8 Rhythm and Speech

FRED CUMMINS

8.1 Laying the foundations: The many senses of rhythm

In a jazz band, the rhythm section of drum and bass provides a regular framework around which the soloists dance and weave, at times conspiring with the beat, at times, pulling away from it in playful or passionate exchange. Rhythm is both the regular grid that provides structure, and the use of that grid to generate, satisfy, or frustrate expectation in time. Whether we use the term "rhythm" in a narrow or extended sense, its use creates a tension between two poles. With the first, we immediately evoke a sense of periodicity, of regularity and recurrence, that serves to heighten expectations and to tie events to particular points in time or space. With the other, we develop the potential for creative expression that lifts off from the grid, and that expresses itself by not being perfectly regular, by omitting the predictable, and switching in the unexpected. Events and accents are interpreted against a background of regularity evoked by an underlying period. Rhythm is more than mere clock time, the invariant sequence of evenly spaced intervals, and yet of such regularity is rhythm born.

It is in music that the concept of rhythm, as distinct from mere periodicity, is at home. In a musical representation of the well known "shave-and-a-haircut – two bits!" motif (Figure 8.1), we can distinguish between the rhythmic pattern of the specific phrase, and the interpretation of this pattern as based on a sequence of evenly spaced (isochronous) beats, which in turn admit of grouping into relatively stronger and weaker positions. Figure 8.1 (right) shows a metrical grouping built over a sequence of eight beats. The numbers indicate the relative rhythmic strength at each point in the sequence. These strengths serve to tune expectation about future events, to focus attention at specific points in time, and to provide a sense of compositional structure to a note sequence (Huron 2006; Large and Riess Jones 1999).

The Handbook of Speech Production, First Edition. Edited by Melissa A. Redford.

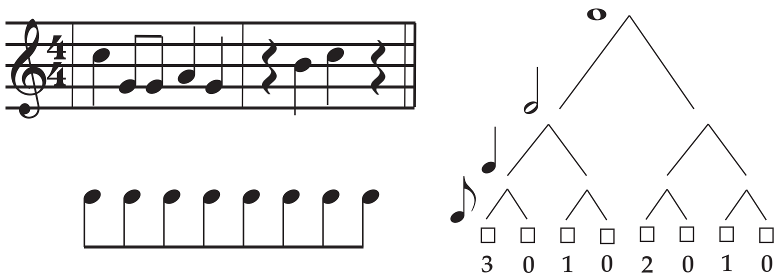

Figure 8.1 Musical representation of the well-known "shave-and-a-haircut – two bits!" motif and beat structure (left) with its metrical interpretation (right).

Speech is continuous with music, but most speech is not musical. When we discuss rhythm in speech, we are not applying a musical concept in an entirely novel domain. The voice is a valued instrument from opera to hip hop. Moving from music, there are intermediate forms of vocal activity in group recitations, prayers, chants, and protest calls that share many of the characteristics of musical performance. These are all collective activities, requiring coordinated timing across individuals. Most forms of choral speaking or recitation employ familiar texts and they are repeated many times, resulting in a highly stylized form of prosody. The characteristic cadences of the American Pledge of Allegiance, for example, will be familiar to many.

The speech of an individual may be rhythmically exaggerated too. Auctioneers often use an idiosyncratic form of patter intended to maintain a constant stream of speech, even when propositional content is limited (Kuiper 1992). Livestock auctioneer competitions provide amusing examples aplenty. Parents reading nursery rhymes to infants will exaggerate rhythm too, using the expectation generated by a strong meter to modulate the attention of the child (Bergeson and Trehub 2002). Trouvain and Barry (2000) provide a thorough analysis of the timing characteristics of the excited speech of race horse commentators. In all cases, rhythmic modulation of the speech goes hand in hand with the modulation of other prosodic characteristics, including speech melody, intensity, and voice quality.

The term "rhythm" has been liberally applied with respect to speech. This chapter will focus primarily on senses of the term that remain close to the musical sense of events critically located in continuous time. It will not treat of the phonology of meter, which understands rhythm as consisting in the atemporal but sequential ordering of strong and weak elements arranged into hierarchical structures (Liberman and Prince 1977). Poetics too must be passed over (Abercrombie 1965), and with it, the formerly canonical art of rhetoric, now sadly in decline.[1]

8.2 The isochrony debate

The manifest similarities between speech and music have led to many attempts to find common underlying principles. One of the first is Joshua Steele's *An Essay towards Establishing the Melody and Measure of Speech to be Expressed and Perpetuated* *by Peculiar Symbols* (1775). Although Steele was concerned mainly with links between the pitch of speech and its links to musical melody, he employed a form of musical notation that also ascribed durations to individual syllables. These gave expression to an underlying assumption, shared by many since, that the impression of near-regular rhythm in speech might be derivable from musical models, and specifically, that some event sequence, such as the onsets of stressed syllables, might be found to be evenly spaced in time. Daniel Jones made this explicit, when he said: "there is a general tendency to make the stress-points of stressed syllables follow each other at equal intervals of time, but … this general tendency is constantly interfered with by the variations in the number and nature of the sounds between successive stress-points" (1918/1956).

An early instrumental study by Classé (1939) served to lend support to this eminently plausible intuition about English speech rhythm. He had subjects read texts (taken from Daniel Jones) into a device called a kymograph that produced a trace of the intensity variation of the speech wave. He measured the intervals between successive syllable onsets (see below), and arrived at findings that were both illuminating and unsurprising. Even spacing between successive stressed syllables emerged as a tendency in the recordings – a tendency greatly encouraged when the lexical material was written with an ear to rhythm, when successive intervals contained phonetically matched segments and syllables, and when they had relatively similar grammatical construction. Any such tendency was disrupted by inter-sentence breaks. This was much as Daniel Jones had surmised, and is to be expected on the basis of English phonology, in which we find both full and greatly reduced syllables, in approximate alternation.

8.2.1 Measurement issues

In his measurements, Classé demonstrated a robust phonetician's instinct that the onset of stressed syllables are important events in the perception of rhythmic progression in spoken utterances. The onsets he measured were indexed, not by the first occurrence of acoustic energy, but by the mid-point in the rise of the amplitude envelope displayed in the kymograph trace.

Determining precisely when something happens is possible only for idealized punctate events of no duration. Real world events take time, and the identification of a moment at which the event is perceived to happen, or to start, is a non-trivial matter. Morton, Marcus, and Frankish (1976) reported that sequences of alternating syllables such as /ba-ma-ba-ma/ were not perceived as isochronous if they were arranged with even spacing from one syllable acoustic onset to the next. To be perceived as evenly spaced, it was necessary for the /ba-ma/ inter-onset interval to be systematically smaller than for /ma-ba/. They introduced the term P-center to describe the perceptual moment of occurrence of a syllable, analogous to the musical notion of a beat.

Subsequent work has demonstrated that the P-center does not correspond to any simple acoustic or articulatory feature, although the rise time, or period of increasing amplitude at the onset, critically affects the perception of the P-center (Scott 1993; De Jong 1994). The P-center can be thought of as an estimate of the beat

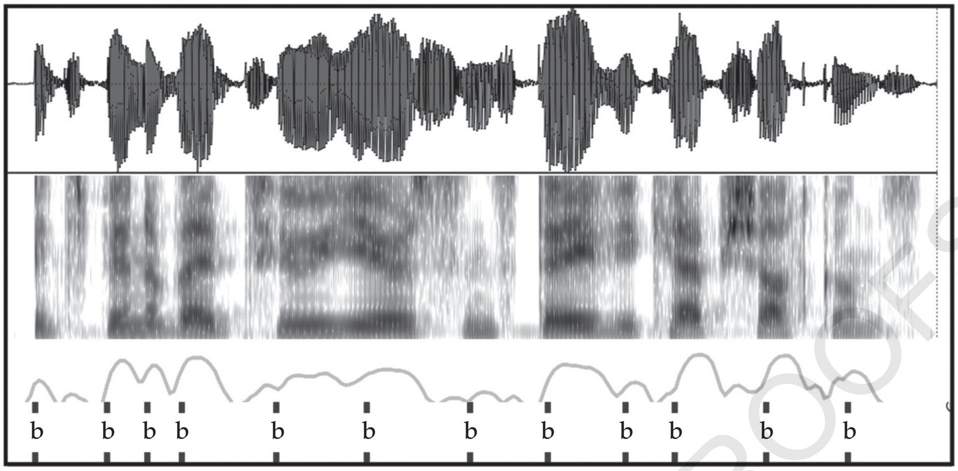

Figure 8.2 P-center estimates are placed at the mid-point of local rises in a smoothed amplitude envelope (bottom) of the filtered signal (top).

location associated with a syllable, and the concept extends naturally to musical tones as well (Vos and Rasch 1981).

A simple algorithm to calculate a P-center estimate, based on prior work by Scott (1993), is provided in Cummins and Port (1998). It is illustrated in Figure 8.2. Speech is first bandpass filtered with cut off frequencies chosen to largely exclude energy directly attributable to the fundamental frequency, and to fricative noise. P-center estimates are placed at the mid-points of local rises in smoothed amplitude envelope of the filtered signal. This algorithm generates estimates based on the physical characteristics of the signal, and the care of the phonetician is still required to assess the relevance of such estimates to the perception of rhythmically salient events.

8.2.2 *Stress-timing and syllable-timing*

As texts vary, so too does the rhythm of the speech they generate. Lloyd James (1940) observed that two kinds of temporal regularity are notable in speaking, which he dubbed machine gun and Morse code styles. It should be noted that these were transitory aspects of speech, and could both be found within the speech of an individual. Kenneth Pike (1945) renamed these patterns as syllable-timed and stress-timed speech, respectively. Those familiar with Martin Luther King's "I have a dream" speech can find reasonably clear examples of each of these in the two phrases "[will be able to] SPEED UP THAT DAY" (syllable timing) and "BLACK men and WHITE men, JEWS and GENtiles, PROTestants and CATHolics" (stress timing).

In the 1960s, these two impressionistic labels acquired a new use, being interpreted as features of whole languages, rather than specific utterances. David Abercrombie generated an enduring linguistic myth when he made the strong typological claim:

> As far as is known, every language in the world is spoken with one kind of rhythm or with the other. In the one kind, known as a syllable-timed rhythm, the periodic recurrence of movement is supplied by the syllable-producing process: the chest-pulses, and hence the syllables, recur at equal intervals of time – they are isochronous. French, Telugu, Yoruba illustrate this mode of co-ordinating the two pulse systems: they are syllable-timed languages. In the other kind, known as a stress-timed rhythm, the periodic recurrence of movement is supplied by the stress-producing process: the stress-pulses, and hence the stressed syllables, are isochronous. English, Russian, Arabic illustrate this other mode: they are stress-timed languages.
>
> (Abercrombie 1967: 97)

This remarkable, and demonstrably false, claim has attracted an undue amount of attention, and has been unquestioningly accepted in some quarters, such that it is regularly repeated as a factual assertion about languages. Despite the slight measurement issues that arise due to uncertainty about the exact location of a beat or pulse (see above), it is a matter of no great difficulty to test Abercrombie's assertion on a linguistic sample. This has been done many times (Classé 1939; Shen and Peterson 1962; Bolinger 1965; O'Connor 1968; Nakatani, O'Connor, and Aston 1981; Crystal and House 1990), and each and every such study has falsified the claim, though many have sought to maintain something of the essence of the claim by appealing to unobservable "perceptual isochrony" (Lehiste 1977; Donovan and Darwin 1979), or by positing an intermediate position between syllable- and stress-timing for specific languages (Balasubramanian 1980; Major 1981; de Manrique and Signorini 1983; Miller 1984). Perhaps the most thorough debunking of the isochrony hypothesis, as Abercrombie's clam has come to be known, was provided by Dauer (1983), who measured inter-stress intervals from readings of texts in English, Thai, Spanish, Italian, and Greek. She found no more inter-stress isochrony in English than in any of the other languages. All languages measured showed a weak tendency for stresses to recur regularly, much as Classé had found in 1939. Dauer persuasively argued that impressionistic accounts of "rhythmic" differences among the languages probably had to do with a variety of factors affecting signal variability, including differences in syllable structure, vowel reduction, and the phonetic realization of stress, rather than with the temporal patterning of stressed syllable onsets. Tellingly, she noted:

> The concept of syllable-timing was originally developed by English speakers to describe a kind of rhythm that is opposite to that of English, that is, it has been defined primarily negatively. However, the label has not been widely accepted by native speakers of those languages described as such.
>
> (Dauer 1983: 60)

A third "rhythm class" has sometimes been claimed, also based on notions of an isochronous timing unit, but in this case it is the Japanese mora, rather than the syllable or stress foot, that has traditionally been claimed to be of equal duration (Port, Dalby, and O'Dell 1987; Han 1994). The mora is often coextensive with the syllable, as in the simple CV form (e.g., ke, ya, etc.). Geminate consonants and long vowels contain two morae, and a nasal may also be a whole mora, so that, for example, Honda has two syllables, but three morae (ho-n-da), and the place name Tokyo has two syllables, but four morae (to-o-k^{j}o-o). Traditional Japanese pedagogy had maintained that morae were of equal duration, and this had been roundly disputed by phoneticians (Beckman 1982). Port et al. demonstrated that words with increasing numbers of morae increase in duration by almost constant increments as morae were added, so that the locally computed average duration of a mora remained constant, with non-local variation in timing distributed over several morae contributing to the net effect. Thus the intuition about even mora timing rested, not on isochrony, but on a statistical property of morae in combination.

One possible reason for the sustained controversy about notional isochrony in speech has been the non-trivial issue of the domain in which isochrony might be observed. Proponents of the direct realist approach to perception have suggested that listeners directly perceive articulatory events, "seeing through," as it were, the acoustic signal to the generative acts from which they arise (Fowler 1979). Articulatory studies have failed to produce evidence for isochrony in this domain however (De Jong 1994). Others have suggested that isochrony is not to be found in the physical signal at all (acoustic or articulatory), but is rather best understood as a perceptual phenomenon (Lehiste 1977). This suggestion seems to remove the hypothesis from the remit of empirical inquiry. Scott, Isard, and de Boysson-Bardies (1985) found that the tendency to perceive events as more regular than they are was generic, not specific to any language or to speech, and so could not be used to support an isochrony hypothesis for English.

Two issues have become confused in this debate. There is first a question of whether speech is rhythmic in the specific sense of providing a sequence of events that are evenly spaced in time. This question, which must usually be answered in the negative, can only be approached on the basis of some specific sample of speech, which may or may not satisfy some criterion of representativeness of a specific language (or dialect, or speaking style, or genre). The second issue is whether languages (abstract entities such as English, Tamil, etc.) fall into two or three distinct classes based on some acoustic properties that might loosely be called "rhythm." This second hypothesis, let us call it the rhythm class hypothesis, has had further development beyond matters pertaining to isochrony.

8.3 The rhythm class hypothesis

Despite the absence of evidence for isochrony in speech, many researchers have sought to defend the supposed dichotomy on grounds other than temporal patterning (Bertinetto 1988). Ramus, Nespor, and Mehler (1999) presented some novel phonetic measures that they thought might justify a presumed classification of languages into stress-timed and syllable-timed families. The authors were heavily committed to the two-way classification, and they had shortly before demonstrated that French newborn infants could discriminate between low-pass filtered speech in Japanese and English, but not between Dutch and English. They could also discriminate between the sets {English, Dutch} and {Spanish, Italian}, but not between the sets {English, Spanish} and {Dutch, Italian}. Of course, these discrimination results in no way confirm that languages fall into two groups, but they are certainly compatible with such a hypothesis, if it were to be established on independent grounds. They arrived at two (correlated) variables, defined over an utterance: the proportion of vocalic intervals (%V) and the standard deviation of the duration of consonantal intervals (ΔC).

Results from eight languages are shown in Figure 8.3 (top). These stem from four speakers per language, reading five short declarative sentences each. At first glance, there appear to be two distinct clusters, and one outlier. The clusters group

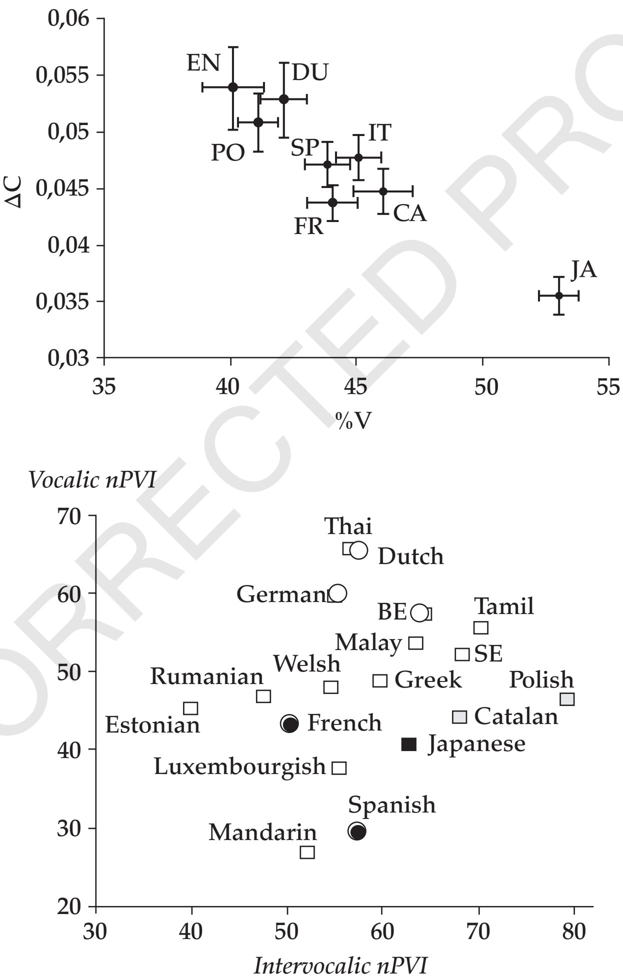

Figure 8.3 Rhythm metrics have been used to discriminate between speech data from different languages. Results from Ramus et al. (1999) and Grabe and Low (2002) are shown in the left and right panels, respectively. Reprinted by permission of De Gruyter.

languages claimed to be stress-timed (English, Dutch, Polish) together, while the so-called syllable-timed languages (French, Spanish, Italian, Catalan) form a second group. Japanese (mora timed) is satisfyingly distant from both groups.

With similar motivation, Grabe and Low (2002) employed a measure of local timing variability originally developed by Francis Nolan, the Pairwise Variability Index, or PVI, that quantifies the degree to which successive units (often, but not necessarily, syllables) differ in duration. Two variants were employed: the raw index (rPVI):

$$\mathrm{rPVI} = \left[\sum_{k=1}^{m-1} \left| d_k - d_{k+1} \right| / (m-1) \right] \tag{8.1}$$

and a normalized form, that uses the average interval length within each pair as a normalization factor:

$$\mathrm{nPVI} = 100 \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right] \tag{8.2}$$

where m is the number of items contained in an utterance, and d_k is the duration of the kth item. The nPVI measure was applied to vowel durations, and the rPVI to the intervals between vowel onsets.

Figure 8.3 (bottom) shows comprehensive results for 18 languages, with data from a single speaker for each language reading set texts in a recording booth. One can read what one likes into the resulting distribution. The authors claimed that the data “support a weak categorical distinction between stress-timing and syllable-timing … [but] … there is considerable overlap between the stress-timed and the syllable-timed group and hitherto unclassified languages” (Grabe and Low 2002: 538). Nolan, from whom the PVI originally stems, has recently applied the measure at both syllable and foot level for four languages (Estonian, English, Mexican Spanish, and Castilian Spanish) (Nolan and Asu 2009). Five speakers of each read a short text to provide the data. There were serious methodological problems in defining units, especially the foot, in comparable fashion across language. Despite these, the author argued that syllable-timing and stress-timing were orthogonal dimension, such that a given language might exhibit characteristics of either, both, or neither.

Several related metrics have subsequently been proposed, any of which might serve to locate languages in a low-dimensional “rhythm-space.” Galves et al. (2002) proposed a sonority-based measure that obviated the need for manual annotation of the speech material. Gibbon and Gut (2001) contributed another, and Wagner and Dellwo (2004) provided yet another variant on the PVI in the service of more or less the same goals. Common to all these approaches is the use of a small (sometimes very small) corpus of read text as the source material that is held to represent the language in question, without consideration of variation within a language. Common to them all is also a rather refined sense of the term "rhythm" that seems to lie quite distant from the core of the term in its musical sense.

The task of identifying objective correlates of speech rhythm is complicated by the fact that perceived temporal properties of speech are influenced by many factors in ways still poorly understood. These include the role of perceived pitch, the presence and strength of accents, and prominences more generally, the duration and distribution of pauses, and the complex effects of speech rate (Dellwo and Wagner 2003; Zvonik and Cummins 2002; Trouvain and Barry 2000; Farnetani and Kori 1990). Arvaniti (2009, 2012) has persuasively argued that studies employing rhythm metrics typically assess the merits of their approach by appeal to the degree to which they support the existing and presumed classification of specific languages. They are typically not at all robust to inter-speaker variation, or elicitation method, rendering their utility in contributing to the rhythm class debate problematic at best.

Several theoretical approaches to speech have suggested that the production and the perception of speech may be very intimately intertwined. This gave rise to the venerable Motor Theory of Speech Perception (Liberman and Mattingly 1985) which posited shared representations, and, with different motivation, to the theory of Articulatory Phonology (Fowler et al. 1980; Browman and Goldstein 1995), which entertains the notion that the abstract units of linguistic contrast that give rise to phonological systematicity are one and the same thing as units of movement, or phonetic gestures. Recent neuroscientific evidence has provided strong evidence that the neural substratum for the production of goal-directed action is not separable from the means by which such actions are perceived (Rizzolatti and Arbib 1998; Goldstein, Byrd, and Saltzman 2006).

Collectively, these approaches and insights suggest that the generation of rhythmic speech may have implications for how speech is perceived. Rhythmic expectation can be construed as a means by which listeners predict what is coming up in the speech signal, and rhythm would thus play a role in the allocation of scarce attentional resources to specific, rhythmically salient, moments in time (Large and Riess Jones 1999). This kind of role for rhythmic structure in speech has been suggested to facilitate the parsing of the speech stream (Cutler and Mehler 1993), and the acquisition of both first (Morgan 1996) and, perhaps, second languages (Wenk 1985).

Whether chasing isochrony, or seeking to underwrite a classification of languages into two or three classes, much of the discussion about rhythm in speech has moved away from the sense of rhythm that is grounded in real time performance, and that is best exemplified by the compulsion to tap one's foot along with a tune. It is to such performative considerations that we now turn.

8.4 Rhythm and fluency

Alterations to speech rhythm are frequently noted in a wide range of speech pathologies, and as a supervening symptom in many kinds of movement and psychological disorders. When the word "rhythm" is employed here, it is typically the case that an extended sense of the term is meant, that overlaps greatly with the notion of "fluency," and that does not admit of a simple operationalization. Prosodically altered speech that gives rise to the perception of altered rhythm may exhibit changes in the distribution and duration of pauses, in the timing of segments or supra-segmental units, in the degree of reduction in unstressed syllables, in the features of the intonational contour, especially in the way in which prominences are signaled, and more besides. Although durational measurements may be employed to illustrate changes in "rhythm," it is clear that the impression of altered speech rhythm does not derive from a single factor alone. Likewise, rhythm is unlikely to be affected in isolation in any given pathology (see, e.g., the multiple alterations found in so-called foreign accent syndrome; Kurowski, Blumstein, and Alexander 1996). Impressionistic labels of "stress-timing" or "syllable-timing" are frequently used to characterize speech with global prosodic alteration, for example, in autistic or schizophrenic individuals (Paul et al. 2008; Goldfarb et al. 1972), or after brain trauma (Knight and Cocks 2007). The literature is heavily biased toward reports of cases in which English is the principal language, which may explain why reports of a change toward syllable-timing are common, but reports of a change from syllable-timing to stress-timing are virtually nonexistent. It has been pointed out that the labels probably do not refer to well-defined language types, and their use in cases of pathological prosody may instead reflect a deviation from canonical, fluent, and expressive speech.

Any sense of rhythm in continuous speech demands that the speech be fluent. This is true of other forms of movement too, and there are many parallels to be drawn between rhythmically disturbed, or dysfluent, speech, and dysfluent movement in other domains. Stuttering provides a domain in which the fluency of speech is threatened, due to difficulties in both the initiation of speech, and its fluent continuous production (Starkweather 1987). Initiation difficulties frequently lead to long pauses, or to multiple attempts to start a single utterance, or a single prosodic unit such as a syllable. This can give rise to repetition, which can also be seen as a frustrated attempt to move on to the next unit of production. Once speech is initiated, characteristic rhythmic disturbances include the prolongation of segments (Yaruss 1997). Stuttering is not simply a timing problem, as evidenced, for example, by a study by Max and Yudman (2003) in which stutterers and non-stutterers performed at entirely equivalent levels in a task that required synchronization of either finger taps or spoken syllables with a metronome. Yet stutterers have been found to display subtle differences compared to non-stutterers on a variety of coordinative tasks, including imitative and shadowing tasks (Starkweather 1987; Nudelman et al. 1987; Williams and Bishop 1992). In many respects, the coordinative and rhythmic problems displayed by stutterers are similar to gross movement deficits seen in patients with Parkinson's disease. This neurological disease, typically linked to pathology of the dopamine system, is readily recognized by the characteristic movement tremor, gait difficulties, and dysfluencies of sufferers. Parkinson's patients frequently display freezing, in which a desired movement, such as walking, is inhibited. For both Parkinson's and stuttering patients, a wide variety of non-specific forms of intervention can help to overcome movement problems (Andrews et al. 1982). These can include moving/speaking at an altered tempo, typically a slower tempo, or by changing the context of production, for example, by getting a frozen walker to step "over" an imaginary stick, or by getting a stutterer to sing, instead of speak.

The study of dysfluency in speech points to the deep relation that obtains between rhythm, fluency, and the coordination of movement. Of the many senses in which the term "rhythm" is applied, one central use is to distinguish between movement sequences that are fluid, skilled, and effortless, in contrast to those which seem disjointed, clumsy, or effortful. Some further insight into rhythm in speech is revealed by consideration of the characteristics of skilled movement, in which rhythm may be usefully viewed as an emergent and gradient phenomenon.

8.5 Rhythm as an emergent phenomenon

In the study of coordinated movement, one of the most profound insights of the last hundred years has been the realization that generic dynamical principles underlie the self-organization of complex systems into simpler, task-specific assemblies suited to specific behavioral goals like walking, reaching, etc. (Latash 2008). Thus, in studying locomotion in the jellyfish, the millipede, the ape, and the bird, common principles can be found, such as the recruitment of multiple body parts into phase-locked coordinative domains in which each limb/effector adopts a fixed cyclic offset with respect to the others (Grillner 1981). A model of this form of coordination had been developed in great detail by Scott Kelso and colleagues, taking the two hands as effectors, and constraining movement such that two fingers are wagged at identical frequencies (Haken, Kelso, and Bunz 1985; Kelso 1995). Just as with multi-legged gaits, the simultaneous cycling of these two effectors can only be performed in a stable fashion when the fingers adopt one of two simple phase relations: either they cycle in synchrony (in phase) or in syncopated (anti-phase) manner. As with gaits, the relative stability of the two forms of coordination depends on rate, and characteristic transitions from the less stable (syncopated) pattern to the more stable (in phase) pattern reliably occur at fast rates. For our present purposes, the importance of this well-studied and modeled system is that it suggests that multiple parts of the body, when performing a periodic task, will spontaneously adopt specific stable configurations, and will shift discretely from one pattern to the other. This is a generic dynamical principle of biomechanical self-organization, and it has been tested in the speech domain as well, despite the manifest dissimilarities between the effectors of locomotion and the articulators of the vocal tract.

In the speech cycling experimental paradigm, a short phrase is repeated in time with an auditory metronome. A canonical example is the targeted speech cycling reported in Cummins and Port (1998), where a short phrase, such as "big for a duck," is repeated along with a series of alternating high and low tones. The high tone sequence cues phrase onset, while the low tone provides a temporal target for the onset of the final stressed syllable (duck). It is quickly apparent that cyclic

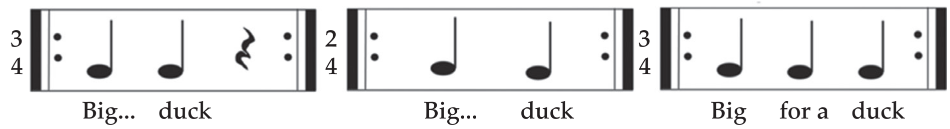

Figure 8.4 Schematic of three stable rhythm patterns produced during the repetition of the phrase *big for a duck*.

repetition like this is highly constrained, and the constraint lies in the temporal relationship between the sequence of syllables, and their organization into larger units, here the foot and the phrase. When the phase (relative time) of the low tone is varied from trial to trial, it becomes clear that some positions of the stressed syllable onset within the repeating phrase cycle are relatively natural, and can be maintained in a stable fashion, while others cannot be so produced. Figure 8.4 shows a schematic representation of the three stable patterns so produced. The last of the three (with a medial phase of 0.66) is less frequent, while the second (phase = 0.5) is the most stable, and the most likely to occur at fast rates. This work suggests that the rhythmic constraints that are apparent in repetitive speech production are of a kind with rhythmic constraints on cyclic movement of the limbs, for example, in juggling, walking, or dancing, and that the temporal patterns arise from generic dynamical principles of self-organization in complex systems, rather than from specific properties of the articulators. It also shows that under appropriate task constraints, speech can, indeed, be produced isochronously.

The idea that speakers/listeners may become mutually entrained to each other during conversational interaction has been suggested on several accounts (Richardson, Dale, and Kirkham 2007; Cummins 2009b, 2009a). Using transcranial magnetic stimulation to reveal weak excitation in muscles, Fadiga et al. (2002) found that there was a highly specific modulation of tongue activation as a function of the speech being perceived by a listener. That is to say, the speech production mechanisms of the listener were being selectively activated, or entrained, by the speech being produced by another. Condon and Sander (1974) observed that the movements of neonates became synchronized with the speech of the mother. Shockley, Santana, and Fowler (2003) documented entrainment in postural sway among standing conversational participants who were engaging in a collaborative speech task.

An experimental variant on choral speaking has been introduced by Cummins (2009b). In a synchronous speech task, two subjects read a prepared text together, attempting to remain in synchrony. This task proves to be easy to do, and on average, asynchrony of approximately 40 ms is observed, rising to a mean asynchrony of 60 ms at phrase onsets (Cummins 2003). Practice does not seem to greatly improve performance (Cummins 2003). The ability of speakers to maintain such tight temporal alignment in the absence of any underlying periodic structure or beat sequence poses something of a challenge, and suggests that entrainment among speakers may provide an alternative way of conceptualizing the role of rhythm in speech. Rhythm thus plausibly has an alternative characterization as a means by which bodily movement becomes entrained across individuals. This view of rhythm also seems to be continuous with its role in music and dance (Cummins 2009a).

8.6 Models

In modeling the form and function of rhythm in speech, two large classes of model can be discerned, and these two classes shadow a long-standing debate in the literature of speech production and motor control more generally about the degree to which temporal structure is controlled (extrinsic, or clock, timing models) or is emergent (intrinsic timing) (Keller 1990; Thelen 1991). Some of this debate has a somewhat anachronistic feel to it today, as we are somewhat more accustomed to working with a plurality of modeling approaches, without insisting on the primacy of one over the other (Lubker 1986). At the heart of the debate, however, lies the important issue of whether time is a controlled variable, measured in the process of perception, and doled out in the act of production, or whether temporal structure is an emergent property of suitably constrained and parameterized dynamical systems.

Models that regard time as a controlled variable typically include a role for a clock, or time- keeping process. One of the most influential of these is the timing model of Wing and Kristofferson (1973) in which a central timekeeper is distinguished from peripheral movement processes. This model has been widely applied in simple repetitive tasks, such as finger tapping. A timekeeper component of this kind is an important element in the EXPLAN model developed by Howell and colleagues, specifically to model dysfluency in speech production, as in stuttering (Howell and Au-Yeung 2002). The explicit computation of temporal intervals has long been a mainstay of rule-based approaches to speech synthesis (Allen, Hunnicutt, and Platt 1987). Including clock or timekeeper components within a model allows one to dictate temporal patterns of arbitrary complexity. In this sphere, rhythmic patterns are privileged primarily because they are simpler than other patterns.

Models in which time is explicitly metered in production or measured in perception belong squarely in the class of cognitivist computational models that represented a preeminent orthodoxy within the cognitive sciences throughout the last two decades of the twentieth century. A large field of alternative accounts has since become prominent, emphasizing the embeddedness of the organism in an environment, the ineliminable role of the body in any perceptual or active processes, and the emergence of domains of lawfulness that transcend the boundaries between brains, bodies, and the world. These are often referred to (somewhat inaccurately) as embodied or enactive theories of cognition, and the tools and concepts of dynamical systems theory find application where cognitivist models employ rule-based transformations over abstract representations. The coordination dynamics of Scott Kelso, and its application in the speech cycling paradigm, were already mentioned above (Kelso 1995; Cummins and Port 1998). A good primer on the basic concepts of dynamical systems is found in Norton (1995).

The emergence of temporal structure, without explicit metering of time, is a characteristic of dynamical systems models. Here, model components are typically oscillatory systems with intrinsic periods, which may be modified in interaction with other such systems. When self-sustaining oscillators interact weakly, they will tend to coordinate their activity, bringing their frequencies into relatively simple relative timing relations, such as 1:1, 2:3, etc. The principles by which oscillating systems tend to coordinate and adopt relatively simple mutual temporal relations are entirely generic, and depend on their dynamical properties, rather than their material substrate (Pikovsky, Rosenblum, and Kurths 2001). Oscillator models provide a natural platform for capturing rhythmic patterning, and they have been widely used in speech studies (O'Dell and Nieminen 1999; Barbosa 2002; Nam and Saltzman 2003). Although these models generate a wide variety of rhythmic phenomena, a limitation in their application to speech has been the absence of clearly defined periodic patterns in actual speech production.

There remains a tension between intrinsic and extrinsic timing approaches that mirrors a larger debate within cognitive science. The computational approaches arising from decades of work within artificial intelligence and cognitive psychology are currently being challenged from a number of quarters by approaches to understanding behavior as a property of embodied beings embedded lawfully in structured environments. This is a very large debate that goes beyond our present concerns.

8.7 Open questions in the study of speech rhythm

The many themes that arise in the study of rhythm ensure that there will always be a rich variety of phenomena to be studied, and a correspondingly plurality in theoretical and modeling approaches employed. This can be confusing to the newcomer, and it is incumbent on researchers within any of these many fields to make explicit their understanding of central concepts such as rhythm, meter, entrainment, coupling, and more. The equation of mere periodicity with the richer set of phenomena deserving of the term "rhythm" constitutes a clear source of conceptual confusion throughout the literature which future work would do well to avoid. Some of the principal areas that have hitherto defined the study of rhythm in speech include the classification of languages, the characterization of speaking styles, the role of rhythm in dysfluencies, and the way in which coordination emerges in speaking. Many of these will continue to be fruitful areas of inquiry, although one might surmise from the above discussion that the vigorous pursuit of a classificatory scheme for languages on rhythmic grounds alone has probably enjoyed an undue amount of attention, with little success.

A large area of recent interest arises from studying rhythm and timing at the dyadic level, or, more generally, as a property of multi-party interaction. If an informal use of the term may be allowed, rhythms emerge in conversational interaction; they arise, are sustained, and disappear again in the ebb and flow of attention and activity among participants. The dynamics of turn-taking has long been hypothesized to be guided by rhythmic principles (Couper-Kuhlen 1993), although empirical studies of the timing of turns has yet to deliver a robust account that is grounded in quantitative observation (Bull 1996). Part of the difficulty encountered lies in the great degree of temporal variability exhibited by pauses in speech (Trouvain and Barry 2000; Zvonik and Cummins 2002), and recent investigation of turn-taking when speakers overlap may open new avenues here (Wlodarczak, Simko, and Wagner, 2012). Beyond turn-taking, the employment of dynamical models may allow the characterization of collective temporal phenomena that are poorly, if at all, identifiable when speech is considered one individual at a time. Synchronized speech represents one emerging topic in this field (Cummins 2012). It is somewhat perplexing that the ubiquitous phenomenon of joint, or choral, speaking has received so little attention by empirical studies, especially when one considers the deep integration of collective speaking practices in educational institutions, houses of worship, sports stadia, and street protests throughout the world.

With the rise of interest in the role of movement in rhythmic behaviors, the study of rhythm in speech is increasingly taking stock of the rich body of work on gestures and whole-body involvement in speaking (McNeill 1992; Goldin-Meadow 1999). Manual gestures, facial movements, even gaze and blinking are potentially co-implicated in the temporal patterning that is speaking (Cassell et al. 1994; Leonard and Cummins 2011; Cummins 2011), and much remains to be uncovered about how these disparate streams are integrated within and across individuals.

Finally, although the approach here has tried to stay close to the core sense of "rhythm" that is home in the domain of music and dance, there is much to be done in fleshing out the continuum that exists between the spoken word and the use of the voice in a musical context. Parallels between music and speech may range from the merely metaphorical to the literal. Anniruddh Patel has contributed a very varied set of studies that have contributed to our understanding of the way in which musical and speech rhythm might relate to one another (Patel et al. 1998; Patel and Daniele 2003), but much remains to be done.

NOTES

1 The notion of "cognitive rhythms" that arose in the 1960s might bear mention in passing, if only to warn newcomers to the field that that particular construct is not theoretically sound and is no longer part of the state of the art (Henderson, Goldman-Eisler, and Skarbek 1966; Goldman-Eisler 1967; Jaffe, Breskin, and Gerstman 1972; Kowal and O'Connell 1985).

REFERENCES

Abercrombie, D. 1965. A phonetician's view of verse structure. In *Studies in Phonetics and Linguistics* 16–25. Oxford: Oxford University Press.

Abercrombie, D. 1967. *Elements of General Phonetics*. Chicago, IL: Aldine.

Allen, J., M. Hunnicutt, and D. Klatt. 1987. *From Text to Speech: The MITalk System*.

Cambridge: Cambridge University Press.
Andrews, G., P. Howie, M. Dozsa, and B.E. Guitar. 1982. Stuttering: Speech pattern characteristics under fluency-inducing conditions. *Journal of Speech and Hearing Research* 25: 208–216.
Arvaniti, A. 2009. Rhythm, timing and the timing of rhythm. *Phonetica* 66(1–2): 46–63.
Arvaniti, A. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40(3): 351–373.
Balasubramanian, T. 1980. Timing in Tamil. *Journal of Phonetics* 8: 449–467.
Barbosa, P.A. 2002. Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production. In *Proceedings of Speech Prosody 2002, Aix-en-Provence*, 163–166.
Beckman, M. 1982. Segment duration and the "mora" in Japanese. *Phonetica* 39: 113–135.
Bergeson, T. and S. Trehub. 2002. Absolute pitch and tempo in mothers songs to infants. *Psychological Science* 13(1): 72–75.
Bertinetto, P.M. 1988. Reflections on the dichotomy "stress" vs. "syllable" timing. *Quaderni del Laboratorio di Linguistica* 2: 59–84 (Scuola Normale Superiore di Pisa).
Bolinger, D. 1965. Pitch accent and sentence rhythm. In I. Abe and T. Kanekiyo (eds.), *Forms of English: Accent, Morpheme, Order*, 139–180. Cambridge, MA: Harvard University Press.
Browman, C.P. and L. Goldstein. 1995. Dynamics and articulatory phonology. In R.F. Port and T. van Gelder (eds.), *Mind as Motion*, 175–193. Cambridge, MA: MIT Press.
Bull, M.C. 1996. An analysis of between-speaker intervals. In *Proceedings of the Edinburgh Linguistics Department Conference '96*, 18–27.
Cassell, J., C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. 1994. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, 413–420.
Classé, A. 1939. *The Rhythm of English Prose*. Oxford: Basil Blackwell.
Condon, W.S. and L.W. Sander. 1974. Synchrony demonstrated between movements of the neonate and adult speech. *Child Development* 45: 456–462.
Couper-Kuhlen, E. 1993. *English Speech Rhythm*. Philadelphia, PA: John Benjamins.
Crystal, T.H. and A.S. House. 1990. Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America* 88(1): 101–112.
Cummins, F. 2003. Practice and performance in speech produced synchronously. *Journal of Phonetics* 31(2): 139–148.
Cummins, F. 2009a. Rhythm as an affordance for the entrainment of movement. *Phonetica* 66(1–2): 15–28.
Cummins, F. 2009b. Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics* 37(1): 16–28.
Cummins, F. 2011. Gaze and blinking in dyadic conversation: A study in coordinated behavior among individuals. *Language & Cognitive Processes*. Published online.
Cummins, F. 2012. Periodic and aperiodic synchronization in skilled action. *Frontiers in Human Neuroscience* 5: 170.
Cummins, F. and R.F. Port. 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26(2): 145–171.
Cutler, A. and J. Mehler. 1993. The periodicity bias. *Journal of Phonetics* 21: 103–108.
Dauer, R.M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11: 51–62.
De Jong, K.J. 1994. The correlation of p-center adjustments with articulatory and

acoustic events. *Attention, Perception, & Psychophysics* 56(4): 447–460.
de Manrique, A.M.B. and A. Signorini. 1983. Segmental duration and rhythm in Spanish. *Journal of Phonetics* 11: 117–128.
Dellwo, V. and P. Wagner. 2003. Relations between language rhythm and speech rate. In *Proceedings of the 15th International Congress of Phonetics Sciences*, 471–474.
Donovan, A. and C.J. Darwin. 1979. The perceived rhythm of speech. In *Proceedings of the Ninth International Congress of Phonetic Sciences*, vol. 2, 268–274.
Fadiga, L., L. Craighero, G. Buccino, and G. Rizzolatti. 2002. Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience* 15(2): 399–402.
Farnetani, E. and S. Kori. 1990. Rhythmic structure in Italian noun phrases: A study on vowel durations. *Phonetica* 47: 50–65.
Fowler, C. 1979. Perceptual centers in speech production and perception. *Attention, Perception, & Psychophysics* 25(5): 375–388.
Fowler, C.A., P. Rubin, R. Remez, and M. Turvey. 1980. Implications for speech production of a general theory of action. In B. Butterworth (ed.), *Language Production*, 373–420. San Diego, CA: Academic Press.
Galves, A., J. Garcia, D. Duarte, and C. Galves. 2002. Sonority as a basis for rhythmic class discrimination. In *Proceedings of Speech Prosody 2002, Aix-en-Provence*.
Gibbon, D. and U. Gut. 2001. Measuring speech rhythm. In *Proceedings of the Seventh European Conference on Speech Communication and Technology*.
Goldfarb, W., N. Goldfarb, P. Braunstein, and H. Scholl. 1972. Speech and language faults of schizophrenic children. *Journal of Autism and Developmental Disorders* 2(3): 219–233.
Goldin-Meadow, S. 1999. The role of gesture in communication and thinking. *Trends in Cognitive Sciences* 3(11): 419–429.
Goldman-Eisler, F. 1967. Sequential temporal patterns and cognitive processes in speech. *Language and Speech* 10: 122–132.
Goldstein, L., D. Byrd, and E. Saltzman. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In M.A. Arbib (ed.), *Action to Language via the Mirror Neuron System*, 215–249. Cambridge: Cambridge University Press.
Grabe, E. and E. Low. 2002. Durational variability in speech and the rhythm class hypothesis. *Laboratory Phonology* 7.
Grillner, S. 1981. Control of locomotion in bipeds, tetrapods, and fish. In V.B. Brooks (ed.), *Handbook of Physiology, Motor Control*. Baltimore, MD: Williams and Wilkins.
Haken, H., J.A.S. Kelso, and H. Bunz. 1985. A theoretical model of phase transitions in human hand movement. *Biological Cybernetics* 51: 347–356.
Han, M.S. 1994. Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America* 96: 73–82.
Henderson, A., F. Goldman-Eisler, and A. Skarbek. 1966. Sequential temporal patterns in spontaneous speech. *Language and Speech* 9: 207–216.
Howell, P. and J. Au-Yeung. 2002. The EXPLAN theory of fluency control applied to the diagnosis of stuttering. In E. Fava (ed.), *Clinical Linguistics: Theory and Applications in Speech Pathology and Therapy*, 75–94. Amsterdam: John Benjamins.
Huron, D. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.
Jaffe, J., S. Breskin, and L.J. Gerstman. 1972. Random generation of apparent speech rhythms. *Language and Speech* 15: 68–71.
Jones, D. 1956. *An Outline of English Phonetics*, 8th edn. Cambridge: Heffer. First published in 1918.
Keller, E. 1990. Speech motor timing. In W.J. Hardcastle and A. Marchal (eds.), *Speech*

Production and Speech Modelling, 343–364. Dordrecht: Kluwer Academic.

Kelso, J.A.S. 1995. *Dynamic Patterns*. Cambridge, MA: MIT Press.

Knight, R. and N. Cocks. 2007. Rhythm in the speech of a person with right hemisphere damage: Applying the pairwise variability index. *International Journal of Speech-Language Pathology* 9(3): 256–264.

Kowal, S.H. and D.C. O'Connell. 1985. Cognitive rhythms reluctantly revisited. *Language and Speech* 28(1): 93–95.

Kuiper, K. 1992. The oral tradition in auction speech. *American Speech* 67(3): 279–289.

Kurowski, K., S. Blumstein, and M. Alexander. 1996. The foreign accent syndrome: A reconsideration. *Brain and Language* 54(1): 1–25.

Large, E.W. and M. Riess Jones. 1999. The dynamics of attending: How people track time-varying events. *Psychological Review* 106(1): 119–159.

Latash, M. 2008. *Synergy*. Oxford: Oxford University Press.

Lehiste, I. 1977. Isochrony reconsidered. *Journal of Phonetics* 5: 253–263.

Leonard, T. and F. Cummins. 2011. The temporal relation between beat gestures and speech. *Language and Cognitive Processes* 26(10): 1457–1471.

Liberman, A.M. and I.G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21: 1–36.

Liberman, M. and A. Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249–336.

Lloyd James, A. 1940. *Speech Signals in Telephony*. London: Pitman.

Lubker, J. 1986. Articulatory timing and the concept of phase. *Journal of Phonetics* 14: 133–137.

Major, R.C. 1981. Stress-timing in Brazilian Portuguese. *Journal of Phonetics* 9: 343–351.

Max, L. and E.M. Yudman. 2003. Accuracy and variability of isochronous rhythmic timing across motor systems in stuttering versus nonstuttering individuals. *Journal of Speech, Language, and Hearing Research* 46: 146–163.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.

Miller, M. 1984. On the perception of rhythm. *Journal of Phonetics* 12: 75–83.

Morgan, J.L. 1996. A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language* 35(5): 666–688.

Morton, J., S. Marcus, and C. Frankish. 1976. Perceptual centers (P-centers). *Psychological Review* 83(5): 405–408.

Nakatani, L.H., K.D. O'Connor, and C.H. Aston. 1981. Prosodic aspects of American English speech rhythm. *Phonetica* 38: 84–106.

Nam, H. and E. Saltzman. 2003. A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 2253–2256.

Nolan, F. and E. Asu. 2009. The pairwise variability index and coexisting rhythms in language. *Phonetica* 66(1–2): 64–77.

Norton, A. 1995. Dynamics: An introduction. In R.F. Port and T. van Gelder (eds.), *Mind as Motion: Explorations in the Dynamics of Cognition*, 45–68. Cambridge, MA: Bradford Books/ MIT Press.

Nudelman, H.B., K.E. Herbrich, B.D. Hoyt, and D.B. Rosenfield. 1987. Dynamic characteristics of vocal frequency tracking in stutterers and nonstutterers. In H.F.M. Peters and W. Hulstijn (eds.), *Speech Motor Dynamics in Stuttering*, 162–169. New York: Springer.

O'Connor, J.D. 1968. The duration of the foot in relation to the number of component sound-segments. *Technical Progress Report* 3: 1–6 (Phonetics Laboratory, University College London).

O'Dell, M. and T. Nieminen. 1999. Coupled oscillator model of speech rhythm. In *Proceedings of the XIVth International Congress of Phonetic Sciences*, vol. 2, 1075–1078.

Patel, A. and J. Daniele. 2003. An empirical comparison of rhythm in language and music. *Cognition* 87(1): B35–B45.

Patel, A., I. Peretz, M. Tramo, and R. Labreque. 1998. Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language* 61(1): 123–144.

Paul, R., N. Bianchi, A. Augustyn, A. Klin, and F. Volkmar. 2008. Production of syllable stress in speakers with autism spectrum disorders. *Research in Autism Spectrum Disorders* 2(1): 110–124.

Pike, K.L. 1945. *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Pikovsky, A., M. Rosenblum, and J. Kurths. 2001. *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge: Cambridge University Press.

Port, R.F., J. Dalby, and M. O'Dell. 1987. Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America* 81(5): 1574–1585.

Ramus, F., M. Nespor, and J. Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3): 265–292.

Richardson, D., R. Dale, and N. Kirkham. 2007. The art of conversation is coordination. *Psychological Science* 18(5): 407.

Rizzolatti, G. and M.A. Arbib. 1998. Language within our grasp. *Trends in Neuroscience* 21(5): 188–194.

Scott, D., S. Isard, and B. de Boysson-Bardies. 1985. Perceptual isochrony in English and in French. *Journal of Phonetics* 19: 351–365.

Scott, S.K. 1993. P-centers in speech: An acoustic analysis. PhD thesis, University College London.

Shen, Y. and G.G. Peterson. 1962. Isochronism in English. In *Studies in Linguistics*, Occasional Papers 9, 1–36. Buffalo, NY: University of Buffalo.

Shockley, K., M. Santana, and C. Fowler. 2003. Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance* 29(2): 326–332.

Starkweather, C.W. 1987. *Fluency and Stuttering*. Englewood Cliffs, NJ: Prentice Hall.

Steele, J. 1775. *An Essay towards Establishing the Melody and Measure of Speech to be Expressed and Perpetuated by Peculiar Symbols*. London: Printed by W. Bowyer and J. Nichols, for J. Almon.

Thelen, E. 1991. Motor aspects of emergent speech: A dynamic approach. In N.A. Krasnegor, D.M. Rumbaugh, R.L. Schiefelbusch, and M. Studdert-Kennedy (eds.), *Biological and Behavioral Determinants of Language Development*, 339–362. Hillsdale, NJ: Lawrence Erlbaum Associates.

Trouvain, J. and W.J. Barry. 2000. The prosody of excitement in horse race commentaries. In *Proceedings of ISCA Workshop (ITRW) on Speech and Emotion, Belfast*.

Vos, J. and R. Rasch. 1981. The perceptual onset of musical tones. *Perception and Psychophysics* 29(4): 323–335.

Wagner, P. and V. Dellwo. 2004. Introducing YARD (Yet Another Rhythm Determination) and reintroducing isochrony to rhythm research. In *Speech Prosody 2004, International Conference*. ISCA.

Wenk, B. 1985. Speech rhythms in second language acquisition. *Language and Speech* 28(2): 157–175.

Williams, H.G. and J.H. Bishop. 1992. Speed and consistency of manual movements of stutterers, articulation-disordered children, and children with normal speech. *Journal of Fluency Disorders* 17: 191–203.

Wing, A.M. and A.B. Kristofferson. 1973. Response delays and the timing of discrete motor responses. *Perception and Psychophysics* 14(1): 5–12.

Wlodarczak, M., J. Simko, and P. Wagner. 2012. Temporal entrainment in overlapped speech: Cross-linguistic study. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association, 2012, Portland, OR*.

Yaruss, J.S. 1997. Clinical measurement of stuttering behaviors. *Contemporary Issues in Communication Science and Disorders* 24: 33–44.

Zvonik, E. and F. Cummins. 2002. Pause duration and variability in read texts. In *Proceedings of the ICSLP, Denver, CO*, 1109–1112.

FURTHER READING

The following works provide a variety of entry points to the diverse senses of rhythm as applied in speech research. Most are broader in topic and should help to situate rhythm research with respect to other, related, fields.

Couper-Kuhlen, E. 1993. *English Speech Rhythm*. Philadelphia, PA: John Benjamins. This work is rich in following intuitions about the role of rhythm in conversation, though somewhat light on empirical investigation.

Dauer, R.M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11: 51–62. Not a book, but if you only ever read one text on the isochrony debate, this would be a good choice.

Huron, D. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press. This book is primarily about music, but it makes explicit links to speech, and to rhythm in speech.

Kelso, J.A.S. 1995. *Dynamic Patterns*. Cambridge, MA: MIT Press. This book summarizes one of the best worked-out examples of the application of dynamical systems modeling to human behavior. Relatively little on speech, but provides a good foundation with which to tackle subsequent work in dynamical modeling of speech.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press. Focuses on the form and function of gestures, which are increasingly important as embodied theories of speech production and perception gain traction.

Patel, A.D. 2008. Music, Language, and the Brain. NY: Oxford University Press. Teases out and makes explicit links between music and speech, with special attention to the form and role of rhythm.