Pamela S. H. Bogart
University of Michigan

Abstract

Rhythm offers a rich framework with which students can improve their control over North American English pronunciation. This article overviews various perspectives of what speech rhythm entails, including research that challenges the supposition of rhythmicity, why to explore rhythm with English language learners nonetheless, and ways to facilitate that exploration of rhythm. An approach that emphasizes experimentation, kinesthetic learning of suprasegmental language features, and learner-led analysis is advocated.

Introduction

As I was preparing this text, my toddler observed me staring at her hand move to and fro as she spoke about her day and pronounced, as illustrated in Figure 1, “See, Mama?—I show my meaning with my hand.—I go back and forth, back and forth.” Her hand moved to a regular beat, punctuating her pauses and the most content-laden words in her speech. I thanked her for offering me an image with which to start this paper.

Figure 1. A toddler’s explanation of speech emphasis.

Indeed, my daughter offered more than an image—she offered a definition of the process and function of rhythm as well as a hint about how we might teach rhythm through body movement.

Rhythm is defined here as a collection of auditory features of spoken North American English language that, with at least somewhat periodic timing, underscores a speaker’s meaning. Learning rhythm involves so many aspects of spoken English that it can provide an efficient path to improving one’s control over the spoken language. This complexity of rhythm also encourages an approach to pronunciation that emphasizes making choices to modulate one’s speech, rather than matching an imagined native speaker ideal. Of course, such an approach is possible within any domain of language, but, in practice, I have found that rhythm’s inherent messiness encourages students to let go of a belief that one must adhere to a single ideal pronunciation pattern. Studying rhythm appeals to multisensory and kinesthetic learning styles, and injects fun into the classroom.

This paper examines competing descriptions of what North American English rhythm is (and whether rhythmicity even exists), discusses the many features of pronunciation and communication that can be learned through a focus on rhythm, and concludes with an overview of techniques for exploring spoken English through rhythm. North American English(es) are frequently referred to with the shorthand “English” in this paper, but this simple term does not suggest that there is only one English, nor does the use of North American English suggest that there is a singular standard on the North American continent.
Descriptions of rhythm

Competing definitions of rhythm in English exist in the literature, well-summarized by Cummins (2002). One might initially explain to students that rhythm is the spacing of strong and weak sounds over time in order to emphasize important information, and then, break down which features of language make listeners perceive a speech sound as strong or weak. The existence and importance of lexical stress and discourse-level stress is generally accepted (Setter, 2006); however, rhythm, the timing with which strong and weak syllables fall, is a concept in contention. It is valuable to share multiple hypotheses about rhythm with students as they learn to experiment with stress and fluency. Genres may also affect timing, with some stylized genres such as public oratory possibly having much more exaggerated or simple rhythm than that in informal conversations.

English as a stress-timed language

One hypothesis about North American English relies on theoretical categories of language rhythm types, contrasting stress-timed languages (e.g., English or German) and syllable-timed languages (e.g., Mandarin or French). In this model, strong beats in a stress-timed language fall equally spaced in time on the stressed syllable of content-rich words with the other syllables compressed in-between, while strong beats in syllable-timed languages fall on every syllable in fast, regular succession (or on every mora in languages such as Japanese). Pronunciation textbooks, such as Grant’s Well Said (2002) or Miller’s Targeting Pronunciation (2006), reflect this hypothesis. The emphasis in these texts is on how some syllables are stressed or unstressed through modulating volume, vowel length, pitch, articulation, and finally timing and on how to identify what to strengthen and weaken. Celce-Murcia, Brinton, and Goodwin (1996) provide a very clear description of this model for teachers.

Erickson, a linguist trained also in musicology, uses Western musical notation to approximately graph the rhythm and intonation of conversational speech. In a 2001 study, Erickson used acoustic measurement software to confirm or disaffirm his sense that English is indeed spoken with strong periodicity, even among multiple speakers in conversation. He measured the distance between the “temporal center” of each stressed syllable, indicated by increased bursts of volume, and found the distance to be consistently 1.7 ± 0.3 seconds between stressed beats, which mapped accurately onto the musical staff he had created to represent the same recorded data. To Erickson, this is proof of the periodicity of stressed elements. In this data and interpretation, it is as if speakers are musicians who do not quite hold onto a steady tempo.

Challenges to the rhythmicity of English

As valuable as the stress-timed contrastive model may be for teaching linking, consonant reduction, and intonation in North American English, a number of recent articles have challenged the accuracy of the stress/syllable/mora timing taxonomy model (e.g., Cauldwell, 2002; Jenkins, 2004; Levis, 2005). In a weaker form of the challenge, it appears that linguistic variables such as syllable structure (what consonant and vowel combinations and what length of clusters are possible), tones (pitch patterns that accord different meanings to phonetically
identical syllables, such as in Mandarin) and the presence of vowel reduction will together predict a tendency toward a more stress-like or syllabic/mora-like timing and that languages simultaneously have traits of each theoretical category in lesser or greater proportions (Setter, 2006).

In a stronger form of the challenge, particularly in Cauldwell’s (2002) research, empirical evidence suggests that no languages have equally timed syllables or stress units. In empirical studies of voice recordings, stressed syllables have not been found to be evenly spaced apart. Multiple researchers have arrived at this conclusion, despite diverse definitions of what to measure. The method of measurement of acoustic features also differs, including both computer analysis of sound waves impressionistic description by a human researcher, using the computer in this latter case primarily for accurate time marking in the recordings.

Reconciling impression and measurement

How is it possible that impressionistically, it seems that English has a regular beat governed by stress, yet this pattern does not seem to hold true in empirical studies? As more computer-aided empirical work, as advocated by Levis (2005) and Jenkins (2004), continues, it may be found that Erickson’s optimistic evidence for periodic rhythm has some merit. Indeed, Cummins (2002) and Port, Cummins, and Gasser (1995) assert that it has not yet been possible to identify what to measure acoustically in the complex system that is rhythm, but that the authors have begun identifying certain elements that do reflect the periodic systematicity that listeners report in much spontaneous speech. Port et al. (1995) actually posit that periodicity be considered a universal of human language, yet one that has multiple parameters, some of which are on or off in various languages. Such a perspective necessarily complicates the task of measuring rhythmic periodicity in any one language, and of learning the patterns of rhythmic periodicity in a second language. Cummins and Port (1998) expand this theory to propose metrical rhythm in speech as being similar to metrical coordination between limbs. They provide an interesting set of evidence to support a theory that speech production adheres to preferred patterns of metric stability.

Such theories that attempt to reconcile the apparent contradiction between an impression that speech is rhythmic, and a lack of empirical evidence are attractive, but not yet proven. If more extensive empirical research continues to disprove an impression of periodicity, English teachers and learners will need to capture the pronunciation elements that contribute to the impression of periodic rhythm and let go of the periodicity hypothesis. Teschner (2004), in a textbook for linguistics students about English pronunciation and stress, focuses on how, where, and why stress is produced in English. Teschner does not address rhythmicity, the distance in time between stressed elements. As such, his detailed description of English stress patterns could inform an approach that rejects rhythmicity. Teschner borrows the tools of poetry scansion to examine trochaic, spondaic, and iambic feet in English words and phrases, but the examination of how stressed elements relate to each other stops there, rather than continuing to the extended-discourse level, one of the possible limitations of rejecting the concept of rhythm.

Given the state of research on speech rhythm today, this paper proposes that teachers might invite students to consider that there seems to be some rhythmicity to English stressed
elements, even though we do not yet quite understand how to describe the complex model. Speech is neither music nor poetry, but both types of formalized verbal expression offer ways to understand how English speakers time their sounds in the service of meaning. The final section of this paper will offer another metaphor for speech rhythm that may accommodate the uncertainty with regard to the more recent empirical work investigating timing of stress and unstress in English.

**Why study rhythm**

Jenkins (2004, 2005) posits a series of levels of pronunciation study, starting with a *lingua franca* core that could best serve communication in English among second language speakers, a core that does not contain many of the suprasegmental features of North American English that come together to characterize rhythm. Even within a North American English context, Levis (2005) challenges the generally-accepted primacy of suprasegmental features for intelligibility, though Field (2005) cites studies that indicate that stress, a suprasegmental feature implicated in rhythm, is measurably important to intelligibility.

With these cautionary perspectives in mind, what is the value of studying rhythm in North American English? Studying pronunciation from the perspective of rhythm encourages students to become analysts of what they hear and produce. For learners in contexts where they do and want to speak with fluent speakers of North American English, rhythm, at least impressionistically, is a salient feature of the spoken language. Given what we do know about rhythm so far, even given competing descriptions of this feature of spoken English, a great deal of productive exploration and analysis is possible. Students can find in the concept of rhythm a frame that intuitively and even kinesthetically links together intonation, articulation, syllabification, volume, chunking, stretching, linking, and reduction in extended discourse. Further, as students explore, they will collide with features of spoken English that are difficult for them and that render fluent delivery difficult, providing opportunities for focused, individualized learning.

Studying rhythm also provides students a way to improve their listening comprehension. When listening for stressed elements, and expecting some predictability about when these elements will fall in time, students are identifying the most content-rich words in an utterance, usually sufficient for understanding that utterance if enough speaker-listener shared context is present. Many students who have practiced this approach to listening report that fast, informal speech used to sound like a blur, but, that with a focus on hearing the stressed elements and an awareness of the somewhat periodic pacing of these elements, key words are rising up from that blur and even what is blurry is more predictable.

Further, studying rhythm implicates grammar. Students need to use grammatical chunks, such as a noun and its modifiers or a verb and its complement, in order to segment their speech coherently. Students may even increase their overall grammatical accuracy in spontaneous speech as a result of a focus on phrasing or chunking as part of rhythm study. Of course, such a hypothesis warrants further empirical investigation, but, until then, teachers and students continue to find that studying rhythm requires greater awareness of and control over multiple features of spoken English.

Because language is an interconnected, complex system of features, one could choose many different starting points, such as intonation or grammar. Rhythm is particularly attractive because it offers new ways of seeing what students have studied from other perspectives. For teachers, rhythm helps to make class activities and individual language experimentation and rehearsal fun. Many students conclude, “I feel a lot more fluent” after even one session experimenting with speech rhythm.

Ways to learn English through rhythm

The description of ways to learn English through rhythm emphasizes experimentation. The focus here is on adult learners in North American ESL programs, learners who have solid literacy skills though the majority of techniques could be readily adapted to other learning contexts and learner populations.

With new students, a recording serves as a useful assessment starting place for both instructor and student. The student rehearses some kind of speaking or conversation, and then either the instructor transcribes an illustrative minute of that recording, or the student does so. Then each independently marks up the transcript in various ways. Table 1 indicates elements useful to analyze.

<table>
<thead>
<tr>
<th>Overall effects</th>
<th>Local features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are some words stressed more than others?</td>
<td>Pitch changes</td>
</tr>
<tr>
<td>Are some syllables longer than others?</td>
<td>Volume changes</td>
</tr>
<tr>
<td>Is emphasis at anything like regular intervals?</td>
<td>Vowel lengthening &amp; accuracy</td>
</tr>
<tr>
<td>Do more important lexical items rise to the surface in my listening?</td>
<td>Vowel reduction</td>
</tr>
<tr>
<td>Is speech organized into chunks that are grammatically or semantically coherent segments?</td>
<td>Consonant and vowel linking</td>
</tr>
</tbody>
</table>

Table 1. Features of interest in student recordings

In a full class, generating individual recordings can be logistically problematic because of the time it takes. In this context, it can be helpful to start with a chapter on stress or focus taken from a pronunciation textbook or to work together as a class to draw from everyone’s background knowledge regarding what elements in extended speech receive stress and how the voice gives stress. Students generally generate a list similar to that in Table 2.

What gets stress
The lexically stressed syllables of content words (e.g. nouns, verbs, adjectives, adverbs)
New information
Contrasting information
Emphasized information
Negative markers

How the voice signals stress
Clearly articulated consonants at syllable onset
Stressed syllable falls on a major beat
Vowel is stretched long & true to its dictionary sound
Stressed syllable is louder & more intense
Pitch changes on stressed syllable

<table>
<thead>
<tr>
<th>What doesn’t get stress</th>
<th>How the voice signals unstress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar &amp; function words</td>
<td>Consonants are linked, elided, or relaxed</td>
</tr>
<tr>
<td>Articles, most pronouns, to be, to have</td>
<td>Syllable avoids a major beat</td>
</tr>
<tr>
<td></td>
<td>Vowel is reduced</td>
</tr>
<tr>
<td></td>
<td>Volume is reduced</td>
</tr>
<tr>
<td></td>
<td>Pitch is flat, probably neutral or low</td>
</tr>
</tbody>
</table>

Table 2. Stress patterns in discourse generated by class brainstorming

The overlap among items in any quadrant is evident, but the goal is that students develop a list that is meaningful to them. In other words, the redundancy that may emerge in the brainstorming process can serve a helpful role. Rhythm becomes relevant when the class looks at the distribution of these stressed elements in time. The concept of a “major beat” presumes a fairly simple repetitive pulse underlying speech, which may be presented to students as a hypothesis and probable oversimplification rather than fact.

**Walking as a metaphor for speech rhythm**

A helpful technique for moving from stress to rhythm is offered in several texts (e.g., Grant, 2002), where students walk, placing stressed syllables on a step down, and fitting in all other syllables in between steps. Gesturing as my toddler does, clapping, tapping pencils or even chopsticks also offer a beat and a kinesthetic connection. Many students report, however, that these hand- and arm-based techniques are more distracting from their communication of meaning, and less intuitive, than walking while conducting a conversation. Walking is also a metaphor for speech rhythm that accommodates rhythm’s acoustic complexity better than western music theory or poetry scansion techniques. Just as one can pause to look at something, speed up or slow down the pace of one’s steps, speakers pause, restart, slow down, or speed up the pace of the rhythmic beat of the stressed elements in speech. Just as one might trip on an obstacle when walking, a speaker might “trip” while speaking, interrupting the tempo, or search for a word or get distracted, slowing the tempo down. When nervous, speakers often accelerate the tempo. In other words, in talking, as in walking, there may be obstacles that interrupt the regular pulse of speech rhythm. There still remain steps or downbeats in a speaker’s thought groups or phrases, even if they are not evenly spaced.

**Kinesthetic play**

In order to apply the metaphor of walking to class activities, teachers can choose from a variety of documents to facilitate student experimentation. During class time or in tutorial sessions, students choose a script with which to work, perhaps from a student self-introduction, or even a short impromptu description of the day’s weather from a student in class. Everyone stands up. Students walk, exaggerating the beat of sentence-level stressed elements by marching around the room, and chorally recite the script repeatedly. Students return to their seats, and each mark the words and pauses that fell on steps. Then student can analyze what kinds of words appear to attract this special attention. It is valuable to underscore how pauses and filler words can fall on stepped beats, thus integrating hesitations into more fluent delivery. Erickson’s (2001) observations that back-channel cues (e.g., right, uh-huh, oh really?) and turn-initiation generally fall on stressed beats during social conversation are helpful background reading for teachers and students. Finally, one can present to students the hypothesis that the speed of the beat underlying our speech is about the same as the speed we most like to walk—and that naturally quick speakers rarely stroll. Students laugh at the theory but do begin to think about timing speech to their own most comfortable walking speed, and this seems to enhance their feet-speech connection.

When physical space allows, students can break into pairs for ambulating conversation, making an effort to feel the relationship between the most content-laden words they are saying and their steps. The pairs would then return to report on the experience and perhaps demonstrate part of their conversation to the class. The process can be repeated in a few minutes of homework each day: walking while generating a stream-of-consciousness set of thoughts in English, perhaps just describing one’s surroundings or anticipating what is coming next in the day. Some students pick up on the relationship between their speech and steps immediately, others do so after some collaborative coaching with another student or the instructor, and a small proportion do not easily find speech rhythm in their feet. For this latter group, it may be more useful to let go of the rhythm framework and to find a different path into the relationship among suprasegmental features of English.

Still, even those students who find that walking does not resonate will sometimes enjoy the silliness of reappropriating objects such as chopsticks to create a rhythmic baseline under speech. A room of 50 individuals clicking 100 chopsticks is a unique “clackety” sound that is as memorable for its humor as for its connection to speech. One activity in the New Ways in Teaching Speaking volume (Bailey & Savage, 1994) proposes developing an “orchestra” by having each student select a unique short set of nonsense syllables to repeat, generating a syncopated, rhythmically complex verbal experience that can illustrate the rhythmic tapestry of possibilities in English—perhaps in all human language.

Listening activities

This variety of possibilities that constitute speech rhythm is exemplified well in listening activities. Students can listen to other speakers or recordings and try to walk or tap at a rate that picks up stressed syllables and pauses. Most National Public Radio (NPR) reporters and interviewees, for example, seem to have a very salient beat to their delivery, across a variety of accents. Students can readily tap or walk along to the regular beats in recordings available on NPR’s website. In order to provide a bridge for such independent listening for relatively regular
beats in speech, it is helpful to find recordings of interest to students, ideally from target communication situations and mark up those recordings as class activities. Recordings from Michigan Corpus of Academic Speech Events (MICASE) are particularly salient for university students. Recordings and accompanying transcripts of classes, lab groups, office hours, speeches, and more are available on the project website (Simpson, Briggs, Ovens & Swales, 2002). MICASE speech patterns are indeed complex, and straightforward illustrations of rhythm are interspersed with fairly irrrhythmic-sounding segments. A number of students have suggested seeking rhythm in situation comedies on television because they enjoy this mode of practicing listening comprehension and increasing everyday vocabulary. While some situation comedies have problematically unrealistic speech patterns (or perhaps realistic only to the situation comedy context), some have conversation that is crafted to sound like everyday unscripted talk, and that does illustrate shared rhythm in interactive speech. An example would be the television series Friends.

**Recorded and transcribed self-introductions**

Working from the speech of others does not sufficiently create a bridge to student speech production. That bridge can be built with student speech models. In classes where students will be introducing themselves to the whole group, the instructor can offer a self-introduction, then record it, create a transcript, and use the text with students to circle words they predict will get strong stress, to underline the stressed syllable in those words, to mark points they the teacher will link words together or elide syllables, and to mark predicted boundaries between thought groups or phrase chunks. Then the class listens to the recording. Students see how well performance matched predictions and discuss the instructor’s hesitations or deviations from the ideal model. The class might do so by tapping out the tempo of stressed elements.

After interest in the teacher’s self-introduction has been exhausted, students are eager to move on to their own. Students use their increased awareness of the features involved in rhythm to create and transcribe recordings of themselves then adjust some features of their pronunciation to arrive at a rhythmic experimentation that gives vocal prominence to the most prominent semantic elements in their self-introduction. Students can work in pairs to serve as extra ears for each other, a pairing that works well (for different reasons) both within and across L1 backgrounds. Students are encouraged to self-assess their awareness of various prominence features and what they are doing differently in their speaking; this process should emphasize and reinforce the learning that is taking place. With self-introductions as models for the techniques of analyzing rhythm, both in listening and speech production, it becomes more straightforward to listen for rhythm in other recordings and to practice rhythm awareness in delivery in a wide variety of speaking contexts. Thus the relationship between use of outside speech models and student speech production is cyclic, not sequential. In recent years, more students have access to video recording equipment to create DVDs of themselves, and students seem particularly adept at describing their speech patterns when they have video footage of their speech behavior.

**Textbook activities**

Pronunciation textbooks offer a third source of material for studying speech rhythm. When working with pronunciation textbooks instead of student-produced speech or recordings of
speech events, a number of activity types are useful. First, textbooks such as Grant (2002) or Miller (2006) provide sets of examples that demonstrate merely that a long word and a short sentence can share the same rhythm and patterns of stress and unstress, a fact often surprising to students. Teachers can adapt this technique and create their own sentences. Approximately rhyming the word and the phrase makes the parallel even more clear (e.g., Ann shouldn’t worry/ambulatory). These phrase and sentence pairs can be read chorally by the class, or can be combined with the kinesthetic activities described above. Another effective activity in several books (e.g., Grant; Miller) starts with a core three-word sentence and adds in function words and low-content modifiers to demonstrate how the distance between the stressed focus words often remains the same. This activity lends itself well to student experimentation with rhythmic speech production.

**Individual attention**

As noted in the prior section on why one would teach pronunciation with a rhythm perspective, each student will encounter different features of spoken English that make it difficult to keep the rhythm going, and this is one diagnostic and formative-assessment advantage of this approach. When students become able to reduce vowels, to link words ending in a dental stop, or to insert a glide to link vowels, they can see, hear, and feel the result in their ability to deliver focus words in a more consistent rhythm. Time with individual learners on these stumbling blocks (turned into learning opportunities) is thus important in realizing the potential for teaching speaking from a rhythm perspective.

**Conclusion**

In sound recordings of real speech events, it often seems that a regular beat does not stand out or is not quite regular or stops and starts. The recent empirical and theoretical challenges to the hypothesis of a stress-timed nature of North American English help to make sense of why this might be the case—some speech probably does not conform to a simple ongoing beat. Nonetheless, it is equally possible to find recorded conversations and monologic speech, both unscripted and scripted, to which one can walk along the sidewalk, with a step on every downbeat, accenting nearly every content-laden stressed word with the sole of the shoe. Impressionistically, at least, there is something to the hypothesis, even if it is on the perception side more than on the production side of speech. More acoustic and intelligibility research will begin to more fully describe the complex system entailed in the distribution of speech sounds.

Pedagogically, what teachers might do for the time being is to offer students a weak version of the stress-timed hypothesis. That is, teachers can share with students that there are several ways to give vocal emphasis to the most content-rich words when speaking, and that sometimes it feels as if these most-emphasized words are equally spaced in time. The stressed-timed hypothesis serves as a useful oversimplification for experimentation. Students can exploit the process of experimentation with periodic rhythm to learn about a variety of features of speech prominence, fluency, grammar and listening comprehension. Beyond a connection among many suprasegmental elements and between speech and other motor skills, experimentation with rhythm invites students to play with how they deliver the language, to become scientists of the language, and to know that, as speakers, they are always making choices about how to deliver
their speech. In time, it is likely that research will reveal how sociolinguistic, discursive, contextual, pragmatic and idiolinguistic variables govern how rhythmic one’s speech is and in what way it is rhythmic, and that students themselves might offer some of these insights before the empirical research evolves enough to create a more accurate picture of the timing of speech sounds.

References


