Journal of Phonetics 55 (2016) 38–52

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

P-centres in natural disyllabic Czech words in a large-scale speech-metronome synchronization experiment

Pavel Šturm*, Jan Volín

Institute of Phonetics, Charles University in Prague, n. J. Palacha 2, Prague 1, 116 38, Czech Republic

A R T I C L E I N F O

Article history:
Received 12 March 2015
Received in revised form
17 November 2015
Accepted 20 November 2015
Available online 17 December 2015

Keywords:
Czech
P-centre
Speech rhythm
Syllable
Synchronization

A B S T R A C T

The difficulty of pinpointing a specific event within words that would correspond to the p-centre is well known. The current experiment, investigating the position of p-centres in Czech, aims to replicate the findings from English and several other languages, and substantially increase the range of phonotactic types and the number of participants. In a speech-metronome synchronization task, 24 subjects pronounced a set of 37 natural disyllabic Czech words of differing complexity at two metronome rates. The beginning of the first vowel (V1) and the moment of the fastest increase in energy within the first syllable were the most consistent synchronization points, but the p-centre occurred earlier than at the V1 initial boundary. Synchronization intervals were significantly influenced by the complexity of the syllabic onset: the p-centre was positioned earlier (further from the V1) as more consonants were included in the onset. The effects of vowel length and final coda were also present, but weaker. In addition, various aspects of human musicality were found to correlate with the ability of speakers to synchronize their articulations with an isochronous auditory sequence.

1. Introduction

1.1. Rhythm in speech

Despite the fact that rhythm is a ubiquitous and familiar phenomenon in both the animate and inanimate world, the role of rhythm in speech has only recently been duly acknowledged. Similarly to other types of human behaviour from breathing and walking to music, dance and collective chanting, speech exploits rhythmicity in order to be more effective, albeit less explicitly. It is common that speakers adjust their temporal organization of speech so that it more closely resembles that of their conversation partner (Cummins, 2009), which may facilitate the communication process and indicate willingness to cooperate. Moreover, it has repeatedly been demonstrated that a natural rhythmic flow of speech is easier to process by the brain, leading to shorter reaction times, than uncommon or unpredictable patterns of prominence contrasts (e.g. Buxton, 1983; Quené & Port, 2005). A theoretical explanation is offered by the neural resonance model of Grossberg (2003) which states that the formation of a percept of any object (syllable, word, etc.) requires highly synchronized activations of neural assemblies in the brain that pertain to the input neural representation on the one hand (analysis of the incoming signal), and the expectational representations on the other (derived from experience and analysis of the context). The argument is discussed more specifically with regard to rhythm in Ghitza and Greenberg (2009). It is no longer surprising, then, that good public speakers tend to be – consciously or subconsciously – more rhythmical than less skilful orators, as such a performance demands less mental effort on the part of the listeners (Kohler, 2009). In a similar vein, specific kinds of rhythm may also contribute to speakers' attractiveness or to the credibility of their propositions (Knight & Cross, 2012).

To put speech rhythm into a larger perspective, numerous experiments on timing control from across disciplines have shown that if we are dealing with recursive actions, rhythmic movements are executed more easily than arrhythmic ones (Cummins, 2009; Kelso, 1995; Port, 2003; Repp, 2005; Turvey, 1990). Regular is not only easier, but it also allows for coordination of actions. For instance, Kelso (1984) asked subjects to oscillate their index fingers to the left and right at various speeds as determined by metronome pulses. The task proved to be most effortless when both fingers moved towards or away from each other, irrespective of tempo,

* Corresponding author. Tel.: +420 221 619 250.
E-mail address: pavel.sturm@ff.cuni.cz (P. Šturm).

0095-4470/$ - see front matter
http://dx.doi.org/10.1016/j.wocn.2015.11.003

whereas other phase relationships were more unstable, with the exception of the synchronous movement of both fingers either to the left or right (but only in the slow tempo). These two modes of wagging seem to work as attractor states towards which the subject's performance converges. Although to a large extent independent, the two hands prefer to work in coordination. More importantly, comparable results were obtained when the subject performed a cyclic action while watching someone else, so it is not merely a result of the physical link between the two hands (Port, Tajima, & Cummins, 1996). Further, applying such principles to speech, Cummins and Port (1998) discovered that in a speech cycling task analogous discrete coordinative patterns emerge. The internal timing of a phrase that was repeated in time with a metronome was not completely autonomous; rather, in that experiment certain stressed syllables inclined to 1/3, 1/2 and 2/3 of the phrase repetition cycle, revealing a harmonic relationship.

Researchers have approached rhythm in speech from various angles. The so-called rhythm metrics, capable of capturing some interesting (structural) differences between languages, individuals or speaking situations, have been used extensively in recent years, although it has been rightly recognized that they provide little information about rhythm as such (e.g. Kohler, 2009). Crucially, these metrics fail to capture the fact that rhythm operates beyond pure duration and, moreover, that rhythm is a perceptual phenomenon (Cummins, 2009; Kohler, 2009; Lee & Todd, 2004; Lehiste, 1977; Volín, 2010). It was acknowledged early on that an acoustically regular arrangement of stressed syllables does not result in perceived isochrony (Lehiste, 1977). If we subscribe to the perceptual view of rhythm, this finding must be neither surprising nor discomforting. On a more general level, the lack of monotonous chains of rhythmic units is to be expected in conversational speech because our speech behaviour is ultimately guided by the requirements of the content (e.g., the structuring of the idea to be expressed; Volín, 2010). More specifically, the metrical "beat" the listener is assumed to perceive in individual words is simply not associated with the acoustic onset of the word or syllable, but its position is determined by a number of acoustic and psychoacoustic factors. The moment of the perceptual emergence of the syllable in the mind of the listener was termed the "P-centre" (Morton, Marcus, & Frankish, 1976).

1.2. Investigation of p-centres

Early inquiries into the nature of p-centres investigated the position of rhythmic stress beats in English. Although the subjects in Rapp's (1971) study repeated nonsense words to a regularly occurring pulse while Allen's (1972) experiment included finger tapping to the presumed beat of a specified syllable in a sentence, in both paradigms the point of synchronization was located near the acoustic onset of the vowel. More importantly, it shifted backward in time (i.e., further ahead from the vowel) in direct proportion to the prevocalic consonantal duration. This finding was corroborated by the classic study of Morton et al. (1976) who found that sequences of digits that were aligned perfectly with respect to their acoustic onsets were not perceived as rhythmical. When instructed to adjust the intervals between the digits for better rhythmicity, the listeners introduced considerable variation into the inter-onset intervals. Taken together, these experiments revealed that departures from acoustic isochrony are systematic and vary in tandem with the duration of the prevocalic material, but nevertheless failed to pinpoint any acoustic marker corresponding to the p-centre, such as the acoustic onset of the word or of the stressed vowel.

Several factors have been identified to play a major role in the determination of p-centres. The duration of the (acoustically salient) prevocalic consonantal portion of a syllable is positively correlated with a backward shift of the p-centre (Allen, 1972; Fowler, 1979; Fowler & Tassinary, 1981; Marcus, 1981; Rapp, 1971). Cooper, Whalen, and Fowler (1986) confirmed that the duration of the onset is the key factor, not its phonetic identity; however, the results of Harsin (1997) based on a wider range of segments suggest that phonetic categories do have an effect on p-centre placement via their acoustic properties, as the size of the effect of onset duration differed for different initial consonants. The relevance of the stressed vowel in terms of its duration was reported by Marcus (1981) and by Fox and Lehiste (1987a), who concurred that longer vowels are associated with later p-centres. The latter investigated tense and lax vowels in English monosyllables and found that it was specifically duration and not vowel quality that contributed to the effect. Marcus (1981) and Fox and Lehiste (1985) reported a significant effect of the final consonant, as did Cooper, Whalen, and Fowler (1988) for both the nucleus and the coda. Finally, Fox and Lehiste (1987b) examined how the location of the stress beat is affected by adding an unstressed prefix or suffix to a monosyllabic word. In both production and perception, an unstressed suffix pulled the p-centre later into the stressed vowel, whereas a prefix exercised a much larger shift in the opposite direction. This demonstrates above all the necessity of investigating more complex stimuli in addition to the simple sequences of stressed monosyllables that have hitherto been the main focus of p-centre research.

A perceptual aspect of the p-centre phenomenon is advocated by Pompino-Marschall (1989), who developed a psychoacoustical, rather than an acoustical, model of p-centres that is based on loudness functions within critical auditory bands. He agrees with Marcus (1981) in assigning greater impact to onsets rather than to the rest of the syllable, but differs from him in establishing interactions between onsets and vowels, and vowels and codas. Moreover, he found that the amplitude envelope of the syllable plays an important role (similarly to Howell, 1984, 1988a,b). In a similar vein the study of Harsin (1997) points to the role of low-frequency energy modulations in the acoustic signal, in particular the velocity peaks at the C–V transitions. P-centres thus seem to be affected not only by the duration of its constituent segments but also by their energy distributions and, by inference, their phonetic identity, which contrasts with the already mentioned data of Cooper et al. (1986) or Fox and Lehiste (1987a). In any case, it is evident that properties of the whole stimulus participate in determining the location of p-centres.

Articulatory correlates of p-centres were investigated in the work of Carol Fowler and her colleagues (Fowler, 1979, 1983; Fowler & Tassinary, 1981; Tuller & Fowler, 1980). Although restricted to only a few subjects and to simple monosyllabic items (e.g., sequences of /sæd mæd stæd stræd/ etc.), her results support the previous findings in a variety of production and perceptual tasks. It is clear that in speech production speakers synchronize some event within a syllable, and presumably within its consonantal onset.