break; the subjects chatted with the experimenter, could stretch their bodies and refresh themselves with a drink. The first block proceeded in the normal tempo and the second in the fast tempo, which resulted in two occurrences of the target items. In addition, filler items (mono- or trisyllabic words with no restriction on the segmental content) were used to provide variation to the task and prevent subjects from lapsing into monotonous behaviour. Each block was also preceded by several training items in which the subjects accustomed themselves to the tempo and task. Items within each block were pseudo-randomized for individual speakers. The duration of the session, comprising 55 items in total, was approximately 15 minutes.

2.3. Extraction of data

Given the purpose of this study two channels were used simultaneously to record the session, one for the metronome beat and one for the spoken production. The stereo recordings were subsequently processed in the programme *Praat* [17], and TextGrid objects were created for annotation. Individual items were segmented into words and phones using the *Prague Labeller* algorithm [18]. Segment boundaries were marked automatically and then manually checked and corrected where necessary. A script based on the detection of an intensity threshold in the metronome channel was used to determine the position of the recorded beats, which allowed us to identify the location of the metrical pulse for each produced word. The duration of each segment was measured, along with the distance of its boundaries from the metrical beat (*zero* = a segment boundary coincides with the beat; *positive numbers* = boundary located after the metronome beat; *negative numbers* = boundary located before the beat). In the subsequent analyses the term synchronization interval (SI) is used for the distance of the first (i.e., stressed) vowel onset and the beat.

3. Results

3.1. Phonological weight of the word

Since all the research in p-centres we have come across relates the position of the 'mental beat' to the onset of the first vowel, we will do the same for the sake of comparison. It should be also pointed out that Czech word stress is fixed to the first syllable – there are no words stressed on the second syllable.

Assuming that each consonant and short vowel weighs 1 phonological unit and long vowels weigh 2 units we could order the target words from the lightest (5 units) to the heaviest (8 units). However, knowing that individual syllable constituents affect the p-centre position with different power and having tied the weight of the intervocalic consonant and the second vowel, we ordered the overall results by the phonological weight of the first syllable and the type of coda of the second syllable. That seems to arrange the results in an internally congruent manner (see Table 2).

Apparently, the synchronization of the first vowel onset with the metronome beat changes with both the complexity of the first syllable and the absence or nature of the word coda. More complex word onsets pull the synchronization point more ahead of the vowel onset (into themselves, so to say.)

first syllable	word coda	word	mean SI (ms)	SI %
CV	∅	ceka:	-3.7	0.7
CV	Son		5.7	-12.7
CV	Obs		18.1	-34.2
CVV	∅	le:ta:	23.2	-14.0
CVV	Son		25.9	-18.4
CVV	Obs		44.4	-29.4
CCV	∅	ʃceka:	36.4	-57.9
CCV	Son		38.5	-60.0
CCV	Obs		72.8	-120.3
CCVV	∅	sle:ta:	65.3	-41.7
CCVV	Son	ste:ka:	64.7	-43.0
CCVV	Obs		57.5	-39.2
CCCVV	∅	sple:ta:	82.1	-57.1
CCCVV	Son		100.2	-71.7
CCCVV	Obs		94.9	-72.5

Table 2. *Mean synchronization intervals (SI) in milliseconds and percentages of vowel durations for individual phonotactic types (n = 36 apart from CCVV, where n = 2 × 36, i.e.,* sle:ta:/ste:ka: *).*

Long vowels seem to have an effect similar to heavier onsets: they prolong the distance between the first (i.e., stressed) vowel onset and the metronome beat.

Although there can be just one or no consonant word finally, this one coda consonant can still be either an obstruent or a sonorant. This distinction apparently matters – adding a sonorant coda (bilabial /m/ throughout our set) influences the synchronization less than adding an obstruent coda (postalevolar /ʃ/ throughout the set), at least in the case of simpler (phonologically lighter) words. Zero codas leave the vowel onset nearer the beat. Figure 1 displays the mean synchronization intervals (SI) as a function of the first syllable structure, Figure 2 as a function of the word coda.

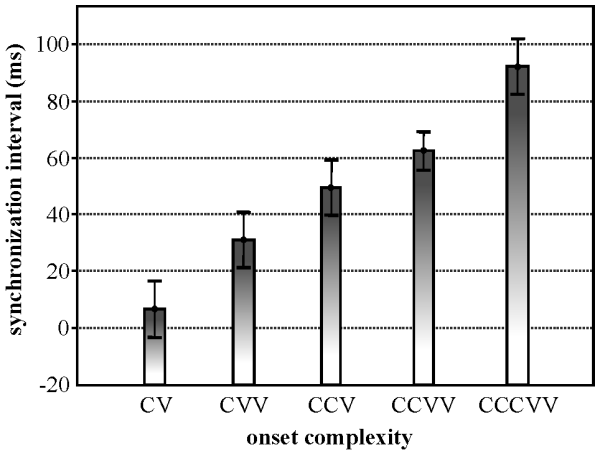

Figure 1. *Mean synchronization intervals for words with various phonological structure of their first syllable. The whiskers delimit the 95% conf. interval.*

Statistical significance of the differences was tested through two-way ANOVA for independent measures with synchronization interval as the dependent variable and ONSET and CODA as independent variables or factors. The main effects of both factors were found significant. For ONSET $F_{(4, 632)} = 43.6$, $p < 0.001$ and for CODA $F_{(2, 632)} = 5.4$, $p < 0.01$. Post-hoc Tukey HSD tests revealed significant differences among all onset complexities with the exception of the CCV which did not differ from CVV and CCVV. Zero codas differed from obstruent codas. No significant interaction was found between the two factors.

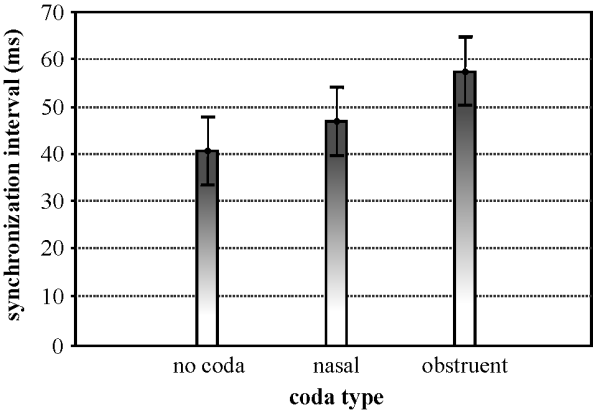

Figure 2. *Mean synchronization intervals for words with various final consonants (zero, nasal, obstruent). The whiskers delimit the 95% conf. interval.*

The situation described above disregards the tempo so it has to be ascertained if both slow and fast modes of synchronization display analogical trends. Inspection of the data showed that in the faster mode the speakers tended to prolong the synchronization interval, while in the slower one the metronome beat tended to be closer to the vowel onset. However, a closer look supported by a two-way ANOVA and post-hoc Tukey tests revealed that the speakers' behaviour was almost identical for simple words. For CV, CVV and CCV the difference between fast and slow tempos produced $p > 0.99$. For CCCVV it was still insignificant with $p \approx 0.52$, and only CCVV produced significant difference ($p < 0.05$). It has to be remembered, though, that the CCVV groups is twice as large as the other groups since it collapses words /ste:ka:/ and /sle:ta:/. (For comparison of these two words see below.) The variance is then lower and the effect is more visible. Clearly, if the effect is sensitive to number of cases in this way, it is not a strong effect. The influence of coda types seemed to be the same for both tempi.

3.2. Pairwise comparisons

Some of the items of our test were designed for direct comparison with others to answer less general questions. The first such comparison concerns the two CCVV words. It was actually carried out prior to the analyses above to make sure that the words /ste:ka:/ and /sle:ta:/ can be collapsed. Although the former has two obstruents in the onset, of which the second is a voiceless plosive, while the latter has got a sonorant sound in the same position, no difference was found in the synchronization interval: $F_{(5, 210)} = 0.44$, $p = 0.817$ (All six forms of these words were entered into the test.) Likewise, no difference was found in their overall duration or of the duration of their consonantal onsets. Internal durational ratios within the onset were different: /k/ was systematically longer than /l/ at the expense of /s/. In summary, although the energy distribution in the onset and the phonetic identities of its constituents differs, the synchronization interval does not.

The pair /ceka:/ × /ʃceka:/ can be observed in Table 2 and it is represented in Figure 1 by the first and the third column. The effect of the additional post-alveolar /ʃ/ is quite substantial. The synchronization moment moves by about 42.5 ms. The durational difference between the /c/ and /ʃc/ onset is about 80 ms, but this value is difficult to measure rigorously since /c/ is a voiceless plosive so the closure phase has to be estimated. Nevertheless, the mean overall duration of the word changes by 35 ms only.

Just as the two previous comparisons addressed the issue of consonantal onset, the following will focus on the vowel length in the first syllable. The relationship of the words /ʃceka:/ × /ste:ka:/ can be estimated from the third and fourth column in Figure 1. Their synchronization differs by about 10 ms which is not a statistically significant result. It is still informative to explore the durational differences in these two items. First of all, the mean duration of /e/ is by 93 ms shorter than that of /e:/, but the durations of the words only differ by 54 ms. It could be inferred that the missing 39 ms are due to the difference between the /st/ and /ʃc/ onsets. That, however is not the case: the difference between the alveolar and palato-alveolar cluster is only 12 ms. The missing time has to be accounted for by durational reorganization of the otherwise identical rest of the word: velar intervocalic consonant, long open vowel and codas. T-tests showed that it is the second vowel that is significantly different ($p < 0.01$). It seems to compensate for what is happening in the first syllable. Interestingly, this complex situation occurs in a pair of words that differ phonologically only in the length of the first vowel. The 10-ms difference in synchronization, which was not found significant actually means a 128-ms distance from the acoustic onset of the word /ʃceka:/ and a 106-ms distance from the acoustic onset of the word /ste:ka:/ suggesting that the first vowel onset is, indeed, a better synchronization point than the very beginning of the word.

Multiple regression analysis

In order to shed more light on the results reported in Table 2, a series of multiple regression analyses were performed with mean synchronization interval (first vowel boundary position) as the dependent variable and several temporal and phonotactic parameters as the predictors. When only TEMPO (slow × fast) and WORD DURATION were taken into account, the explanation power was negligible ($R^2_{adj} = 0.10$), partly because the two factors express similar facts. However, slightly better results were obtained for TEMPO and ONSET DURATION, assuming to reflect the phonotactic complexity of the onset, with $R^2_{adj} = 0.17$. Next, we extended the model with the variable of V1 DURATION, which further increased the amount of variance explained ($F_{(3,643} = 55.3$, $R^2_{adj} = 0.20$, $p < 0.001$). Interestingly, when a phonological category was used instead of a phonetic variable (vowel length instead of vowel duration), the power of the model did not decrease but, on the contrary, slightly increased: $F_{(3,643} = 60.2$, $R^2_{adj} = 0.22$, $p < 0.001$.

Adding information about the coda, whether in the form of CODA DURATION or CODA TYPE, did not make much difference. In both cases the model gained one more percent of explanatory power: $F_{(4,642} = 49.3$, $R^2_{adj} = 0.23$, $p < 0.001$.

If we wanted to use the regression analysis to create an analogy to Marcus' classic formula: $P = 0.65x + 0.25y + k$ (in [6]), where P is the synchronization interval, x is the onset duration and y is the rhyme duration (in the case of our first syllable only vowel), we would get $P = 0.37x + 0.20y - 27.2$. This analogy is not out of proportion, but it has to be remembered that it only explains 16 % of variance and that the methodology leading to it is not comparable with the Marcus' procedure. Be that as it may, the model with the information about coda (z in the following formula) and the whole word duration (w in the following formula) would have:

$$P = 0.61x + 0.50y + 0.27z - 0.29w + 73.6$$

4. Discussion

In a task of synchronizing words with an isochronous auditory sequence, the moment of the acoustic event can be considered the surface manifestation of the p-centre [14]. Our respondents clearly modified their behaviour according to the structure of the word and aligned the words differently for different consonant-vowel (CV) strings and different arrays of features attached to these segments. To answer the first question from the introduction then, we can state that our results based on Czech, typically classified as syllable-timed (but cf. [19]), resonate with previous research on stress-timed English and provide descriptors that can be further tested.

The current analyses are based on durational or structural characteristics only. They explain some of the variance in the data and show the trend for longer and more complex structures to push the p-centre farther ahead of the onset of the first vowel. Word-final segments exert smaller influence on the outcome. They display greater variation and little effect in regression analyses. Little, however, does not mean zero. The triplets of words in our set that differed solely in the presence or absence of the word-final consonant or its manner of production did not behave uniformly. Therefore, the answer to the second question from the introduction is positive as well.

The title of [9] stated that p-centre positions were unaffected by phonetic categorization. We supported that claim using different methodology by showing that our speakers' behaviour was identical for two (triplets of) words that differed in their two-consonant onset. In one of them it consisted of a sibilant + lateral approximant, in the other it was the same sibilant + voiceless plosive. The energy distribution was not equal, but the duration of the onset was. This seems to be the essential property of the item to condition the outcome. Our corroboration concerns the syllable onset only, though. It cannot be generalized for codas where we saw the difference between sonorant /m/ and post-alveolar /ʃ/. The issue will be further investigated in the nearest future as we have already prepared more stimuli for forthcoming testing.

The final question in the introduction reacted to findings in [14], where the respondent produced different patterns of behaviour for different tempi. Our results revealed no interactions between slow and fast mode of testing. That does not provide any evidence against [14] since our methods were incomparable, but rather suggests that for the pace of approximately 4 syllables per second and 6 syllables per second we can expect rather proportionate trends. Our linear regression found word duration as an expression of rate useful.

5. Acknowledgements

This work was supported by the Charles University Grant Agency (GAUK) under Grant 834213, and by the Charles University in Prague programme for science development P10-Linguistics.

6. References

[1] Buxton, H., "Temporal predictability in the perception of English speech", in A. Cutler and D. R. Ladd [Eds], Prosody: Models and Measurements, 111–121, Berlin: Springer-Verlag, 1983.

[2] Dilley, L. C. and Pitt, M. A., "Altering context speech rate can cause words to appear or disappear", *Psychological Science*, 21: 1664–1670, 2010.

[3] Cross, I., "Rhythms of persuasion: The perception of periodicity in oratory", in *Proceedings of Perspectives on Rhythm and Timing*, 27, Glasgow, 2012.

[4] Kohler, K. J., "Rhythm in speech and language: a new research paradigm", *Phonetica*, 66: 29–46, 2009.

[5] Morton, J., Marcus, S., & Frankish, C., "Perceptual centres (P-centres)". *Psychological Review*, 83: 405-408, 1976

[6] Marcus, S., "Acoustic determinants of perceptual center (p-center) location", *Perception & Psychophysics*, 30(3): 247–256, 1981.

[7] Fox, R., and Lehiste, I., "The effect of vowel quality variations on stress-beat location", *Journal of Phonetics*, 15: 1-13, 1987.

[8] Howell, P., "An acoustic determinant of perceived and produced anisochrony", in *Proceedings of the 10th ICPhS*, 429–433, Dordrecht, 1984.

[9] Cooper, A.M., Whalen, D.H., Fowler, C., "P-Centers are unaffected by phonetic categorization", *Perception & Psychophysics*, 39, 187-196, 1986.

[10] Pompino-Marschall, B., "On the psychoacoustic nature of the P-centre phenomenon", *Journal of Phonetics*, 17: 175–192, 1989.

[11] Fowler, C. A., Whalen, D. H. and Cooper, A. M., "Perceived timing is produced timing: A reply to Howell", *Perception & Psychophysics*, 43: 94–98, 1988.

[12] de Jong, K., "Acoustic and articulatory correlates of p-centre perception", *UCLA Working Papers in Phonetics*, 81: 66-75, 1992

[13] Patel, A., Löfquist, A. and Naito, W., "The acoustics and kinematics of regularly timed speech: a database and method for the study of the P-Centre problem", in *Proceedings of 14th ICPhS*, 1: 405–408, San Francisco, 1999.

[14] Barbosa, P. A., Arantes, P., Meireles, A. R., Vieira, J. M., "Abstractness in speech-metronome synchronisation: P-centres as cyclic attractors", in *Proc. of 9th European Conf. on Speech Communication and Technology (Interspeech 2005)*, 1441–1444, Lisbon, 2005.

[15] Cutler, A. and Otake T., "Rhythmic categories in spoken-word recognition", *Journal of Memory and Language*, 46: 296–322, 2002.

[16] Kohno, M., "Two different systems for rhythm processing and their hierarchical relation", in *Proceedings of the 13th ICPhS*, 1: 94-97, Stockholm, 1995.

[17] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer (Version 5.2.37)", 2011, accessed on September 3, 2011 from http://www.praat.org/.

[18] Pollák, P., Volín, J. and Skarnitzl, R.: "HMM-Based Phonetic Segmentation in Praat Environment", in *Proceedings of the XIIth International Conference Speech and Computer – SPECOM 2007*, 537–541, Moscow, 2007.

[19] Dankovičová, J. and Dellwo, V.: "Czech speech rhythm and the rhythm class hypothesis", in *Proceedings of the 16th ICPhS*: 1241-1244, Saarbrücken, 2007.