INVITED TALK

THE RIPPLES OF RHYTHM: IMPLICATIONS FOR ESL INSTRUCTION

Wayne B. Dickerson, University of Illinois at Urbana-Champaign

> The work of Brazil, Coulthard, & Johns (1980), Bolinger (1986), Cauldwell (2002), Wells (2006), and others has led to a growing consensus about spontaneous English phrases: Their rhythm consists predominantly of only one pitch accent (nucleus) or two pitch accents (onset and nucleus) that alternate with unaccented syllables. If we accept these findings, then our pronunciation teaching will differ from our traditional TESOL approach that has been so profoundly shaped by Prator's (1951) version of stress timing. This paper explores the implications of this radically different model of rhythm for the content and presentation of ESL/EFL pronunciation instruction.

INTRODUCTION

Most ESL instructors would agree that we teach pronunciation to help learners communicate intelligibly in oral English. To achieve this goal, many types of language content go into the curriculum. If we consider only the pronunciation part of such a curriculum, as represented by recent pronunciation textbooks, the range of relevant topics is amazingly wide. The prospect of covering so many areas in whatever time is available is truly daunting. In decades of teaching pronunciation, I have lost track of how many times I have dropped, added, and reordered topics in an effort to create a course with the most effective and well-motivated mix of content for my students.

This creative process has led me to the conclusion that there is only one pronunciation topic that rises to the top of a list of priorities because of its centrality to the goal of helping learners communicate intelligibly in oral English. The topic is rhythm. I give it top billing because I believe it alone has the potential to organize the content of a pronunciation course in a way that keeps instructors' and learners' attention on the stated goal of intelligible oral English.

To support this claim about the potential of rhythm, I start by identifying what I mean by rhythm and why it is pivotal to any effort to improve the intelligibility of learners' oral communication. We will then be in a position to consider the implications of adopting and promoting this model: What have we been doing well in our teaching of rhythm, and where and why have we missed opportunities to offer important guidance to our students?

THE TWO-PEAK PROFILE

Lucy Pickering (2018) summarizes David Brazil's model of the rhythm found in spontaneous English:

> It can be useful to think of the typical structure of a tone unit in English as comprising three to seven words and containing one or two prominences. (p. 23)

Pickering's summary applies equally to the conclusions of other researchers like Kenneth Pike (Fries 1945, p. 64), Dwight Bolinger (1961, p. 135), Richard Cauldwell (2013, p. 39), and John Wells (2006, p. 192), whose work attempts to characterize the language found in naturally occurring talk.

This summary does not describe stress-timed rhythm, a linguistic hypothesis about how English rhythm works that, when repeatedly tested by phoneticians, was found to be universally unsupported (Dauer, 1983; Roach, 1982). (See Cauldwell (2002) for an overview of the topic and Arvaniti (2012, p. 351ff) for a detailed examination of the many studies that tested the hypothesis.)

Clifford Prator, Jr. (1951) introduced the model of stress timing to ESL/EFL instructors through his ESL pronunciation textbook. However, he did so with a simplification of his own that was foreign to the original linguistic hypothesis. He taught his students to use a pitch accent on **every** content word in a phrase. It is this now-discredited model in a distorted form that has become the de facto standard in our language-teaching field and now appears in almost every ESL/EFL pronunciation text, teacher's guide, and teacher-preparation course (Dickerson, 2015).

Given the centrality of rhythm to intelligible oral communication, as argued in this paper, there could not be a worse place in phonology to distort the linguistic facts. Since rhythm is a feature of every spoken phrase, stress-timed rhythm yields a form of oral English that no one speaks natively. For more than half a century, with no malice of purpose—indeed with every good intention, our profession has nevertheless been teaching students around the world a non-English (and often unrecognizable) way to communicate. ESL/EFL learners who have acquired an ability to speak English well have done so despite the instruction they have had in rhythm.

By contrast, the model of rhythm advanced here is one that has evolved partly from the fieldwork of phoneticians examining spontaneous speech. It reflects the actual rhythm of English. Just as rhythm is the worst place in phonology to distort the linguistic facts about so basic a feature, it is also the best place in phonology to get the facts right for the sake of our students and their communicative effectiveness.

To elaborate on this model of rhythm in spontaneous speech, I want to introduce the pedagogical terminology I use when teaching ESL/EFL students (and will use in this paper). To start with, since the model had no name—no counterpart to "stress-timed rhythm," I drew on the metaphor of a mountain range in profile and called it the **two-peak profile.** The name emphasizes peak-valley alternation, not timing, as the fundamental nature of English rhythm.
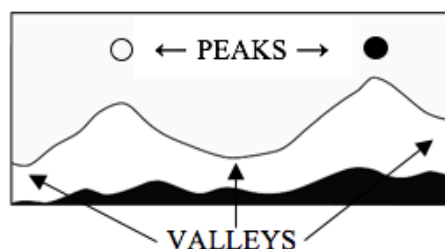
In our pronunciation materials, we refer to the second pitch accent as the **primary peak,** or simply the **primary**. It is commonly referred to as the **tonic** or the **nucleus** in linguistic literature. We mark this peak with a filled bullet over the accented vowel. The first pitch accent, if there is one,

is the **anchor peak**, or simply the **anchor**. In phonetics, it is known as the **onset** or **onset syllable** (Wells, 2006, p. 8; Tench, 1996, p. 135). We mark this peak with a hollow bullet over the accented vowel. Following are examples of the two-peak profile from speech samples.

(1) Have you had lunch yet?   (3) Usually he won't comment on them.

(2) Twenty twenty's a leap year.   (4) I bought these peaches in Savannah yesterday.

Before, between, and after these peaks are what we call **valleys**—syllables that carry no pitch prominence. Sometimes, to draw special attention to valleys, we mark them with subscripted swooshes, as shown above. But even without swooshes, all non-peak syllables are deaccented. There are potentially five positions in the two-peak profile: two peaks and three valleys. While there must be a primary peak, all other parts are optional. For instance, example (1) has no anchor nor medial valley. Example (3) is missing an initial valley. Examples (2) and (4) have all five parts.

Although the two-peak profile has a long history in phonetics (see Pike in Fries, 1945, p. 64; Bolinger (1961, p. 135), we think of it as new because it is largely unknown in TESOL circles. But not entirely so. It first appeared in David Brazil's pedagogical materials published in 1994. More recently, Gorsuch, Meyers, Pickering, and Griffee built on Brazil's work in their *English Communication for International Teaching Assistants* (2013). These are pioneering efforts, but resistance to an alternative model of rhythm is formidable; the old model of stress-timed rhythm remains strongly entrenched in our field. Furthermore, over the past three decades, it has been hard to get excited about a model for which no full rule had emerged to predict the first peak.

Is there a rule for where the anchor occurs? We have rules for nearly all of phonology—for placing the primary peak, for how to stress multiword numbers, for how final consonant clusters are simplified, for why an intonation pattern rises or falls at the end of a phrase, and so on (Dickerson, 2004). Given that language is profoundly regular, a rule must also exist for the anchor. A stronger case for assuming that an anchor-placement rule exists is that native speakers of English behave in a rule-governed way: They can tell when an anchor is used neutrally—without drawing special attention—and when it is used emphatically (Dickerson, 2015).

The need for this rule is felt acutely by adult learners who want to take responsibility for their learning and do not want to depend on teachers as crutches for where to put pitch accents (Dickerson, 2004). They need an anchor-placement rule to use in their private self-practice to check and correct their own accuracy just as they need other pronunciation rules. Furthermore, given how commonly we use emphasis in everyday language, learners also need the rule by which to determine when a co-speaker is speaking non-emphatically and emphatically. These learners also need the rule in order to know where to place anchors with the meaning they intend.

To address this felt need and to encourage greater interest in and adoption of the two-peak profile, in 2013 my co-author, Laura Hahn, and I set our sights on working out the rule for the dominant uses of the anchor, translating it into a pedagogical form, and trialing it with classes of university-level students and international teaching assistants. The second edition of our textbook, *Speechcraft: Discourse Pronunciation for Academic Communication* (in press, University of Michigan Press) is built around the two-peak profile and accent-placement rules for the primary and anchor peaks.

THE LINCHPIN LINKING SOUND TO MEANING

Intelligible spontaneous speech—the goal of pronunciation teaching—is speech that the **listener** can understand in the rapid give-and-take of conversation. Since the listener is the judge of what is intelligible, what must a spoken phrase contain to make it so? The answer that research gives is that the phonetic cues to the core message—the **focus**—must stand out to the listener (Wells, 2006, p. 234). The listener, in turn, must know how to decipher the phonetic cues to discover the focus if the communication is to be successful.

Focus is a term that has found its way into many pronunciation textbooks (Gilbert, 2012; Gorsuch, et al. 2013; Grant, 2017). Originally synonymous with the word carrying the primary peak, it meant what is new or in contrast (Ladd, 1996, pp. 225ff). Although we now understand that English rhythm has one or two peaks per phrase, the concept of focus is still tethered to peaks. In the case of two peaks, it now starts with the cue in the anchor word and ends with the cue in the primary word. In the case of one peak, the cue for the start and end of the focus is in the same primary peak. To refer to both cases, Wells uses the term *focus domain* (2006, p. 116), and Brazil et al. prefer the term *tonic segment* (1980, pp. 39-42).

To foreground the focus of a phrase for the listener, the speaker uses pitch accents—the peaks— of rhythm to identify the words and constructions that are most germane to the message. The focus itself is not in the phrase but is instead in the pragmatic interpretation of the phonetic cues in the phrase.

<div align="center">

●   ○   ●

(1) Have you had lunch yet?   (2) Twenty twenty's a leap year.

</div>

For example, the speaker's peak in (1) should tell the listener that the heart of the question is *lunch,* not the suprasegmentally enhanced vowel in *lunch.* Similarly, in (2), the speaker's pitch accents on the main vowels in the second *twenty* and in *leap* should lead the listener to interpret the focus as the multiword number *twenty twenty* and the compound noun *leap year.*

I find the current definition of the focus, namely, the cue(s) in the anchor-word-through-the-primary-word string, particularly valuable conceptually and pedagogically because it identifies the part of a phrase that the listener must understand in order for the phrase to be intelligible. This string allows us to describe the essence of the message better than using the primary peak alone. In fact, for pedagogical purposes, we prefer to call this string the **essence** or **semantic essence** of the phrase rather than the focus.

Since noticeable pitch accents are essential to a listener's grasp of the speaker's message, they should also be essential to what we teach learners to produce and interpret. That is, the objective of any course that aspires to help learners communicate intelligibly in spoken English should be the clear production of the two-peak profile and the facile discovery of the focus in that profile.

The reason we assign this rhythmic profile the highest priority in pronunciation instruction is its facilitative role in communication and in teaching the sound system. First, its peaks highlight the part of a phrase—the focus—that most directly signals the speaker's message and that is most relevant to the listener's understanding of that message. Second, and equally important, rhythm is the gravitational center of the English sound system; the entire phonology is concentrated on

implementing that rhythm. In fact, much of phonology is dependent on the precedence of rhythm. For example, as critically important as intonation is to the intelligibility of the focus, its pitches hang on the peaks of rhythm.

IMPLICATIONS FOR INSTRUCTION: PERSPECTIVE

What are the consequences that flow from adopting the two-peak profile for our ESL/EFL instruction? Some consequences affect how we present pronunciation instruction to learners.

1. **A well-motivated starting point.** If the reason we teach pronunciation is to help learners make their phrases easy for listeners to understand by meeting their expectations, then instructors should state this objective at the start of each semester. Furthermore, if the way to reach this objective is to teach learners to pronounce what listeners are listening for, namely, the focus signaled in the two-peak profile, then it stands to reason that instruction on the two-peak profile must start each semester, as well.

Rhythm as a starting point is novel to most learners because it has not been part of their prior ESL/EFL learning. Rhythm also has a nebulous quality because it does not seem so concrete as other parts of the sound system, like consonant and vowel segments or lexical stress. Even intonation seems easier to grasp than rhythm. Rhythm is all the more startling because the claims we make for it are so consequential, namely, its centrality to the intelligibility of every spoken utterance. How could something so relatively unknown in learners' experience with English and so seemingly intangible really be so important? There is shock value in this starting point that, for many, is actually intriguing. It says: "This instruction will be different from what you may have expected. Nevertheless it will turn out to be really important to your success with English."

2. **A consistent, unifying rationale for instruction.** An overarching benefit of recognizing that all phonology works toward the single objective of creating noticeable pitch accents is that, as we take up each new topic, we can relate the relevance of the new topic to the master objective of creating a rhythm that communicates the focus of the message. By returning again and again to rhythm as the point of reference for every new topic, we create a thematic unity of subject matter that would otherwise seem like disparate, unrelated parts of the sound system.

While some students accept each topic with equanimity as it comes along, we have found that most learners feel decidedly better about studying a topic when they know how it affects their personal communicative effectiveness. Instructors, too, have a stronger sense of mission when they articulate the connection between what they teach in each lesson and the goal of their instruction, namely, intelligible communication achieved through an effective rhythm.

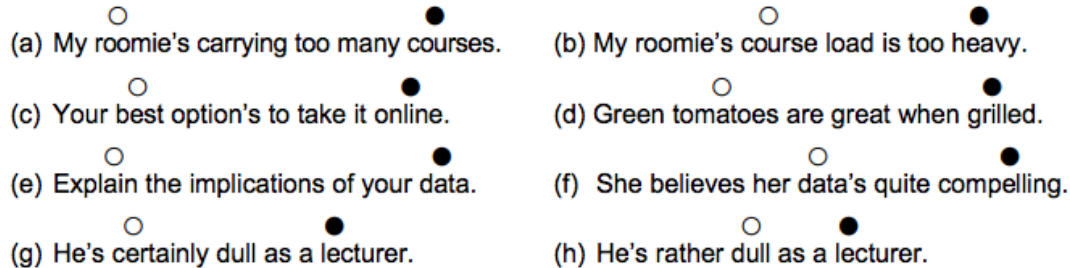## IMPLICATIONS FOR INSTRUCTION: WHAT, WHEN, AND HOW WE TEACH

Other implications for teaching the two-peak profile go beyond orientation and motivation. They entail adding and reorganizing content and teaching with new techniques and emphases. What have we done well in the area of rhythm, and where are the gaps we could fill to prepare our students even better to use the two-peak profile in their interactions with English speakers?

3. The role of grammar in making pronunciation decisions. The terms **content words** and **function words** (Pike, 1945, p. 118) are staples in discussions of pronunciation. Since these categories are defined by parts of speech, they tell us that grammar is fundamental to pronunciation choices. This should come as no surprise. Since language is for communicating meaning, both grammar and pronunciation point toward the same goal. Grammar does it by arranging parts of speech; pronunciation does it by arranging sounds, stresses, and pitch.

The traditional, although tacit, assumption about the convergence of grammar and pronunciation is that the two map onto each other so well that basic parts of speech are adequate to describe and predict how the sound system works. ESL/EFL learners' background in grammar should therefore be a satisfactory preparation for pronunciation work; it should be enough that students can recognize, for example, content words as nouns, adjectives, verbs, and adverbs.

OLD (and inadequate) goal: Use general categories of grammar to predict pronunciation.

The reality is that the match between parts of speech and pronunciation choices, while close, is not one-to-one. There is no part of speech that pitch accents prefer categorically. As examples, look at a sampling of anchor placements among content words. In (a) the anchor is on the first noun, but in (b) it is not; the first noun is in a valley. In (c) the anchor is on the first adjective, but in (d) it is not; the first adjective is in a valley. In (e) the anchor is on the first verb, but in (f) it is not; the first verb is in a valley. In (g) the anchor is on the first adverb, but in (h) it is not; the first adverb is in a valley. An ability to recognize general parts of speech is, unfortunately, not sufficient to understand nor to predict the behavior of the anchor peak.

○ ●	○ ●
(a) My roomie's carrying too many courses.	(b) My roomie's course load is too heavy.
○ ●	○ ●
(c) Your best option's to take it online.	(d) Green tomatoes are great when grilled.
○ ●	○ ●
(e) Explain the implications of your data.	(f) She believes her data's quite compelling.
○ ●	○ ●
(g) He's certainly dull as a lecturer.	(h) He's rather dull as a lecturer.

Often the grammatical function of a part of speech is germane to a pronunciation choice. For example, the subject noun carries the anchor in (a), but the possessive noun in (b) does not. This point leads us to a more adequate goal for preparing learners for pronunciation work.

NEW goal: Teach parts of speech with enough detail to make good pitch-accent decisions.

Accordingly, in our instruction, we ask students early in the semester to review parts of speech and their functions. It is not necessary to brush up on all of English grammar when the review is specifically tailored to pronunciation needs. This review (an appendix in *Speechcraft*) is done out of class as self-study and followed up with an assessment of learners' grammatical readiness for the study of pronunciation.

Not only do students perform better after doing a part-of-speech review, but the targeted study also helps to level the playing field for class members. Those whose formal exposure to English

grammar is only marginally adequate are not disadvantaged by their weak background, and even those who have a good grasp of English grammar refine their understanding.

4. An identifiable language style with its own characteristics. Few pronunciation texts are explicit in identifying the target style of English they present. Even fewer attempt to teach the features of a particular style. Perhaps the assumption is that details of style do not matter as long as the language taught is educated speech.

OLD (and inadequate) practice: Teach English pronunciation without specifying a style and its features.

The problem comes down to one's definition of educated speech. Judging from current pronunciation texts, educated oral English is not dominated by the two-peak profile. It puts no constraints on phrase length. Nor does it require compression devices in production. This is not what listeners wish to hear when spoken to. Their choice of style is the one they themselves make when speaking, namely, spontaneous educated speech. This style is characterized by (a) the familiar two-peak profile, (b) short phrases, and (c) an abundance of valley compression.

As we are coming to understand, this trio of spontaneous-speech features promotes intelligibility. Early sections of this paper described the centrality of the two-peak profile to identify for listeners cues to the beginning and end of the focus. The rhythmic profile, however, does not perform its pragmatic function alone; it requires the help of two other features of spontaneous speech— constraints on phrase length and pervasive compression phenomena—both of which address the same issue, namely, the limitations of the listener's memory.

The listener's goal is to snatch the essence of a phrase out of the air as the speaker talks. The target of the listener's attention is everything from the anchor word through the primary word, including any in-between valley words. It is the in-between stretch that can tax memory because it makes a difference to the listener how much time the speaker takes to get from the anchor word to the primary word. For the sake of memory, the shorter the better, lest the listener forget the first peak before the second arrives, slowing comprehension (Kjellin, 1999; Munro & Derwing, 1998). Anything that shortens inter-peak time benefits memory.

A **short phrase length** helps memory because the closer the anchor and primary peaks are to each other, the quicker the focus can be revealed to the listener. **Valley compression** also helps memory because it directly minimizes time spent in the valley between the anchor and primary peaks. Each of these features deserves attention. Let us consider first how we help learners create well-sized phrases. The topic of compression comes in the following section.

Although English grammar does not limit the length of a spoken phrase, cognitive factors like the capacity of speakers to take in, and listeners to process, speech on the fly do impose limits. The phrase-length solution to the problem of retaining a memory of the whole focus is to keep the anchor and primary peaks close together by admitting at most only five words between them. When the maximum number of words in a phrase is held to seven, inter-peak time is controlled.

While limits on phrase length come naturally when speaking spontaneously, limits are not automatic for learners who are practicing with a written text. Unaware of how short spoken

language phrases really are, learners need guidance. We offer it to them by starting with an example of spontaneous speech to make three points. First, phrases are typically no longer than seven words and average only three to four words. Second, pauses between phrases occur at identifiable places that can be discovered by examining spontaneous speech. Third, to speak intelligibly, learners must practice using phrase-boundary guidelines to convert written language into phrases that simulate spontaneous speech. These three points are not customarily part of ESL/EFL pronunciation instruction despite their being necessary to intelligible speech.

To model appropriately sized phrases, all instructional materials should conform to constraints on phrase length so that students continue to learn by example. Then, whenever they compose materials to be spoken, they will be more likely to remember to adhere to good models and to practice written-to-spoken conversion techniques. The point we emphasize is that when spoken phrases are sized correctly, the two-peak profile fits more comfortably on them, and listeners get the speaker's point more readily, than when phrases are overly long and strain listeners' memory and processing capabilities.

NEW goal: Adopt spontaneous speech as the oral target; teach written-to-spoken conversion techniques to simulate spontaneous phrases; practice the two-peak profile with short phrases.

Of course not all talk is spontaneous, particularly in academic and professional settings where presenting papers and proposals is expected. I strongly believe that the best preparation for formal speaking is learning to converse well in everyday interactions. The characteristics of extemporaneous speaking are basic to good public speaking—good peak-valley contrasts in the two-peak profile on short phrases that exhibit comfortable compression in their valleys.

5. Natural Speech Phenomena. Spontaneous speech is characterized not only by the two-peak profile and by short phrases but also by the pervasive use of compression devices (assimilation, trimming, reduction, and linking). These devices provide another solution to the listener's challenge of catching and remembering the whole focus of a phrase in the rapid flow of speech. By miniaturizing each valley syllable to the extent possible, the speaker can quickly deliver to the listener cues to the focus in a single continuous string. In turn, the listener is more likely to recognize the focus and take it in fully when it comes in the context of a familiar rhythm used in short phrases (Kjellin, 1999, pp. 23f).

In the repertoire of pronunciation topics, natural speech phenomena were among the last to be added to pronunciation textbooks. Often relegated to a lesson late in the text, the emphasis was on the compression of function words (Dale & Poms, 1994). The mistaken impression, arising from stress timing, is that all content words will carry a pitch accent and therefore require no compression.

From the perspective of the two-peak profile, we now understand that every word that can contribute a peak to the two-peak profile can also contribute its syllables to valleys. Even so, compressed speech has not yet achieved parity with other pronunciation topics. Two issues stand in the way.

First, a widely accepted conclusion about natural speech phenomena is that we must teach these devices mainly for purposes of perception (Levis, 2018, pp. 148-149). Learners' struggle to

understand an incoming speech stream that has been reshaped by compression devices is painfully obvious, and research has confirmed their problem. But to downplay learners' responsibility to produce such a speech stream for listeners is to say that the intelligibility of the learners' speech is not important to listeners. This is the opposite of the goal of pronunciation instruction, as stated in the first point above, namely, to help learners make their phrases easy for listeners to understand by meeting listeners' expectations.

OLD (but incomplete) goal: Teach learners to understand compressed speech but not to produce it.

When we elevate learners' difficulty to understand a speaker's compressed speech but dismiss listeners' frustration trying to understand learners' uncompressed speech, we expose our bias and belie our belief that intelligibility is a reciprocal requirement for speakers and listeners alike.

Second, although natural speech devices are part of all modern pronunciation texts (Miller, 2006), they are typically taught as discrete topics to improve the naturalness and flow of speech, not as an integral part of a coordinated effort to make it possible for the listener to grasp the focus in one take. A better way to approach these devices is to understand them as contributing to the listener's success at understanding the core meaning of a phrase. This means introducing these devices together as soon as learners begin producing the two-peak profile, namely, early in instruction. From a practical point of view, how does it help learners to use the rhythm model, introduced early in instruction, if consonant and vowel compression comes late in instruction or is scattered among learners' pronunciation lessons?

NEW goal: Teach consonant and vowel compression techniques in connection with the two-peak profile as equally important for perception and production.

6.   Rhythm as a pacing tool. A myth about English that we hear repeatedly from our students is that fluent speech is fast speech. This assumption leads some to try to say everything as quickly as they can. Despite the frequency of this error, it is rare that pronunciation texts comment on pacing one's oral delivery except to urge students to hurry their articulation of function words.

OLD (and inadequate) practice: Speed up function words; slow down content words.

In point of fact, every language is delivered at a range of speeds (Cauldwell, 2013, pp. 94ff). The belief that a target language is spoken especially fast reflects more the learners' inability to process it as quickly as it is spoken than it does the actual articulation speed of the new language.

A more helpful rebuttal to the myth of speed is this fact: No single speed is appropriate for all situations, nor is it appropriate for all parts of an English phrase. Peak syllables sound slow when they are stretched out, and valleys syllables sound fast when they are compressed. Hurrying and slowing are rhythm skills. They need to be practiced in conjunction with the two-peak profile because it serves as a traffic signal for when to speed up and when to slow down in a phrase. When the signal is obeyed, each phrase is delivered with speeds expected of intelligible speech.

**NEW goal:** Include in a description of the two-peak profile its use to control the speed of speech.

7. Switching patterns: The Heart of the Two-Peak Profile. A two-peak profile has, at most, two valley-to-peak changes in prominence (first-valley-to-anchor and middle-valley-to-primary) and two peak-to-valley changes in prominence (anchor-to-middle-valley and primary-to-last-valley). We call these **switching patterns** because they are sudden flips in prominence that often happen from one syllable to the next. Without these two rapid switches—peak-to-valley and valley-to-peak—there would be no two-peak profile.

In general, this point sounds like something from stress-timed rhythm instruction: For decades, pronunciation teachers have encouraged their students to put peaks on content words and push function words into valleys. The idea of contrast is not new, nor is it wrong.

OLD (inadequate) goal: Contrast peaks (on all content words) with valleys (on function words).

However in two respects, making a difference between content words and function words is not enough for the two-peak profile. Roger Kingdon (1958, p. 160) identifies one issue: "The difference in prominence between stressed and unstressed syllables is greater in English than in many languages." That is, many students will find that the degree of peak-valley difference needed for English switching patterns lies outside their customary range. When speaking English, their peaks must be more distinct from their valleys in pitch, duration, and intensity.

The second issue becomes clear with some actual examples. Here are two valley-peak switches:
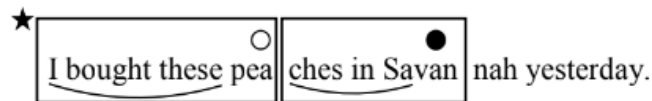
And here are two peak-valley switches:
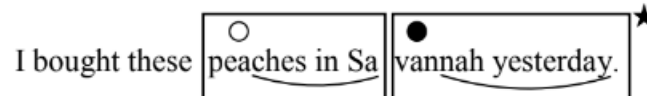
Learners accustomed to stress-timed rhythm phrases may be startled to see **whole** content words in valleys, such as the words *bought* and *yesterday* (see the stars). The vowel in each syllable of these content words must be suppressed.

NEW goal: Monitor the size of contrasts and use of valley vowels for content words in valleys.

An attention to the magnitude of switches and to the suppression of content-word vowels is still not enough to prepare learners to use vowels fully to implement a rhythm that communicates clearly to listeners. The story of vowels continues.

8. The Prosody of Vowels. We have noted the traditional emphasis on developing learners' skill at distinguishing phonemes such as /iy/ and /ɪ/ so that listeners can hear the difference between words like *reach* and *rich, cheap* and *chip.* We have also highlighted the importance of including work on vowel compression to reduce the time of valleys. Our focus in both cases is on guiding learners' control of the **articulatory** features in the lower half of the vowel circle below.

PROSODY

Kingdon's (1958, p. 160) caution about the need for adequate contrast between peaks and valleys points us to another area of vowel work that is generally missing from pronunciation instruction. When we contrast a peak with a valley and vice versa, what part of a syllable do we change? The answer is the **vowel** part. But what aspect of vowels do we change when we switch from a peak to a valley and vice versa? We do not manipulate their articulatory features but their **prosodic** features—those in the upper half of the vowel circle.

Contrast, as in a switching pattern, is the exclusive domain of vowel prosody. Whether we use the two switching patterns to create word rhythms around a single peak (peak-valley, valley-peak-valley, or valley-peak) or phrase rhythms around potentially two peaks, we do so by adjusting the duration, pitch, and intensity of vowels. Although they create different rhythm patterns, switching patterns at these two levels of structure are identical. This is why it makes sense to teach switching patterns at the word level; they transfer directly to the phrase level. That is, a peak at the word level can create one of the peaks at the phrase level. Of course, there are no switching patterns in words found in phrase-level valleys because there are no peaks there.

Since prosody-based switching patterns using vowels are the means by which a speaker tells a listener where the focus of the phrase begins and ends, I was obliged to adjust my goal for vowels once more, this time to include vowel prosody.

NEW goal: Teach vowel articulation and v*owel prosody* together*.*

Since manipulating prosody is new to vowel instruction, some example exercise types may be helpful. As noted, the place to begin to meet this expanded goal is at the word level. We work on the **perception** of prosodic vowel differences first by asking learners to listen to a pair of vowels in adjacent syllables. Here is an exercise on the diphthong in *now.* Of course learners are naturally drawn to the most noticeable syllable. To encourage them to pay attention to the full range of vowel

prosody, we ask them to pick out the least noticeable syllable. (More items than these would be used with learners, of course.)

I'll say a word twice. Identify its <u>quietest</u> syllable by saying 1st or 2nd. For example, if I say **óutbòund**, **óutbòund,** you would say **2nd, for the second syllable.** But if I say **bòw dówn, bòw dówn,** you would say **1st, for the first syllable.**

óutbòund,	2nd		òutshóut,	1st
bòw dówn,	1st		tównhòuse,	2nd
lóudmòuth,	2nd		fòuled óut,	1st
chòw dówn,	1st		cóuntdòwn,	2nd

Next we work on the **production** of prosodic vowel differences at the word level. The focus is on pronouncing an adequate difference of the same vowel. If the target vowel is the same as above, then we can use items from the perception exercise for production purposes. We monitor for articulation and for prosodic contrast. For example, is the first vowel in *loudmouth* sufficiently different in its prosodic features from the second vowel to register as a peak-valley contrast?

Each time we introduce a vowel, we include perception and production exercises that offer learners practice identifying and modulating vowel prosody.

From isolated words and constructions (e.g., *cóuntdòwn, fòund óut*) we move to contextualized phrases but still contrasting the same vowels. Here is an excerpt from a dialogue about a student who was in a traffic accident (Dickerson & Hahn, in press). It illustrates how peak-valley and valley-peak switching patterns using vowel suprasegmentals fit into and promote the two-peak profile. Two instructors are talking.

A: Did you hear about Maria? She was in a traffic accident.

B: I'm sorry to hear that. What happened? How is she?

A: She was walking along Peach Street, | composing a text message, | and strayed into a bike lane. She collided with a cyclist | and came home with a few bruises | but no broken bones. That was three weeks ago.

Instead of starting with this dialogue as an interaction, we start with word pairs from the dialogue where the same vowel is used in both words, creating a peak-valley switch and a valley-peak switch between a primary-peak vowel and a lesser-stressed vowel. That is, we again hold the vowel constant so students can focus their attention on the change in prosody.

| **PEAK-VALLEY CONTRASTS**	**VALLEY-PEAK CONTRASTS**
●	○ ●
[She was in a] traffic accident /æ - æ/	[came home with a] few bruises /uw - uw/
○ ●	○ ●
[She was walking along] Peach Street /iy - iy/	[but no] broken bones /ow - ow/
○ ●	○ ●
[composing a] text message /ɛ - ɛ/	[That was] three weeks [ago] /iy - iy/

After this kind of practice, students use the dialogue itself. Contrasts at the level of two-word phrases are the kind that can lead to a clear rendition of the two-peak profile in the dialogue.[i]

9. Contrasts with Word Stress. Contrasts in vowel prosody are, of course, also contrasts in word stress. However, by starting with vowels in words, a smaller context, we can isolate prosodic issues with the components of stress—pitch, intensity, and duration—and deal with them more easily than we can with word stress in phrases, a larger context. This is the ideal environment to work with those students whose native languages have left them less sensitive or insensitive to the prosodic features that are relevant to English stress (Levis, 2018, pp. 109-114).

Before continuing with word stress, let us change the notation system from pedagogical bullets to conventional stress marks, which allows us a more refined look at word stress. By using the quaternary stress-marking system (´ ˆ ` ˘), we can represent the stress of the two-peak profile as well as the stress of words. Since there are different stress-marking systems, note how the quaternary stress marks are used instead of bullets.

PRIMARY PEAK   ANCHOR PEAK   VALLEY STRESSES
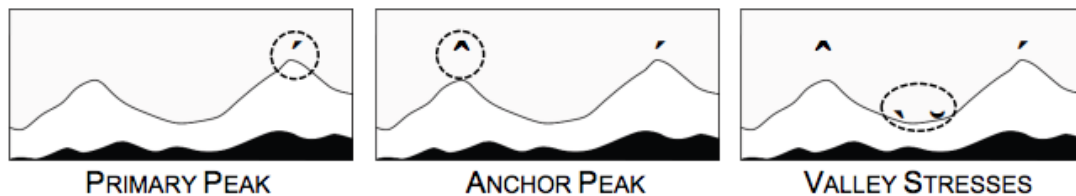
For the two-peak profile, the acute stress mark is used for the vowel that carries the primary peak, as *lunch* does in *Have you had lúnch yet?* The circumflex stress mark is for the vowel that carries the anchor peak, like the first vowel of *usually* in *Ûsually he won't cómment on them.* These are the two levels we call **maximized** vowels. Valley syllables are also of two kinds. Those that have a lightly stressed vowel are marked with a grave stress mark, and those that have an unstressed vowel are marked with a breve stress mark. Lightly stressed vowels are those in the words *Ì, bòught,* and *thèse,* and unstressed vowels are those in *-chĕs ĭn Săv-* and *-ă* in *Ì bòught thèse pêachĕs ĭn Săvánnă.* These are the two levels of **minimized** vowels.

For the citation form of a word, as found in a dictionary, only three levels of stress are needed. Since a multisyllabic word has only one heavy stress, there is a primary peak but nothing comparable to the anchor peak at the word level; no circumflex stress mark is used at this level. All non-primary peak syllables are valley syllables, containing either lightly stressed vowels indicated with a grave stress mark or unstressed vowels indicated with a breve stress mark.

Resuming our topic of word-stress, what do we teach when we teach word stress? The typical answer to this question is that we teach the **position** of the main stress in a word. For example, is the main stress of *fantastic* on the first syllable, the middle syllable, or the last syllable? If my students are able to hear, predict, and produce the heavy stress of a word on the right syllable, I applaud their success. They have accomplished something important that transfers directly to one aspect of rhythm: The primary peak is positioned on the main stress of a multisyllabic word. So the position of the main stress in a word is a worthy part of what we want our students to learn. It is also content typically found in modern pronunciation texts.

OLD (but incomplete) goal: Teach the position of the main stress in a word.

However, when examined from the perspective of our end goal of helping students communicate intelligibly, this objective is incomplete. It says nothing about how the focus of a phrase is signaled in rhythm. By stopping with the position of the main stress, I missed another connection between word stress and phrase rhythm. In effect, I was no better than a dictionary.

Dictionaries tell learners that the main stress of *popularity* is on the third syllable. In citation form, that stress is marked as a primary stress—*populárity*—because the citation form of a word is a one-word phrase with a primary peak on the main stress. Is the primary stress appropriate everywhere in the two-peak profile? Of course not. It is appropriate when the word carries the primary peak, but it is not appropriate for the anchor because we drop the level of stress for the anchor. A primary stress is certainly not appropriate for a word in a valley, where there is no pitch prominence. The main stress becomes a deaccented—minimized—vowel in a valley.

The citation form of a dictionary word does nothing to help students with the **level** of the main stress in the two-peak profile. Finding only the position of main stress in a word is the weakness of the old goal for word-stress work. A more adequate goal with respect to word stress must also include giving students practice modifying the level of stress in a word to match its place in the two-peak profile.

NEW goal: Teach the position of word stress and the *flexibility to adjust the level of word stress* at this position according to the role the word plays in rhythm.

The limitations of the old goal reflect the limitations of the old model of rhythm in which every content word in a phrase receives a heavy stress and no content word would be found entirely in a valley. In the actual rhythm of spontaneous speech, where there are usually only two maximized vowels in a phrase, the stress of all other content words is pervasively minimized. Reality at the phrase level is that the stress of any word must be flexible enough to fit anywhere in the two-peak profile. So how do we prepare learners to be flexible with word stress?

First we teach learners how to predict the peaks (and valleys) of the two-peak profile. That is because the rhythm profile is the template that tells learners the level of stress to use for all multisyllabic words in a phrase. Then comes practice that targets the skill of adjusting stress levels appropriately. When the prosody of vowels was the target, our first step in practice was to hold the vowel sound constant and vary the prosody of this one vowel. Now that word stress is the target, we hold the word constant and vary the location of this one word in the two-peak profile.

One exercise type from *Speechcraft* puts the same word in two different positions. The target word is *good*. Students fill in the blanks to describe themselves. Then they read the three lines. Their practice contrasts a maximized vowel in *good* (anchor) with a minimized vowel in *good* (valley).

Describe yourself by filling in the blanks with some of the words below. Then read the three lines aloud. Focus on contrasting the stress of the word good.

I'm *good* at ___chitchat___ . But I'm not so *good* at ___listening___ , | and I don't want to try.

I'm good at ___languages___ . But I'm not so good at ___math___ , | and I don't want to try.

I'm *good* at _____ . But I'm not so *good* at _____ , | and I don't want to try.

math	science	sports	research	writing
organizing	music	chitchat	drawing	listening
languages	soccer	dancing	cooking	teaching

Another exercise type, called "Adjusting Stresses to Match the Two-Peak Profile," puts the same words in all three possible positions of the two-peak profile to practice three levels of stress—for the anchor peak, the primary peak, and a valley. After identifying the target words and constructions, learners practice adjusting their stresses until they can do so smoothly. Here an instructor is cautioning a class about getting the most from their literature review.

Highlight the words and constructions that are repeated with three levels of main stress. Listen to the dialogue and practice it until you can easily adjust the level of stress in the flow of the discourse.

Situation: An instructor's caution

Here's a fact about footnotes: When reading research, | footnotes don't seem important |

because they are footnotes. But don't ignore them. Sometimes one that doesn't appear

important | contains a fact | that's important to your work, | a fact you've been looking for.

The motivation for asking learners to identify words with three levels of stress (anchor, primary, and destressed valley) is to engage them in something like a treasure hunt. With their understanding of the two-peak profile, the task is not difficult. Yet it draws their attention to the different stress requirements of the rhythm. In this monologue, learners find the words *fact, footnote,* and *important,* each in the three stress positions of the two-peak profile. Exercises of this kind, with supporting audio, develop learners' flexibility to vary the stress of a word in a phrase according its role in the two-peak profile.

Work on adjusting the level of stress is a word-stress topic that connects stress levels directly to

the two-peak profile. Stress level, not stress position, signals the start, middle, and end of the focus of a phrase. The contrasting peaks and valleys of the focus must stand out clearly if listeners are going to find the phrase intelligible.

10. Pitch and its Patterns. The role of pitch that is most familiar to ESL/EFL instructors is in pitch patterns or intonation patterns. These patterns are important because they tell us how to interpret what we hear in the focus of a phrase: Is this core meaning a question or a statement? Is it finished or is there more to come? Is it friendly, businesslike, neutral, or insistent? Is it a true question or a request for confirmation? However, intonation patterns are a secondary role of pitch, not its primary role, and therefore not entirely adequate to promote the two-peak profile.

OLD (and incomplete) goal: Teach the pitch of pitch patterns.

The primary role of pitch, along with intensity and duration, is to highlight for listeners the cues that mark the beginning and end of the focus. We recognize these peaks best as listeners, not because speakers maximize the pitch, intensity, and duration of peak vowels, but because they maximize the **difference** between peak vowels and nearby valley vowels. These acoustic markers are the yin and yang (☯) of rhythm.

When it comes to pitch accents, we tend to concentrate more on maximizing the suprasegmentals, including the pitch, of their accented vowels than on minimizing the suprasegmentals of adjacent unaccented vowels. To restore the right balance, we must remember that both ends of the suprasegmental continua deserve equal attention in order to create the difference between these extremes that English listeners expect.

NEW goal: Prioritize the focus first by preserving its yin-yang balance: Teach peaks by contrasting them with valleys. Then teach pitch patterns to help listeners interpret the focus.

If peaks do not stand out, what is the point of an intonation pattern to a listener who does not know what the core message is?

CONCLUSION

This paper began with the assumption that the purpose of oral language is to communicate meaning. If this assumption is accurate, then how does oral language advance this purpose? For English, the answer seems to be that the peaks and valleys of its phrase rhythm identify for listeners the basic pragmatic structure called the focus of the phrase, its core meaning.

This answer says, in effect, that phrase rhythm is the most fundamental linguistic pattern in oral English because of its centrality to intelligible communication. Furthermore, in one way or another, the rest of phonology works to create and build on this rhythm for the ultimate purpose of communicating meaning. That is, all of phonology, including rhythm, is driving toward the goal of making intelligible oral communication possible.

Arriving at these conclusions was the result of taking a fresh look at my assumptions. The fact is that, like many ESL/EFL instructors, I learned my stress-timed-rhythm lessons so well that they created blind spots in my understanding of the communicative role of rhythm.

However, the real turning point for me was the exercise of trying to teach the two-peak profile to my ESL students. It forced me to reexamine everything I was saying about pronunciation, to identify implications of this rhythmic profile for my teaching, and to start implementing changes in content and practice that I saw were needed for the sake of my students.

The critical changes involve having students review parts of speech to facilitate their use of accent rules, starting the course with the two-peak profile, keeping in front of students the centrality of this profile for communicating the semantic essence of a phrase, emphasizing that the defining features of spontaneous speech—short phrases, only one or two peaks, and compressed valleys— help their listeners grasp the semantic essence quickly, and training students in flexibly creating peaks and valleys by manipulating the suprasegmentals of vowels.

The efforts that my pronunciation team and I made to introduce these changes across our entire syllabus have been rewarded by the progress our students have shown. Since they know what they are doing and why, they seem more determined to change their oral skills in the ways we recommend. Our impression, yet to be confirmed empirically, is that they are improving their intelligibility much more quickly now than in the era before the two-peak profile and before being inundated by the myriad ripples of this model of rhythm.

ABOUT THE AUTHOR

Wayne Dickerson is professor emeritus in the Department of Linguistics at the University of Illinois at Urbana-Champaign where he directed the MA-TESOL program and taught courses in pedagogical English phonology (online and on campus) for TESOL candidates and ESL pronunciation for university-level students. He has since helped to develop an online pedagogical phonology course for candidates in the MA-TESOL program at Wheaton College. His research focuses on pedagogical applications of phonetics and phonology, pronunciation pedagogy, the value of orthography for learners, phonological variability, and pronunciation assessment. His two pronunciation textbooks are *Stress in the Speech Stream: The Rhythm of Spoken English* (1989/2004), and associated audio recordings, University of Illinois Press, and (with co-author Laura D. Hahn) *Speechcraft: Discourse Pronunciation for Advanced Learners* (1999), and associated workbooks and audio recordings, University of Michigan Press.

Contact information:
7 Hale Haven Court
Savoy, IL 61874 U.S.A.
E-Mail: dickrson@illinois.edu

REFERENCES

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40,* 351-363.

Bolinger, D. (1961). Three analogies. *Hispanica, 44,* 135-136.

Bolinger, D. (1986). *Intonation and its parts.* Stanford, CA: Stanford University Press.

Brazil, D. (1994). *Pronunciation for advanced learners of English. Student's book.* Cambridge: Cambridge University Press.

Brazil, D., Coulthard, M., & Johns, C. (1980). *Discourse intonation and language teaching.* London: Longman.

Cauldwell, R. (2002). The functional irrhythmicality of spontaneous speech: A discourse view of speech rhythms. *Apples 2,* 1-24.

Cauldwell, R. (2013). *Phonology for listening, teaching the stream of speech.* Birmingham, UK: Speech in Action.

Dale, P. & Poms, L. (1994). *English pronunciation for international students.* Englewood Cliffs, NJ: Prentice-Hall Regents.

Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11,* 51-62.

Dickerson, W. (2004). *Stress in the speech stream, the rhythm of spoken English.* Urbana, IL: University of Illinois Press. [CD-ROM] (Original work published 1989).

Dickerson, W. (2015). A NAIL in the coffin of stress-timed rhythm. In J. Levis, R. Mohamed, M. Qian & Z. Zhou (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 184-196). Ames, IA: Iowa State University. https://apling.engl.iastate.edu/alt-content/uploads/2015/05/PSLLT_6th_Proceedings_2014.pdf

Dickerson, W. & Hahn, L. (in press). *Speechcraft: Discourse pronunciation for academic communication* (2nd ed.). Ann Arbor: University of Michigan Press.

Fries, C. (1945). *Teaching & learning English as a foreign language.* Ann Arbor: University of Michigan Press.

Gilbert, J. (2012). *Clear speech: Pronunciation and listening comprehension in North American English* (4th ed.). Cambridge: Cambridge University Press.

Gleason, Jr., H. (1961). *An introduction to descriptive linguistics* (rev. ed.) New York: Holt, Rinehart and Winston.

Gorsuch, G., Meyers, C., Pickering, L., & Griffee, D. (2013). *English communication for international teaching assistants* (2nd ed.). Long Grove: IL: Waveland Press.

Grant, L. (2017). *Pronunciation for clear communication* (4th ed.). Boston: National Geographic Cengage.

Hahn, L. & Dickerson, W. (1999). *Speechcraft: Discourse pronunciation for advanced learners*. Ann Arbor: University of Michigan Press.

Kingdon, R. (1958). *The groundwork of English intonation.* London: Longmans, Green and Co.

Kjellin, O. (1999). Accent addition: Prosody and perception facilitate second language learning. In O. Fujimura, B.D. Joseph, & B. Palek (Eds.), *Proceedings of LP '98 (Linguistics and Phonetics Conference)* at Ohio State University, Columbus, OH, September 1998 (pp. 373-398). Prague: The Karolinum Press. http://olle-kjellin.com/SpeechDoctor/ProcLP98.html.

Ladd, D. (1996). *Intonational phonology.* Cambridge: Cambridge University Press.

Levis, J. (2018). *Intelligibility, oral communication, and the teaching of pronunciation.* Cambridge: Cambridge University Press.

Miller, S. (2006). *Targeting pronunciation, communicating clearly in English* (2nd ed.). Boston: Houghton Mifflin Company.

Munro, M., & Derwing, T. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning, 48,* 159-182.

Pickering, L. (2018). *Discourse intonation: A discourse-pragmatic approach to teaching the pronunciation of English.* Ann Arbor: University of Michigan Press.

Pike, K. (1945). *The intonation of American English.* Ann Arbor: University of Michigan Press.

Prator, C., Jr. (1951). *Manual of American English pronunciation for adult foreign students.* Berkeley, CA: University of California Press.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic Controversies* (pp. 73-79). London: Edward Arnold.

Tench, P. (1996). *The intonation systems of English.* London: Cassell.

Wells, J. (2006). *English intonation: An introduction.* Cambridge: Cambridge University Press.

---

[i] Note that all of the examples above contrast peak and valley versions of full vowels. The other contrast, not illustrated, is between a peak vowel and an unstressed vowel. Peak-valley switches—*tráffĭc, wálkĭng,* and *brúisĕs*—and valley-peak switches—*Măría, cŏmpóse,* and *cŏllíde*—are also practiced in vowel lessons.