11 The P-Center Effect and the Domain of Beat Perception in Speech

TAMARA RATHCKE

11.1 The Perceptual Center in Speech

The notion of the perceptual center (or the P-center) dates back to the beginnings of speech rhythm research that focused on temporal isochrony (Morton et al., 1976), though the concept has not lost its appeal to the present day (Lin and De Jong, 2023; Zoefel et al., 2023). The P-center is defined as the subjectively perceived moment of occurrence, highlighting that acoustic and perceptual onsets of rhythmic events do not necessarily co-occur (Morton et al., 1976). Instead, the P-center seems to lag behind the acoustic onset of the corresponding rhythmic event, such as a (monosyllabic) word. The discovery was made with a recording of English digits from one to nine that were evenly concatenated to create an isochronous rhythmic sequence. The evenly spaced concatenation, however, sounded irregular to the experimenters (and other listeners). The sequence could only be made regular once the digits were concatenated with reference to the perceived, rather than acoustic, isochrony (Morton et al., 1976). It was observed that the perceptual onsets of the concatenated digits deviated systematically, but somewhat inconsistently, from their acoustic counterparts. They did not coincide with local peaks in signal amplitude or fundamental frequency (Morton et al., 1976).

Figure 11.1 illustrates this idea with the recording of a speaker producing the words *bad*, *mad*, *sad*, *had*, *ad*, *pad* at a steady pace as cued by a 2.5 Hz metronome (with an interval of 400 ms between beat onsets). If we measure the resulting intervals between successive word onsets, the produced sequence of words is irregular. And yet it sounds isochronous, just as the speaker intended to produce it. Under the P-center view, perceptual isochrony of this example derives from an even spacing of the P-centers of rhythmic speech events. The concept has also been applied to music (London et al., 2019; Vos and Rasch, 1981), and some discussions of the P-center propose that it constitutes the level of the beat in speech, thus linking the rhythmic structure of speech and music (Allen, 1972; Cumming et al., 2015; Harsin, 1997; Harsin and Green, 1994; Hoequist, 1983; Scott, 1998). The beat in music is defined as an underlying grid of equal time intervals that provides temporal structure to musical notes (Savage et al., 2015). It is uncontroversially regular in contrast to speech timing that shows no evidence for isochrony and

https://doi.org/10.1017/9781009295888.014 Published online by Cambridge University Press

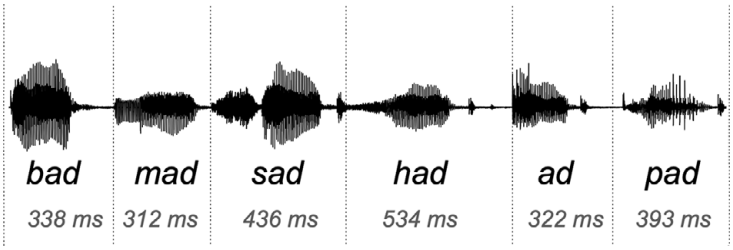


Figure 11.1 Illustration of the P-center effect.
The word sequence consisting of six English words (*bad, mad, sad, had, ad*, and *pad*) was produced in sync with a metronome at 2.5 Hz (or 400 ms between beat onsets) and sounds highly regular. However, the resulting inter-onset intervals between successive word onsets vary between minimally 312 ms (*mad-sad*) and maximally 534 ms (*had-ad*), demonstrating a discrepancy between (irregular) acoustics and (regular) perception typical of the P-center effect.

very limited evidence for regularity (Arvaniti, 2009; Dauer, 1983; Rathcke and Smith, 2015), at least on the surface of measurable acoustics (Turk and Shattuck-Hufnagel, 2013). The P-center effect indicates that equal time intervals may exist in speech after all, and that they are perceptual rather than acoustic in nature.

11.2 Methods of Examining the P-Center

Following its discovery, the P-center was extensively researched using a great variety of methods with the ultimate goal of developing an algorithm that would automatically identify P-center location in speech. One method gave rise to the example shown in Figure 11.1. In this task, participants are asked to produce a series of words or syllables in time with a (real or imagined) metronome (e.g., Chow et al., 2015; Fowler, 1979; Fox and Lehiste, 1987a; Marcus, 1981; Šturm and Volín, 2016; Tuller and Fowler, 1980). The subsequent analyses focus either on determining the alignment point of the metronome beat and the speech signal or on identifying the magnitude of discrepancies between timings of words with identical versus varied phonological structure. Another commonly used task is based on the perceptual adjustment for isochrony (e.g., Cooper et al., 1988; Harsin, 1997; Marcus, 1981; Pompino-Marschall, 1989; Scott, 1998). In a version of this task, listeners are given one base word repeating with a fixed inter-onset interval and asked to adjust the timing of a following, phonologically different word such that it matches the regular inter-onset intervals of the preceding base word repetitions. The subsequent analyses of listeners' adjustments examine temporal deviations between the inter-onset intervals established by the base words and the perceptually matched words that deviate from the base word in their phonological structure. Another variant of the task asks listeners to align words to metronome beats (Pompino-Marschall, 1989). Finally, participants have also been asked to tap along with a designated syllable of a looped sequence of words (Allen, 1972). The timing of the tap can then be analyzed, determining the location of the syllable's P-center.

https://doi.org/10.1017/9781009295888.014 Published online by Cambridge University Press

Having discovered the P-center effect with a series of isolated monosyllabic words, Morton et al. (1976: 405) were cautious to add that the properties (and the existence) of the P-center may well be "subject to phonological, semantic, or syntactic influences" that play a role in natural speech. These influences have not yet been empirically addressed (see Chapter 22). While it has also been suggested that the P-center effect may explain the apparent lack of isochrony in speech acoustics (e.g., Lehiste, 1977; Morton et al., 1976), no studies have examined the P-center in connected speech. Across all methods mentioned above, speech materials consist of isolated, real or nonce, words presented with an intervening pause. The materials can be varied with regards to the identity of onset consonants and the phonological complexity on their clusters (e.g., *seed, bead, lead, blead*), the vowel quality in the syllable nucleus (e.g., *bad, bed, bid*), the presence or absence of coda consonants, and the complexity of codas (e.g., *see, seek, seeks*), but they have been generally restricted to mono- or bisyllabic words. Evidence from related work on beat perception in natural connected speech indicates that the subjectively perceived onset of the beat in spoken sentences indeed deviates from the acoustic onset of phonological syllables (Lin and Rathcke, 2020; Rathcke et al., 2021), providing preliminary support for the P-center effect in speech of higher complexity than one-word utterances (though without demonstrating an isochronous distribution of P-centers in natural speech).

11.3 On the Location of the P-Center

There is no generally accepted model of the exact P-center location (Villing et al., 2011). The evidence on which factors affect it and in what ways is mixed. It is mostly assumed to lie somewhere close to the vowel onset (Barbosa et al., 2005; Franich, 2018; Hoequist, 1983; Marcus, 1981) and to be mostly affected by syllable onsets rather than codas (Howell, 1988; Marcus, 1981; Pompino-Marschall, 1989; Scott and Howell, 1992; Šturm and Volín, 2016), though these generalizations may be limited to Germanic and possibly Romance languages that have predominantly been studied to date (Šturm and Volín, 2016). In Cantonese, however, monosyllabic words produced in time with a metronome do not show the tendency for the beat to lag behind the acoustic syllable onset (Chow et al., 2015). The speech-to-metronome synchronization in this tonal language is tightly timed to syllable-initial consonants rather than vowels (Chow et al., 2015), casting doubts on the existence of the P-center in Cantonese and, more generally, on the effect being a cross-linguistic universal as previously suggested (Hoequist, 1983).

Overall, the effect has been documented in nontonal languages including English (e.g., Cooper et al., 1988; Fowler, 1979; Fox, 1987; Harsin, 1997; Tuller and Fowler, 1980), German (Pompino-Marschall, 1989; Pompino-Marschall et al., 1989), Brazilian Portuguese (Barbosa et al., 2005), Spanish (Hoequist, 1983), Czech (Šturm and Volín, 2016), and Japanese (Fox, 1987; Hoequist, 1983). Extending typological diversity, the effect has recently been ascertained in Medumba, a tonal language of the Bantu family (Franich, 2018), and Mandarin Chinese (Lin and De Jong, 2023). This finding suggests that the lack of the P-center effect in Cantonese cannot be due to lexical tone. Chow et al. (2015) explain their

https://doi.org/10.1017/9781009295888.014 Published online by Cambridge University Press

result with reference to syllable structure in Cantonese whose onsets are either empty or occupied by one or maximally two (an obstruent plus a glide) consonants. The authors suggest that the phonotactic restriction may have the acoustic consequence that the prevocalic part in Cantonese syllables is relatively short and minimally variable, leading to vowel onsets being less reliable acoustic landmarks than onsets of syllable-initial consonants. However, the syllable structure of Cantonese is quite comparable to that of Japanese (Kubozono, 1989; Otake et al., 1993), yet Japanese speakers (Hoequist, 1983) and listeners (Fox, 1987) display the P-center effect in production and perception comparable to the one found with English speakers and listeners. Moreover, Mandarin Chinese has an even more restricted syllable phonotactics than Cantonese (e.g., Zhao and Berent, 2016), though a recent production study indicates that the P-center of Mandarin Chinese is located close to the acoustic vowel onset, just as in nontonal languages. Language-specific syllable phonotactics is thus less likely to be the main reason for the cross-linguistic differences in the P-center effect.

While Cantonese boasts a complex tone system with several dynamic and level tones, Medumba has a two-way contrast (Franich, 2018) and Mandarin has a four-way contrast (Lin and De Jong, 2023). It is unclear whether this difference in tonal inventory can account for the discrepancy in P-center findings across the three tonal languages. It is also unclear whether pitch plays any role in influencing the location of the P-center. While Chow et al. (2015) did not observe any differences between P-centers of Cantonese words carrying different tones and Lin and De Jong (2023) only examined syllables with tone-1, Franich (2018) measured differently timed P-centers in words carrying a low versus high tone, with high tones leading to earlier P-centers. In their seminal study, Morton et al. (1976) did not find an effect of pitch on P-center location in English. This finding was confirmed in a recent study with more complex English materials (Lin and Rathcke, 2020; Rathcke and Lin, 2023), though there remains a possibility that pitch may shape beat perception in some (not necessarily tonal) languages.

In languages that clearly demonstrate the P-center, its location seems to be affected by the properties of the whole syllable or word, though the effects of onset, nucleus, and coda are neither similar in magnitude nor additive, and the evidence documenting the (phonological versus phonetic) nature of P-center shifts is mixed. Early work by Marcus (1981) experimented with natural and manipulated versions of monosyllabic words for English digits and found that their P-center was located later in the syllable if the duration of the onset was shorter, or if the vowel or coda duration was longer. Fox and Lehiste (1987a) asked if such durational influences on P-center shifts were phonological rather than phonetic in nature, given that many phonological contrasts (e.g., tense versus lax vowels in English) go hand in hand with timing alternations (long versus short). They conducted a study into the effect of vowel quality as opposed to vowel duration on P-center location, examining English monosyllables with lax versus tense vowel nuclei. The results indicated little role of vowel phonology in shifting the location of the P-center within a syllable, confirming that the nature of the P-center effect was purely phonetic rather than phonological. An opposite conclusion was reached by Šturm and Volín (2016) who demonstrated that P-center location in bisyllabic words of Czech was strongly affected by the phonological vowel length rather than their phonetic duration.

Cooper et al. (1986) studied the phonetic influence of syllable onsets and nuclei by varying the duration of fricative noise in a fricative-vowel syllable, the duration of acoustic silence in a

https://doi.org/10.1017/9781009295888.014 Published online by Cambridge University Press

fricative-stop-vowel syllable, or the duration of the vowel itself. The perception of the P-center in the resulting stimuli was mostly affected by the duration of the syllable-initial consonant(s) and, to a lesser extent, by the duration of the vowel, showing temporal shifts similar to those documented by Marcus (1981). Following up on this work, Cooper et al. (1988) examined the role of syllable rime in more detail, testing the hypothesis put forward by Marcus (1981) that the rime behaves as a unit such that durational variability in the vowel versus the coda does not exert an independent influence on the location of the P-center. Two experiments systematically manipulated the duration of the vowel in a vowel-consonant syllable with either covarying or constant duration of the rime. The results did not provide evidence in support of the hypothesis by Marcus (1981). Instead, they suggested that both constituents of the rime (i.e., vowels and codas) had comparable effects on affecting P-center location.

A series of experiments with more varied materials conducted by Pompino-Marschall (1989), however, showed that the phonetic effects of segment duration on P-center location were not as linear and additive as suggested by earlier research. Rather, the duration of the syllable onset, vowel, and coda interacted in complex ways, jointly determining the direction and the magnitude of shifts in P-center location. Adding to this complexity, Harsin (1997) provided further evidence that the durational effect of syllable onsets on P-center location did not equally apply across a wide range of consonants but was moderated by the phonological category of the onset. Specifically, syllables with sonorants versus obstruents of the same duration differed in their P-centers and did not display a unified effect of consonant lengthening on a later location of the P-center that had been generally shown in earlier work with more limited materials (Allen, 1972; Cooper et al., 1988; Fowler, 1979; Marcus, 1981). Given that sonorants and obstruents show remarkable differences in their energy distributions and amplitude envelopes, subsequent work focused primarily on the attempts to model P-center location as a function of spectro-temporal properties of a syllable, even though experimental evidence to this end had been rather mixed (Harsin, 1997; Marcus, 1981; Morton et al., 1976; Tuller and Fowler, 1980).

Testing an acoustic account of the P-center in their original work, Morton et al. (1976) excluded local peaks in signal amplitude or in fundamental frequency as suitable signal-driven anchors of the center location. Subsequent studies further elaborated that the P-center did not coincide with any acoustic landmarks in speech (e.g., Cooper et al., 1986; Marcus, 1981). However, Howell (1988) and Scott and Howell (1992) revisited the acoustic account of the P-center and proposed a model based on the amplitude envelope and a syllabic "center of gravity," suggesting that perceptual judgments are linked to the distribution of the energy in a syllable (see Chapter 3). In this model, the center of gravity refers to the moment when the energy peak of a syllable is reached, typically at the consonant-vowel transition. The slope of the energy rise toward the center of gravity is assumed to encode onset consonants and is crucial to the calculations of P-center location. If the energy contour rises quickly right from the syllable onset (as for some syllable-initial fricatives), the P-center occurs earlier; if it shows a more gradual increase, the P-center is located later (the concept came to be widely known as syllable rise time, e.g., Goswami et al., 2011; Leong et al., 2011). This model stands in contrast to the most recent acoustic representation of P-center location that somewhat downplays the energy of some consonants – notably fricatives – in order to derive the P-center (Šturm and Volín, 2016). According to Šturm and Volín (2016), the P-center is best represented as the

https://doi.org/10.1017/9781009295888.014 Published online by Cambridge University Press

moment of the fastest energy change (maxD) occurring at the consonant-vowel transitions within a syllable, though for the algorithm to perform well, the high energy of some consonants such as fricatives ought to be significantly downplayed and smoothed (Šturm and Volín, 2016: 42). Previous attempts to localize the P-center at the midpoint of the amplitude rise time did not have that feature (Cummins and Port, 1998). The two algorithms are available for researchers with an interest in the study of the P-center, either from the first author's website (Cummins and Port, 1998) or upon individual request (Šturm and Volín, 2016).

Despite some differences, both algorithms of P-center location operate within the domain of a syllable and sample acoustic properties of the local amplitude envelope delimited by the syllable boundaries. There is, however, some evidence that the P-center can also be affected by a preceding or following syllable. For example, Fox and Lehiste (1987b) showed that the P-center shifts to a later location if an additional syllable is suffixed to form a bisyllabic word. In contrast, it shifts (even more substantially) to an earlier location if an additional syllable is prefixed. While Šturm and Volín (2016) focused specifically on bisyllabic words, they did not compare them to monosyllables, so it is unclear whether or not their *maxD* algorithm should account for polysyllabic complexity and in what ways. In recent work, we applied the algorithm to more varied and naturally complex materials in English and examined the potential of the *maxD*-derived landmark to predict the location of finger taps produced during a task requiring participants to synchronize with the beat of repeated sentences (Rathcke et al., 2021). The results showed that the maxD landmark was statistically as good a predictor of finger-tap locations as vowel onsets. Even though both the task and the stimuli of this sensorimotor synchronization experiment by far exceeded the complexity of more traditional P-center paradigms, the findings confirmed that the P-center effect existed in the perception of rhythmic beat structure in natural English speech. That is, there is a discrepancy between the acoustic onset of a syllable and the perceived onset of the syllable beat.

11.4 Explanations of the P-Center Effect

The original interest in the effect was grounded in the idea that rhythm meant isochrony and motivated by the search for some temporal constancy in language. The discovery of the effect gave rise to the hypothesis that isochrony in language might be perceptual and not acoustic (Lehiste, 1977; Morton et al., 1976). Even though the idea that speech rhythm can be defined purely on the basis of duration and timing has received much criticism (Arvaniti, 2009; Kohler, 2009) and is not unanimously shared (White and Malisz, 2020), the P-center effect maintains its relevance to speech rhythm research as it signifies a notable discrepancy between speech acoustics and perception. Such discrepancy is not unique to the P-center but generally applies to a range of speech perception phenomena that show nonlinear relationships with physical input properties (e.g., Dilley and Pitt, 2010; Goldstone and Hendrickson, 2010; Warren, 1968).

As noted by Morton et al. (1976: 408), the concept of the P-center has "no explanatory power" of its own as it simply describes one temporal aspect of speech perception. Not surprisingly, approaches to explaining the origin of the P-center effect somewhat mirror approaches to understanding speech perception in general. A prominent account in this regard

https://doi.org/10.1017/9781009295888.014 Published online by Cambridge University Press

is the proposal put forward by Fowler and colleagues (Fowler, 1979, 1986, 1994; Tuller and Fowler, 1980). It follows the motor theory of speech perception (Liberman and Mattingly, 1985), assuming that articulatory gestures constitute perceptual units in connected speech and that P-centers track the kinematic signal of speech production or, more specifically, the temporal regularity of vowel gestures (see Chapter 2). Accordingly, the P-center effect originates in the fact that "listeners extract information from the acoustic signal that specifies articulatory timing" (Fowler et al., 1988: 94). While articulatory recordings do not straightforwardly support the motor account of the P-center (De Jong, 1992, 1994; Pompino-Marschall et al., 1989), the key suggestion that beat perception in speech may be locked to vowels is further found in other discussions of the P-center effect (Barbosa et al., 2005; Rathcke et al., 2021). For example, Barbosa et al. (2005) hypothesize that P-centers can best be understood as a surface manifestation of the perceptual task to predict upcoming vowel onsets in a sequence of syllables. Similarly, Rathcke et al. (2021) discuss that vowels have a special importance for speech perception as they shape the sonority contour of the acoustic speech signal. The resulting fluctuations in signal sonority may support speech segmentation and assist first-language acquisition (Räsänen et al., 2018). Naturally evolved drummed languages such as Amazonian Bora also make use of such sonority fluctuations and map rhythmic units onto intervocalic intervals (Seifart et al., 2018), further corroborating the special role of vowels for the perception of speech rhythm. In languages spoken around the world, the nucleus (most frequently occupied by a vowel) represents the only obligatory constituent of a syllable and is often reflected in a local sonority maximum (Blevins, 1995).

Rathcke et al. (2021) also note that during rhythmic synchronization with the beat of natural sentences, it was particularly the very first vowel of a sentence that showed anticipation – in other words, a vowel occurring after an acoustic silence. All subsequent vowels – in other words, vowels embedded in a meaningful sentence – were synchronized with more precisely and in a less anticipatory way. Notably, previous studies of P-center location experimented with isolated words concatenated using silent pauses. It is therefore not implausible to hypothesize that the P-center may reflect a temporal prediction of a vowel onset that is expected to occur after a silent pause. This explanation of the P-center effect is in line with current evidence of negative mean asynchrony obtained in rhythmic synchronization tasks with a variety of auditory stimuli (Aschersleben, 2002; Repp and Su, 2013). Accordingly, measurable anticipation of regular auditory prompts occurs specifically when those prompts are interspersed with acoustic silences but is attenuated, or even completely removed, in complex, continuous rhythmic contexts such as music (see Chapter 6). This hypothesis of P-center origin requires experimental testing in future research.

Alternative accounts of the P-center assume that the effect is rather acoustic (Howell, 1988; Scott and Howell, 1992; Šturm and Volín, 2016; Vos and Rasch, 1981) or psychoacoustic (Harsin, 1993, 1997; Pompino-Marschall, 1989) in nature. These accounts highlight that the P-center is neither unique to speech nor completely independent of the spectro-temporal features of the stimuli tested. Psychoacoustic models define P-centers with reference to critical-band audio frequency regions that matter for the human auditory system (see Chapter 3). Accordingly, the P-center effect arises due to a salient acoustic energy change within the critical audio frequency bands of an entire syllable. The P-center itself can then be best modeled as a

https://doi.org/10.1017/9781009295888.014 Published online by Cambridge University Press

tracker of acoustic changes at relevant frequencies to which a perceptual weighting function is applied. The weighting function integrates the knowledge of critical bands as well as perceptual thresholds that need to be reached for the P-center effect to arise at a given point in time.

Purely acoustic models tend to abstract away from the complexities of critical frequency bands and nonlocal influences on P-center location. These models see the origin of the effect in the perceptual system sampling amplitude envelopes at onsets of auditory input units (e.g., syllables, tones, or metronome clicks) and responding particularly sensitively to salient points of the maximal rate of change. Most recent installments of this account further suggest that the perceptual system may not be simply attracted to the local moments of the fastest energy change but is sensitive to the overall rate of change in the amplitude envelope (i.e., slope and rise time). Accordingly, only some acoustic signals lend themselves readily to the perception of a clear P-center at a certain point in time (Villing et al., 2011), while others may be perceived like a "broad slur" (Benadon, 2014). A notion of a "beat bin" has been put forward to account for the fact that the clarity of the P-center tends to vary across different types of auditory input (Danielsen et al., 2019), though this idea has not yet been comprehensively investigated, particularly in speech.

11.5 Unresolved Issues and Future Directions

Apart from the controversies surrounding the exact P-center location, difficulties with the development of suitable P-center algorithms, and limited availability of cross-linguistic evidence, the current understanding of the effect faces one key issue – individual variability. As Pompino-Marschall (1989) notes, listener performance in P-center tasks can differ rather substantially. Early work even reported on difficulties in determining P-center locations with inexperienced listeners (Cooper et al., 1988; Fox and Lehiste, 1987b; Morton et al., 1976). Several studies reviewed above are based on data from no more than three or four participants, which, in the presence of large individual variability, suggests that the difficulty in establishing P-center location may be even greater than currently appreciated, though it may also be more meaningful than currently assumed. Studying the P-center effect with a more representative sample of 23 Cantonese participants, Chow et al. (2015: 63) also noted that the participants "behaved quite differently from one another." The answer to the question of which individual listener traits and characteristics moderate the perceptual variability may potentially help to better explain the origins and the nature of the effect. Some preliminary findings indicate that musical training (or, more generally, musical aptitude) plays a role in rhythmic tasks such as P-center paradigms (Rathcke et al., 2021; Šturm and Volín, 2016). Participants with higher levels of musical training show reduced variability of P-center responses (Šturm and Volín, 2016) and higher accuracy in rhythmic synchronization with vowel onsets (Rathcke et al., 2021). Having studied the effect with highly skilled musicians, Villing et al. (2011: 1626) found consistent results across a range of tasks and argued that P-centers demonstrated "a reliable and universal percept" – a conclusion that probably owed a lot to the homogeneity of the participating group of listeners.

However, the role of musical aptitude and training is possibly limited to the perception of participants whose native languages do not have lexical tone, as Chow et al. (2015) did not

https://doi.org/10.1017/9781009295888.014 Published online by Cambridge University Press

observe any systematic differences in P-center location among musically trained and untrained Cantonese participants. Little is known if, and how, native language(s) of listeners shape(s) their beat perception in speech and other complex auditory signals, which is a fruitful avenue to explore in future studies of the P-center effect.

Box 11.1 Chapter Overview

Summary

The P-center refers to the perceptual moment of occurrence of a speech unit and has been hypothesized to represent the beat in spoken language. It can be found among many other controversial concepts surrounding the idea of rhythm in speech and language. Over decades of study, the exact location and the nature of the P-center have remained largely unresolved, though the concept itself has retained its potential to inform future research.

Implications

The P-center effect has direct implications for the construction of speech stimuli, specifically for those experiments that work with concatenated monosyllables interspersed by silent pauses. If the P-center effect is not considered, an acoustically constant inter-onset interval connecting a string of phonologically variable monosyllables may be perceived as irregular, while a (slightly) jittered concatenation results in a good approximation of perceptual regularity.

Gains

The key contribution of the P-center to the current understanding of rhythm in speech and language is rather profound as it establishes that the perception of temporal structure in speech, just as the perception of spectral and other features, deviates from the acoustic signal in complex ways and is not universal but language-specific. As such, the P-center signifies that a purely acoustic study of speech rhythm is likely to be futile.

REFERENCES

Allen, G. D. (1972). The location of rhythmic stress beats in English: An experimental study I. *Language and Speech*, 15(1), 72–100. https://doi.org/10.1177/002383097201500110

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66(1–2), 46–63. https://doi.org/10.1159/000208930

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition*, 48(1), 66–79. https://doi.org/10.1006/brcg.2001.1304

Barbosa, P. A., Arantes, P., Meireles, A. R., and Vieira, J. M. (2005). Abstractness in speech–metronome synchronisation: P-centres as cyclic attractors. Interspeech 2005, pp. 1441–1444. https://doi.org/10.21437/Interspeech.2005-512

https://doi.org/10.1017/9781009295888.014 Published online by Cambridge University Press

Benadon, F. (2014). Metrical perception of trisyllabic speech rhythms. *Psychological Research*, 78(1), 113–123. https://doi.org/10.1007/s00426-013-0480-1

Blevins, J. (1995). The syllable in phonological theory. In J. A. Goldsmith (ed.), *The handbook of phonological theory* (pp. 206–244). Blackwell.

Chow, I., Belyk, M., Tran, V., and Brown, S. (2015). Syllable synchronization and the P-center in Cantonese. *Journal of Phonetics*, 49, 55–66. https://doi.org/10.1016/j.wocn.2014.10.006

Cooper, A. M., Whalen, D. H., and Fowler, C. A. (1986). P-centers are unaffected by phonetic categorization. *Perception & Psychophysics*, 39(3), 187–196. https://doi.org/10.3758/BF03212490

Cooper, A. M., Whalen, D. H., and Fowler, C. A. (1988). The syllable's rhyme affects its P-center as a unit. *Journal of Phonetics*, 16(2), 231–241. https://doi.org/10.1016/S0095-4470(19)30489-9

Cumming, R., Wilson, A., Leong, V., Colling, L. J., and Goswami, U. (2015). Awareness of rhythm patterns in speech and music in children with specific language impairments. *Frontiers in Human Neuroscience*, 9. www.frontiersin.org/articles/10.3389/fnhum.2015.00672

Cummins, F., and Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, 26(2), 145–171. https://doi.org/10.1006/jpho.1998.0070

Danielsen, A., Nymoen, K., Anderson, E., et al. (2019). Where is the beat in that note? Effects of attack, duration, and frequency on the perceived timing of musical and quasi-musical sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 45(3), 402–418. https://doi.org/10.1037/xhp0000611

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11(1), 51–62. https://doi.org/10.1016/S0095-4470(19)30776-4

De Jong, K. J. (1992). Acoustic and articulatory predictors of p-center perception. *Journal of the Acoustical Society of America*, 91(4), 2339–2339. https://doi.org/10.1121/1.403496

De Jong, K. J. (1994). The correlation of P-center adjustments with articulatory and acoustic events. *Perception & Psychophysics*, 56(4), 447–460. https://doi.org/10.3758/BF03206736

Dilley, L. C., and Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, 21(11), 1664–1670. https://doi.org/10.1177/0956797610384743

Fowler, C. A. (1979). "Perceptual centers" in speech production and perception. *Perception & Psychophysics*, 25(5), 375–388. https://doi.org/10.3758/BF03199846

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics*, 14(1), 3–28. https://doi.org/10.1016/S0095-4470(19)30607-2

Fowler, C. A. (1994). English speech rhythm. *Language and Speech*, 37(1), 67–76. https://doi.org/10.1177/002383099403700105

Fowler, C. A., Whalen, D. H., and Cooper, F. S. (1988). Perceived timing is produced timing: A reply to Howell. *Perception & Psychophysics*, 44(4), 386–392.

Fox, R. A. (1987). Perceived P-center location in English and Japanese (working paper). Department of Linguistics, Ohio State University. https://kb.osu.edu/handle/1811/81333

Fox, R. A., and Lehiste, I. (1987a). Effect of unstressed affixes on stress-beat location in speech production and perception. *Perceptual and Motor Skills*, 65(1), 35–44. https://doi.org/10.2466/pms.1987.65.1.35

Fox, R. A., and Lehiste, I. (1987b). The effect of vowel quality variations on stress-beat location. *Journal of Phonetics*, 15(1), 1–13. https://doi.org/10.1016/S0095-4470(19)30532-7

Franich, K. (2018). Tonal and morphophonological effects on the location of perceptual centers (P-centers): Evidence from a Bantu language. *Journal of Phonetics*, 67, 21–33. https://doi.org/10.1016/j.wocn.2017.11.001

Goldstone, R. L., and Hendrickson, A. T. (2010). Categorical perception. *WIREs Cognitive Science*, 1(1), 69–78. https://doi.org/10.1002/wcs.26

Goswami, U., Fosker, T., Huss, M., Mead, N., and Szűcs, D. (2011). Rise time and formant transition duration in the discrimination of speech sounds: The Ba-Wa distinction in developmental dyslexia:

https://doi.org/10.1017/9781009295888.014 Published online by Cambridge University Press

Auditory discrimination in dyslexia. *Developmental Science*, 14(1), 34–43. https://doi.org/10.1111/j.1467-7687.2010.00955.x

Harsin, C. A. (1993). Acoustics of perceptual centers in speech. *Journal of the Acoustical Society of America*, 94(3), 1864–1864. https://doi.org/10.1121/1.407646

Harsin, C. A. (1997). Perceptual-center modeling is affected by including acoustic rate-of-change modulations. *Perception & Psychophysics*, 59(2), 243–251. https://doi.org/10.3758/BF03211892

Harsin, C. A., and Green, K. P. (1994). Perceptual centers as an index of speech rhythm. *Journal of the Acoustical Society of America*, 96(5), 3350–3350. https://doi.org/10.1121/1.410633

Hoequist, C. E. (1983). The perceptual center and rhythm categories. *Language and Speech*, 26(4), 367–376. https://doi.org/10.1177/002383098302600404

Howell, P. (1988). Prediction of P-center location from the distribution of energy in the amplitude envelope: I. *Perception & Psychophysics*, 43, 90–93. https://doi.org/10.3758/BF03208978

Kohler, K. J. (2009). Rhythm in speech and language. *Phonetica*, 66(1–2), 29–45. https://doi.org/10.1159/000208929

Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech*, 32(3), 249–278. https://doi.org/10.1177/002383098903200304

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5(3), 253–263. https://doi.org/10.1016/S0095-4470(19)31139-8

Leong, V., Hämäläinen, J., Soltész, F., and Goswami, U. (2011). Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language*, 64(1), 59–73. https://doi.org/10.1016/j.jml.2010.09.003

Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36. https://doi.org/10.1016/0010-0277(85)90021-6

Lin, C.-Y., and Rathcke, T. (2020). How to hit that beat: Testing acoustic anchors of rhythmic movement with speech. Proceedings of Speech Prosody 2020, pp. 1–5. https://doi.org/10.21437/SpeechProsody.2020-1

Lin, Y.-J., and De Jong, K. (2023). The perceptual center in Mandarin Chinese syllables. *Journal of Phonetics*, 99, 101245. https://doi.org/10.1016/j.wocn.2023.101245

London, J., Nymoen, K., Langerød, M. T., et al. A. (2019). A comparison of methods for investigating the perceptual center of musical sounds. *Attention, Perception, & Psychophysics*, 81(6), 2088–2101. https://doi.org/10.3758/s13414-019-01747-y

Marcus, S. M. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics*, 30(3), 247–256. https://doi.org/10.3758/BF03214280

Morton, J., Marcus, S., and Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, 83, 405–408. https://doi.org/10.1037/0033-295X.83.5.405

Otake, T., Hatano, G., Cutler, A., and Mehler, J. (1993). Mora or Syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32(2), 258–278. https://doi.org/10.1006/jmla.1993.1014

Pompino-Marschall, B. (1989). On the psychoacoustic nature of the P-center phenomenon. *Journal of Phonetics*, 17(3), 175–192. https://doi.org/10.1016/S0095-4470(19)30428-0

Pompino-Marschall, B., Kühnert, B., and Tillmann, H. G. (1989). P centers, C centers, or what else? *Journal of the Acoustical Society of America*, 85(S1), S28–S28. https://doi.org/10.1121/1.2026894

Räsänen, O., Doyle, G., and Frank, M. C. (2018). Pre-linguistic segmentation of speech into syllable-like units. *Cognition*, 171, 130–150. https://doi.org/10.1016/j.cognition.2017.11.003

Rathcke, T., and Lin, C.-Y. (2023). An acoustic study of rhythmic synchronization with natural English speech. *Journal of Phonetics*, 100, 101263. https://doi.org/10.1016/j.wocn.2023.101263

Rathcke, T., and Smith, R. H. (2015). Speech timing and linguistic rhythm: On the acoustic bases of rhythm typologies. *Journal of the Acoustical Society of America*, 137(5), 2834–2845. https://doi.org/10.1121/1.4919322

https://doi.org/10.1017/9781009295888.014 Published online by Cambridge University Press

Rathcke, T., Lin, C.-Y., Falk, S., and Dalla Bella, S. (2021). Tapping into linguistic rhythm. *Laboratory Phonology*, 12(1), 1. https://doi.org/10.5334/labphon.248

Repp, B. H., and Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, 20(3), 403–452. https://doi.org/10.3758/s13423-012-0371-2

Savage, P. E., Brown, S., Sakai, E., and Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, 112(29), 8987–8992. https://doi.org/10.1073/pnas.1414495112

Scott, S. K. (1998). The point of P-centres. *Psychological Research*, 61(1), 4–11. https://doi.org/10.1007/PL00008162

Scott, S. K., and Howell, P. (1992). Perceptual centers in speech: An acoustic analysis. *Journal of the Acoustical Society of America*, 92(4), 2443–2443. https://doi.org/10.1121/1.404580

Seifart, F., Meyer, J., Grawunder, S., and Dentel, L. (2018). Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. *Royal Society Open Science*, 5(4), 170354. https://doi.org/10.1098/rsos.170354

Šturm, P., and Volín, J. (2016). P-centres in natural disyllabic Czech words in a large-scale speech–metronome synchronization experiment. *Journal of Phonetics*, 55, 38–52. https://doi.org/10.1016/j.wocn.2015.11.003

Tuller, B., and Fowler, C. A. (1980). Some articulatory correlates of perceptual isochrony. *Perception & Psychophysics*, 27(4), 277–283. https://doi.org/10.3758/BF03206115

Turk, A., and Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology*, 4(1), 93–118. https://doi.org/10.1515/lp-2013-0005

Villing, R. C., Repp, B. H., Ward, T. E., and Timoney, J. M. (2011). Measuring perceptual centers using the phase correction response. *Attention, Perception, & Psychophysics*, 73(5), 1614–1629. https://doi.org/10.3758/s13414-011-0110-1

Vos, J., and Rasch, R. (1981). The perceptual onset of musical tones. *Perception & Psychophysics*, 29(4), 323–335. https://doi.org/10.3758/BF03207341

Warren, R. M. (1968). Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin*, 70(4), 261–270. https://doi.org/10.1037/h0026275

White, L., and Malisz, Z. (2020). Speech rhythm and timing. In C. Gussenhoven and A. Chen (eds.), *Oxford Handbook of Language Prosody* (pp. 167–182). Oxford University Press.

Zhao, X., and Berent, I. (2016). Universal restrictions on syllable structure: Evidence from Mandarin Chinese. *Journal of Psycholinguistic Research*, 45(4), 795–811. https://doi.org/10.1007/s10936-015-9375-1

Zoefel, B., Gilbert, R. A., and Davis, M. H. (2023). Intelligibility improves perception of timing changes in speech. *PLoS ONE*, 18(1), e0279024. https://doi.org/10.1371/journal.pone.0279024

https://doi.org/10.1017/9781009295888.014 Published online by Cambridge University Press