Journal of Phonetics 49 (2015) 55–66

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

Syllable synchronization and the P-center in Cantonese

Ivan Chow *, Michel Belyk, Vance Tran, Steven Brown

Department of Psychology, Neuroscience & Behavior, McMaster University, 1280 Main Street West Hamilton, Ont., Canada L8S 4L8

ARTICLE INFO

Article history:
Received 19 November 2013
Received in revised form
1 October 2014
Accepted 16 October 2014

Keywords:
Cantonese
Lexical tone language
P-center
Syllable synchronization
Speech rhythm

ABSTRACT

In speech rhythm analysis, it is important to localize the perceptual center (P-center) of syllables in order to establish a basis for measuring syllabic duration. P-center research has focused primarily on Germanic languages, for which syllables tend to begin with multiple consonants. In Cantonese, the syllable-initial position contains no more than one consonant, making it less prone to durational variation. Studies of various language types using syllables beginning with a singleton indicate that the P-center is generally localized close to vowel onset within the initial consonant–vowel transition. In the present study, Cantonese speakers were found to localize the P-center to the initial consonant–vowel transition when repeating a sequence of identical syllables in a synchronized manner against an audible metronome. However, contrary to previous findings, the metronome beat was aligned close to the onset of the syllable-initial consonant. With increasing repetitions, syllable-initial consonant onset became more closely aligned with the beat while vowel onset became further removed. Similar behavioral patterns are found in sensorimotor synchronization studies (e.g., for finger tapping), suggesting that Cantonese speakers use syllable-initial consonant onset as an articulatory reference point in speech synchronization. Further investigation is needed to elucidate the prosodic and cognitive basis of this behavioral discrepancy.

1. Introduction

To understand speech rhythm, it is essential to examine how timing units align with the syllable (Liberman, Shankweiler, Fischer, & Carter, 1974; Port, 2003). Central to this understanding is the identification of the "perceptual center," or P-center, which is the temporal reference point at which a syllable is perceived to occur (Cooper, Whalen, & Fowler, 1986; Fraisse, 1974; Hoequist, 1983; Marcus, 1981; Morton, Marcus, & Frankish, 1976). It follows that the localization of the P-center within the syllable would provide a basis for measuring syllabic duration (Pompino-Marschall, 1989, 1991).

A paradigm in which a sensorimotor activity, such as finger tapping, is repeated while synchronizing to a visual or auditory metronome signal is commonly used in psychology to study human sensorimotor synchronization (SMS) behavior (see Repp, 2005 and Repp & Su, 2013 for comprehensive reviews). Based on such a paradigm, the present study was designed to observe how Cantonese speakers repeat series of identical monosyllables while synchronizing to an audible metronome click. Without explicitly instructing participants to align the metronome click to a specific acoustic point of reference within the syllable (e.g., syllable-initial consonant onset or vowel onset[1]), we attempted to localize the P-center by measuring the timing differences between the temporal reference point (i.e., the time of the metronome click) and the acoustic landmarks used by speakers to mark the syllables' point of occurrence. A similar experimental design was undertaken in P-center studies on Brazilian Portuguese by Barbosa, Arantes, Meireles, and Vieira (2005). Admittedly, this approach is not common in prosody research. As such, it may be subject to some forms of adaptation specific to syllable synchronization that are potentially not used in normal speech. However, if normal speech had "rhythm," it would be useful to establish a relationship between an isochronous time reference and the point at which a syllable is

Abbreviations: P-center, perceptual center; NMA, negative mean asynchrony.

* Corresponding author. Tel.: +1 905 525 9140x21443.
E-mail address: chowiv@mcmaster.ca (I. Chow).

[1] In the context of this paper, "onset" is taken to mean the instantaneous point at which a consonant or vowel begins. Onset is not used in the traditional phonological sense to signify the part of a syllable that occurs before the rime.

0095-4470/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2014.10.006

perceived to occur (i.e., the P-center). The measurement of syllabic duration, then, provides a foundation for performing quantitative analyses on speech rhythm (cf. Allen, 1972; Fraisse, 1982; Klatt, 1976; Lehiste, 1977).

The exact location of the P-center and its corresponding acoustic landmarks continue to be the subject of debate. However, for syllables beginning with a consonant, it is generally agreed that the P-center is localized close to vowel onset, within the transition between the onset of the syllable-initial consonant and the vowel (Barbosa et al., 2005; Hoequist, 1983; Marcus, 1981; Patel, Löfqvist, & Naito, 1999). Research on the P-center is generally conducted from either the perspective of speech perception or speech production. As pointed out by Barbosa et al. (2005), speech perception is linked with speech production in the case of speech synchronization tasks. Thus, the two are either treated as somewhat autonomous systems (Howell, 1988; Pompino-Marschall, 1989; Scott, 1993) or as intimately intertwined (Fowler, 1979, 1983; de Jong, 1992, 1994; Morton et al., 1976). In fact, in order to perform speech synchronization, which involves both processes, it would make sense for both speech perception and production to make reference to the same speech event (Fowler, 1979; de Jong, 1994). This in turn supports the assumption that research from either perspective would yield similar results and that the P-center location would essentially be analogous in both speech perception and production.

In studies that took the speech perception approach (Cooper et al., 1986; de Jong, 1994; Marcus, 1981; Patel et al., 1999; Pompino-Marschall, 1989; Scott, 1998), participants were asked to listen to a pre-recorded isochronous sequence of syllables with different initial consonants and rimes. The methodology of these studies differs mainly in the use of simultaneous or interpolating metronome signals as a temporal reference during the experiment. In the test sequence, syllables were aligned in such a way that the inter-onset intervals (IOI) were constant. Participants were asked to make judgments as to whether the syllable sequence was isochronous or not, either between subsequent syllables or between syllables and the metronome. Based on this judgment, they then adjusted the temporal positions of the syllables until the sequence was perceived as isochronous. The position of the P-center was determined by locating the point in the syllables that coincided with the newly adjusted isochronous interval.

Cooper et al. (1986) carried out a series of speech perception experiments by asking participants to adjust the temporal positions of an isochronous series of syllables with initial fricatives (e.g., /s/) until the series was perceived as isochronous. The duration of these fricatives was varied to create different experimental conditions in order to observe the effects of syllable-initial consonant duration on the P-center location. Variations in the duration of the syllable-initial fricative also triggered changes in the categorical perception of the pre-vocalic consonant in such a way that the fricative at the long end of the spectrum would be perceived as a fricative /s/, while the same fricative would be perceived as an affricate /ts/ at the short end. Results showed that the P-center was not affected by the perceptual (phonetic) categorization of the pre-vocalic consonant. Instead, it was largely influenced by the duration of that consonant and, to a lesser extent, the duration of the syllabic rime, which agrees with findings of Marcus (1981) and Rapp-Holmgren (1971). Using an alternating series of synthetic syllables and metronome signals, Pompino-Marschall (1989) showed similar findings in terms of the syllabic rime. However, in contrast to the work of Cooper et al. (1986) and Hoequist (1983, next paragraph), Pompino-Marschall found a non-linear relationship between the syllable-initial consonant duration and the location of the P-center.

In research that assumed the speech production perspective, participants were asked to produce a sequence of syllables while synchronizing with a visual or auditory metronome signal (e.g., Barbosa et al., 2005; Cummins & Port, 1998, 2009; de Jong, 2001; Hoequist, 1983; Tajima & Port, 2003). De Jong (2001) used the syllable synchronization paradigm to study the temporal relations between open (consonant–vowel) and closed (vowel–consonant) syllables, whereas Cummins and Port's (1998) research concentrated on observing whether speakers were able to synchronize their speech against external signals that were in phase, anti-phase, or out of phase. The latter did not provide concrete results as to exactly which part of the syllable was used by speakers to align with the metronome signal. Likewise, based on the same "speech cycling" research paradigm (Cummins & Port, 1998, 2009), Tajima and Port (2003) conducted a comparative study to observe how stress patterns of different languages affect the subdivision of timing units between prominent syllables within prosodic feet for English and within words for Japanese. Since the P-center was not the focus of their study, Tajima and Port (2003) assumed that the P-center would be localized to the vowel onset and used it as a reference point for examining the temporal alignment of syllables within a speech cycle. They did not investigate where the P-center was localized, nor did they make a comparison of the P-center location between English and Japanese speakers. Both Hoequist (1983) and Patel et al. (1999) used a paradigm in which participants were asked to repeat alternating monosyllables in an isochronous manner. Hoequist created two test conditions and conducted a comparative experiment between speakers of English, Spanish, and Japanese in which participants were asked to perform the repetition task in phase with a metronome pulse in one condition and without a metronome pulse in the other. However, data collected from the repetition task with a metronome pulse were not included in the statistical analyses. In Patel et al.'s (1999) study, participants were simply asked to produce a sequence of isochronous syllables without the help of a metronome. The P-center was then determined by identifying the location within the syllables that coincided with the point at which isochrony was established. In Barbosa et al. (2005), a speaker of Brazilian Portuguese was asked to repeat 21 series of identical consonant–vowel syllables while synchronizing to metronome signals of four different frequencies (80, 108, 138, and 208 beats per minute (bpm)).

Despite differences in experimental paradigms and the native language of participants, Barbosa et al. (2005), Hoequist (1983), and Patel et al. (1999) generally conclude that the P-center occurs close to vowel onset within the transition between the syllable-initial consonant and the vowel. However, they were unsuccessful in identifying any acoustic or articulatory landmark that precisely aligns with the P-center. On the other hand, Hoequist (1983) found a slight but statistically insignificant language difference in which the alignment of vowel onset to the P-center was less accurate for Spanish speakers compared to that for English and Japanese speakers. Similar to Pompino-Marschall (1989), Barbosa et al. (2005) found the location of the P-center to be largely determined by