English lexical stress, prominence and rhythm

Book or Report Section

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

Publisher: Routledge

English lexical stress, prominence and rhythm

Jane Setter and Boikanyego Sebina

The University of Reading, UK

j.e.setter@reading.ac.uk; b.sebina@pgr.reading.ac.uk

Abstract

English speech rhythm is closely associated with the patterns of lexical stress and prominence in a stream of speech. Older varieties of English (OVEs), such as British and American English, which usually operate as the model in English language teaching, are often described as 'stress-timed', meaning the time between stressed syllables is more or less equal, in comparison with 'syllable-timed' languages (e.g., French or Cantonese), for which the time between successive syllable onsets is more or less equal. The usefulness of this distinction, however, has been disputed; e.g., Cauldwell (2002) talks about 'functional irrythmicality' in English speech.

Cutler (1984) explains that native speakers of English focus on stressed syllables when listening to a stream of speech as part of the decoding process; i.e., for native speakers, lexical stress and the rhythm of the incoming signal play an important part in perception. Couper-Kuhlen and colleagues (e.g., Auer, Couper-Kuhlen, & Müller, 1999) have shown that speech rhythm plays an important part in the coordination of turn-taking in conversation. Anderson-Hsieh and Venkatagiri (1994) argue that speakers' intelligibility will be affected if they do not sufficiently weaken English unstressed syllables. Such research indicates that the differences in the lexical stress and/or speech rhythm patterns of learners of English, or speakers of New Varieties of English (NVEs) which are not 'stress-timed', could create difficulties in comprehension and cooperative interaction for native speakers of OVEs and also, plausibly, for other speakers of English if they are using similar strategies. However, whether the majority of speakers of English in the world have a speaker of an OVE as their target interlocutor is coming increasingly under question.

This chapter gives an overview of English lexical stress, prominence and speech rhythm in OVEs, including theoretical approaches to their description, and includes suggestions for pedagogical approaches for the English language classroom.

Introduction/definitions

Lexical stress, prominence and rhythm are very much intertwined in speech and pronunciation. They are suprasegmental elements, 'suprasegmental' referring to units larger than the individual sound segment, phoneme or speech sound. Lexical stress in English relates to individual syllables in words and can be defined as 'the force of articulation with which a syllable is uttered' (Arnold, 1957, p. 222) and the auditory effect this has on the listener. This force makes certain syllables more prominent than others, i.e., it makes them stand out more in the stream of speech. These syllables are referred to as being stressed or accented.

At the individual word level, stress begins to manifest itself in words of two or more syllables, one-syllable words being intrinsically stressed if spoken on their own in their citation form. Lexical stress

rules, therefore, concern which syllable in a bi- or multi-syllabic word is pronounced with the greatest force of articulation. However, the rules are somewhat cumbersome and not easily learned.

Syllables in English words are made prominent by manipulating a combination of syllable length (duration), loudness (intensity), changes in pitch, and vowel quality (Roach, 2009). Stressed syllables tend to be longer than unstressed ones, louder than unstressed ones, often have a higher pitch than unstressed ones – or there is an audible pitch difference – and must have a full vowel rather than a reduced/weak vowel such as schwa (/ə/). It should be noted that 'pitch' here is used to refer to both the articulation and the auditory effect. For example, in the noun <desert>, the vowel in the first syllable is stressed and will sound longer, louder and higher in pitch than the second syllable. It will contain a full vowel, whereas the second syllable will not. In the General British accent (see Cruttenden, 2014), we can give the transcription /ˈdɛzət/, with the diacritic /ˈ/ indicating the first syllable is stressed. Compare this with the noun <dessert>, which can be transcribed /dəˈzɜːt/, in which the vowel in the second syllable is stressed, i.e., it is longer, louder and pitch prominent, and contains a full vowel. Waveforms of these words spoken by the first author indicate the differences:

FIGURE 1 HERE

FIGURE 2 HERE

Speech rhythm in English arises from the placement of stressed syllables in utterances normally greater than individual words and from the listener's perception of the placement of these syllables. We are now, therefore, describing rhythmic stress – or 'sentence stress' – rather than lexical/word stress. Prominent syllables may be in single- or multi-syllabic words dependent on the communicative message the speaker wishes to convey, and tend to be in lexical (content) words, such as nouns, main verbs, adverbs and adjectives, rather than functional (grammar) ones, such as determiners, prepositions, conjunctions, auxiliary verbs and pronouns.

For example, in the utterance <tell me she likes desserts made with icecream>, the lexical words are <tell>, <likes>, <desserts>, <made> and <icecream>. Theoretically, the rhythmic stresses are likely to fall on the following syllables (1):

1. ˈtell me she ˈlikes deˈsserts ˈmade with ˈicecream

We can divide this utterance into groups known as the 'foot', each foot beginning with the onset of a stressed syllable. This utterance, therefore, has five feet (2):

2. |ˈtell me she |ˈlikes de |ˈsserts | ˈmade with | ˈicecream

In addition, according to the theory of stress-timing, the time the speaker takes to produce each foot – i.e., the syllables between each stress – will be equal; that is, it will take the same amount of time to say the three syllables <tell me she> from the onset of the first syllable <tell> as it will to say the two syllables <likes de-> from the onset of the syllable <likes> and the single syllable <-sserts>. This is known as 'isochrony' (Abercrombie, 1967).

However, in actual speech, it is entirely possible to produce this utterance with three rhythmic stresses, for example (3), or even fewer:

3. |ˈtell me she likes de |ˈsserts made with | ˈicecream

In (3), it will theoretically take the same amount of time to say <tell me she likes de-> (five syllables) as to say <-sserts made with> (three syllables).

We refer to the lexical words in this utterance as 'stressable', i.e., it is likely that the lexical words will be rhythmically stressed rather than the function words (1), but it is not necessary that they all are (3). However, it is also possible for the non-lexical words to become prominent, depending on prior context and the intentions of the speaker.

    4. |ˈtell me | ˈshe likes de |ˈsserts made with | ˈicecream

In (4), <she> is made prominent to draw the listener's attention to that word, in contrast with another person implicated in the discourse, for example. This will have an effect on the rhythm of the utterance in that we now have the feet <tell me> (two syllables) and <she likes de-> (three syllables) but, in theory, it should not affect the stress-timing; i.e., theoretically, the time it takes to say each foot will still be equal.

Anderson-Hsieh (1992, p. 51) writes: 'Suprasegmentals provide the framework for utterances, they direct the listener's attention to important information in the discourse, and they help to establish cultural harmony between the speaker and listener.' Speakers of varieties of English which use lexical stress and speech rhythm in the way described above will therefore expect stresses and rhythm to be reliable indicators of meaning in an utterance. Deviation from the expected forms can have serious consequences for intelligibility, comprehensibility, and cultural harmony. However, how one predicts lexical stress in new words and whether English really is stress-timed are problematic issues.

Historical perspectives

Lexical stress

Historically, definitions of lexical stress in English centre on the fact that, in multi-syllabic words, some syllables have greater force of articulation than others. Intensity, or loudness, is usually given as a salient feature. For example, Bloomfield (1933), Bloch and Trager (1942) and Pike (1947) all mention intensity or loudness; Jones (1949) and Arnold (1957) refer to the force of utterance of a syllable. This is often followed by a discussion of levels of stress; Jones (1956, p. 142) explains that it is possible to distinguish 'up to four degrees of stress', but notes that it is adequate to refer to three or even two levels. In Jones's *English Pronouncing Dictionary* (EPD)*,* first published in 1917 (current 18th edition, 2011), two levels of stressed syllable are indicated: primary and secondary stress, although this could be counted as three if one takes 'unstressed' to be a level of stress. Similarly, Kenyon and Knott's *Pronouncing Dictionary of American English* (1953) indicates primary and secondary stress, using the term 'accent' in preference to 'stress' (p. xviii). The first edition of Wells's *Longman Pronunciation Dictionary* (1990) indicates tertiary stress, but this was later dropped.

The main stress in a word is associated with 'pitch prominence' (Arnold 1957, p. 225), i.e., there is a noticeable change in pitch when a stressed syllable is uttered. We will refer to this type of stress as primary stress.

As an example, let us consider the words <photograph>, <photography> and <photographic>, as given in (5):

5. a. ˈphotograph
   b. phoˈtography
   c. ˌphotoˈgraphic

The stress mark (ˈ) on <photograph> (5a) indicates that primary stress is on the first syllable. This is likely to be pitch prominent in comparison with the others and, if spoken in citation form, will normally be the onset of a falling intonation pattern. The stress mark on the second syllable of <photography> (5b) indicates the first syllable is unstressed and the linguistically prominent material begins on the second syllable; similarly, this syllable will be the onset of a falling intonation pattern if spoken in citation form. See the chapters on intonation for more information about tone choice.

The stress pattern in <photographic> (5c) is different in that both primary (ˈ) and secondary (ˌ) stress are indicated. Arnold (1957, p. 225) argues that secondary stress 'lacks the pitch prominence always associated with a principle stress'. Therefore, the first syllable of <photographic> is likely to be louder than the unstressed syllables, but not stand out as much as the third syllable which is indicated to be primary and pitch prominent. In general, one can also expect stressed syllables to be greater in length than unstressed ones, and incrementally so; we can therefore expect the third syllable of <photographic> to be longest in the word, followed by the first.

Lexical stress is also linked intrinsically to vowel quality in English; a stressed syllable must contain a full vowel, not a weak one. An unstressed syllable, however, may contain a full vowel or a reduced/weak vowel. There is a lengthy discussion of this in Arnold (1957), where he refers to 'LENIS' and 'FORTIS' vowels (small capitals in the original). LENIS vowels are those 'which are *normally* rhythmically weak' (Arnold, 1957, p. 235, italics in the original): /i, ə, u, o/; FORTIS vowels are those 'which are *normally* rhythmically strong' (Arnold, 1957, p. 235, italics in the original): /iː, e, æ, ɑː, ɔ, ɔː, uː, ʌ, əː, ei, əu, ai, au, ɔi, ɛə/ (vowel symbols from the original). Arnold then covers the rules of assignation of stress to syllables in words, based on word type and whether a syllable contains a FORTIS or a LENIS vowel; this complex exposition continues from page 237 to page 441, i.e., 204 pages. Jones (1949) comments on the lack of useful rules for English lexical stress, and indicates that foreign learners must simply learn the stress of a word when they learn the word.

Rhythm

The first thorough treatment of the notion that English has stressed syllables which are equally spaced in time is attributed to Steele, 'who in 1775 [… tried] to represent speech rhythm by means of musical notation, dividing stretches of speech into "cadences" equivalent to musical bars, each of them beginning with a strong syllable' (Knowles, 1987, p. 144). Although later work on rhythmic grouping in English dispensed with the musical notation, the idea of dividing up groups of syllables beginning with a strong stress (i.e., a foot) has persevered.

Classe investigated the reality of isochronous stress in English prose, commenting that, in normal speech, 'the realisation of a perfectly isochronous rhythm would sometimes produce ludicrous results' (Classe, 1939, p. 70). What is viewed as permissible is the idea of English speech being composed of 'more or less isochronous groups' (1939, p. 133). He concludes in one section that 'perfect isochronism can only be realised when very definite conditions are fulfilled […] similarity of phonetic structure of the groups, including number of syllables' and 'similarity of grammatical structure of the groups, and similarity of connexion between the groups' (Classe, 1939, p. 100).

As Classe notes, these conditions are not commonly met in ordinary speech but are found in poetry.

The controversial terms 'stress-timed' and 'syllable-timed' first appear in Pike (1945). Using what he calls the 'rhythm unit' as a basis for analysis, Pike states:

> The timing of rhythmic units produces a rhythmic succession which is an extremely important characteristic of English phonological structure. The units tend to follow one another in such a way that the lapse of time between the beginning of their prominent syllables is somewhat uniform.
> (Pike, 1945, p. 34)

It is this uniformity of spacing of stressed syllables which gives rise to 'the obscuring of vowels' (Pike, 1945, p. 35), i.e., the tendency of unstressed vowels to weaken.

Pike's rhythm unit conforms to what some phoneticians have called the tone unit (see, e.g., Roach, 2009, p. 144; Brazil, 1994, p. 3). In his examples, each rhythm unit is given only one stress which is tonic/nuclear. The focus is on grammatical phrases or clauses, which links rhythm strongly to the way that grammatical relationships are expressed in the language and is a view held by others (e.g., Lehiste, 1977; Cutler & Isard, 1980; Pisoni, 1981; Halliday, 1989). Pike's main claim concerning stress-timing seems to be that the tonic syllables of rhythm units are equally spaced, and that the nature of a stressed-timed language is that it contains rhythm units which themselves contain one main stress. He refers to this as a stress-timed unit (Pike, 1945, p. 35). It is noted, however, that all the stressed syllables in a possible rhythm unit containing more than one stressed syllable have 'equal lapses of time' between them (Pike, 1945, p. 34). This is the more usual interpretation of stress-timing.

By contrast, syllable-timing is defined by Pike as follows:

> a rhythm which is more closely related to the syllable; […] it is the syllables, instead of the stresses, which tend to come at more or less evenly recurrent intervals.
> (Pike, 1945, p. 35)

Pike asserts that the syllables are 'less likely to be shortened and modified' (1945, p. 35), giving Spanish as an example of a syllable-timed language.

Abercrombie (1967, p. 97) claims that, '[a]s far as is known, every language in the world is spoken with one kind of rhythm or with the other', and attributes differences between the rhythms of French, Telugu and Yoruba on the one hand (syllable-timed) and English, Russian and Arabic on the other (stress-timed) to the way the pulmonic airstream mechanism is utilised during speech. Following work by Stetson (1957, in Kelso & Munhall 1988), later discredited by Ladefoged (1967), Abercrombie (1967) claims that, in syllable-timed languages, it is the chest pulses producing the airstream for the syllables which are equally spaced, which leads to the isochrony of syllables in these languages. Conversely, in stress-timed languages, it is 'the stress-pulses, and hence the stressed syllables' which are isochronous (Abercrombie 1967, p. 97), leading to the necessity to adjust the length and/or speech rate of successive syllables. Abercrombie (1967, p. 98) uses the following sentence to demonstrate the variation:

⎸Which is the ⎸train for ⎸Crewe, ⎸please?

Tapping one's finger for each individual syllable, the syllables would be unequally spaced; however, if one taps just the initial syllable, which is stressed, the taps are equally spaced. Abercrombie concludes:

> […] there is considerable variation in syllable-length in a language spoken with a stress-timed rhythm, whereas in a language spoken with a syllable-timed rhythm the syllables tend to be of equal length.
> (Abercrombie 1967, p. 98)

He notes that few learners of English succeed in mastering stress-timed rhythm without training.

Halliday states that English is a foot-timed language, in which 'all **feet** are more or less the same length' (Halliday 1989: 50, bold in the original). This is in comparison with French and Hindi, which are examples of languages in which 'all **syllables** are more or less the same length' (Halliday, 1989, p. 50). Although basically seeming to uphold the notion of isochrony in English, Halliday (1989, p. 50) draws our attention to the phrase 'more or less'. In some contexts, he says, it is possible to have feet which are exactly the same length. In most contexts in ordinary conversation, however, the lengths of the feet are not likely to be the same, with a two-syllable foot being slightly longer than a one-syllable foot, 'but nothing like twice as long' (Halliday, 1989, p. 50). In addition, he claims that 'the relative duration of syllables in the foot is also entirely systematic; it is determined by a combination of grammatical and phonological factors' (Halliday, 1989, p. 51).

To end his treatment of English speech rhythm, like Abercrombie (1967), Halliday notes the difficulties faced by learners of English in 'getting the rhythm right' (1989, p. 52), especially for those from a syllable-timed language background, commenting that it can cause unintelligibility.

Critical issues and topics

Lexical stress

Various psychological studies of speech perception demonstrate that deviations from typical English stress patterns can cause difficulty in the correct parsing of a message. This is in comparison with speakers of languages with fixed stress, as lexical stress may not be stored in the mental lexicon; see, e.g., Pepercamp and Dupoux (1992, cited in Field 2005).

Cutler (1984) explains that English 'word stress patterns are an integral part of the phonological representations of words in the mental lexicon' (1984, p. 78) and that something which comes close to that representation must be produced by the speaker. It is, therefore, crucial that this close approximation to the mental representation has correct stressing. If this is not achieved, the listener will at the very best have difficulty reconstructing the message. Cutler & Norris (1988) suggest that lexical access is initiated by the occurrence of a stressed syllable, claiming that the high frequency of English content words starting with stressed syllables means that this strategy works very well in English. This is also supported by Grosjean and Gee (1987), who claim that 'stressed syllables (and only they) are used to initiate lexical search' (1987, p. 144), and Fear, Cutler and Butterfield (1995), who highlight the role of strong and weak syllables in speech perception.

If the stressed syllables in a stream of English speech are incorrectly placed, native speakers may process the message as something completely different. Field (2005) discovered that altering the lexical stress in English bi-syllabic words had a negative effect on both native and non-native listeners, but that this effect was lessened if the syllables each contained a full vowel. He further

discovered that moving the stress to the left – i.e., earlier – in words which normally had stress on the second syllable had a lesser impact than moving it to the right on words normally stressed on the first syllable. He gives the example of <followed> stressed on the second syllable rather than the first, suggesting that, if the listener perceived <load> or <flowed>, this would 'shape [their] expectations as to what was likely to follow' in a way which could seriously impede overall intelligibility (Field, 2005, p. 418).

Rhythm

The notion of stress-timing in English has been disputed, both in terms of whether there is such a thing as a stress-timed vs. syllable timed distinction in languages, and also whether English is, in fact, stress-timed at all. Crystal (1996, p. 8) claims that the distinction 'is an extremely crude one, and in its bare form almost certainly wrong'; Marks (1999, p. 194) concludes 'that stress-timing is something of a myth'; Faber (1986, p. 206) refers to it as 'extremely suspect'; Cauldwell (2002) describes English as 'irrhythmical'. Knowles (1987) claims that listeners are 'willing to treat considerably different intervals as perceptually equal' (p. 145).

Reporting on inconclusive experiments designed to measure the presence or absence of stress-timing in metric feet, Lehiste (1977) found inconclusive support for isochrony. However, she explains that, in English, 'the listener expects isochrony' (Lehiste, 1977, p. 262). She concludes that isochrony is an integrated part of the grammar of English by demonstrating that speakers use it to disambiguate ambiguous sentences, and reports that similar findings were made by O'Malley, Kloker and Dara-Abrams (1973), who looked at how sections in parentheses in algebraic expressions are indicated in speech – e.g., <a + (b x c) vs. (a + b) x c>.

Roach (1982) examined spontaneous speech recordings of speakers of Abercombie's (1967) six languages to see whether it is possible to clearly assign them to one category or the other. Two of Abercrombie's claims are tested:

> (i) there is considerable variation in syllable length in a language spoken with stress-timed rhythm whereas in a language spoken with syllable-timed rhythm the syllables tend to be equal in length
> (ii) in syllable-timed languages, stress pulses are unevenly spaced.
> (Abercrombie, 1967, p. 98)

Concerning (i), Roach measured the standard deviation of the syllable durations in milliseconds and found that the claim was not supported. He also dismissed claim (ii). Hypothesising that, in order for the claim to be valid, 'syllable-timed languages would exhibit a wider range of percentage deviations in inter-stress intervals than would stress-timed (the latter being more nearly isochronous)' (1982, p. 77), Roach finds that the so-called stress-timed languages demonstrate the greater range of deviance, concluding that one cannot assign languages to either category solely by measuring the time intervals. Like Lehiste (1977), he concludes that 'a language is syllable-timed if it *sounds* syllable-timed' (Roach, 1982, p. 78), and that 'all languages display both sorts of timing' (Roach, 1982, p. 78). This is supported by Miller (1984), whose experiment on the perception of the timing of different languages by groups of French and English phoneticians and non-phoneticians leads her to conclude that 'each language displays features of both types [of timing] in different proportions' (Miller, 1984, p. 82). Roach (1982) also notes that the same speaker may demonstrate different types of timing, with perhaps more formal language being more rhythmical. If one listens to well-known and, often, rehearsed speeches, such as Martin Luther King's <I have a dream> (delivered on August 28th 1963 – see https://www.youtube.com/watch?v=n82rgdbM9G4 for an

excerpt), there is certainly a strongly rhythmic, almost poetic quality to many of them. This type of speech is referred to by Cauldwell (2002) as having 'elected' rhythmicality (see below).

Comparing English, Thai, Spanish, Italian and Greek speakers producing continuous speech, as in Roach (1982), Dauer (1982) finds that inter-stress intervals in English are no more isochronous than those in Spanish, and that the rhythmic difference between the two languages 'has nothing to do with the durations of inter-stress intervals' (Dauer, 1983, p. 54). To account for the fact that, e.g., English and Spanish sound different rhythmically, Dauer (1983) looks at syllable structure, vowel reduction and stress/accent. She finds that there is 'a greater variety in permissible syllable types' in stress-timed languages (Dauer, 1983, p. 55) and that open syllables (CV, where C stands for 'consonant' and V for 'vowel') are found to predominate in Spanish and French, whereas in English there is much more variation among different syllable types. Dauer also finds that 'there is a strong tendency for "heavy" syllables […] to be stressed and "light" syllables […] to be unstressed' in stress-timed languages (1983, p. 55); 'light' syllables are those structured CV or V, whereas heavy syllables have a consonantal coda (i.e., one or more consonants at the end, including a glide) and may contain consonant clusters (e.g., <stray> /streɪ/ has a cluster of three consonants at the beginning). She notes that the two most common syllable types in English are CVC and CV, with stressed syllables tending to be CVC. Unstressed CV syllables often contain one of a small set of shortened weakened vowels. These features are similar in Arabic and Thai, also considered to be stress-timed. Spanish, however is different, with CV syllables predominating, 'whether stressed or unstressed' (Dauer, 1983, p. 56) and fewer variations in syllable duration. Concerning stress, Dauer (1983) claims that syllable-timed languages tend to have no lexical stress, whereas stress-timed languages do.

Preferring the term 'stress-based' (as used by both Allen (1975) and O'Connor (1973)), like Roach (1982), Dauer suggests a continuum on which languages may be placed depending on how stress-based their rhythm is. Her suggestion is as follows:

					⟶ Stress-based
Japanese	French	Spanish	Greek	Portuguese	English

Figure 3. Dauer's stress-based continuum (1983: 60)

As instrumental studies of rhythm in speech cannot be relied upon when assigning languages to a rhythmic group, Ramus, Nespor and Mehler (1999) refer to Dauer's account of how several factors contribute independently 'the phonological account of rhythm' (Ramus et al., 1999, p. 269). However, this phonological account 'does not explain how rhythm is extracted from the speech signal by the perceptual system' (Ramus et al., 1999, p. 269). Ramus et al. (1999) took instrumental measurements based on consonant and vowel segmentation for eight languages (English, Polish, Dutch, French, Spanish, Italian, Catalan and Japanese) and show that the phonological properties mentioned by Dauer (1983) have 'reliable phonetic correlates that can be measured in the speech signal, and that these correlates predict the rhythm classes' (Ramus et al., 1999, p. 275). They conclude that the more complex a language's syllable structure is, the more likely it is to be nearer the stress-based end of the continuum.

Current contributions and research

Lexical stress and prominence

Recent research into lexical stress and prominence has looked at how speakers realise lexical stress in Lombard speech, i.e., speech produced in noisy environments (see Lombard, 1911). Arciuli,

Simpson, Vogel and Ballard (2014), for example, found that English speakers increase lexical stress contrasts rather than producing all syllables more loudly, although there is an overall increase in speech intensity in noisy conditions. This indicates that speakers place greater importance on prominent syllables in speech production.

Rhythm

As mentioned earlier, Cauldwell's (2002) view is that English is functionally irrythmical. While he does not dispute the occurrence or salience of prominent syllables – in fact, they are the core to the approach presented in *Phonology for Listening* (Cauldwell, 2013) – he refers to regularities in speech rhythm as 'elected', as in poetry, or 'coincidental', i.e., 'the side effects of other higher-order choices made by speakers' (Cauldwell, 2002, p. 1). This is largely because spontaneous speech does not contain long enough units to be perceptually rhythmical. It is Cauldwell's contention that elected regular rhythms occur for a social purpose; although he does not explain what that social purpose could be, Martin Luther King's famous speech could be an example of that phenomenon. Asserting that a 'lack of a regular rhythm in speech production is essential for effective communication' (Cauldwell, 2002, p. 20), he also claims that, were English truly stress-timed, this would distract the speaker and the listener from the meaningful content of the utterance as too much attention would be paid to producing or perceiving formal elements.

As well as contributing to intelligibility, work by Couper-Kuhlen and colleagues demonstrates that speech rhythm is important in conversational turn-taking (e.g., Couper-Kuhlen 1991, 1993; Auer, Couper-Kuhlen & Müller 1999); interlocutors time their responses to fit with the rhythmic pattern of the preceding turn to the extent of leaving short pauses at the start of a turn, if necessary, to match the rhythmic stress of prominent syllables. This could be an example of Cauldwell's (2002) social purpose, or Anderson-Hsieh's (1992) 'cultural harmony'. Szczepek Reed (2010) shows that, in interactions in English between speakers of so-called stress-timed and syllable-timed languages, syllable-timed language speakers harmonise with stress-timed speakers mainly for the first few turns before moving back to syllable-timing. Her 2012 paper considers the implications of rhythm and turn-taking for learners of English.

Main research methods

The main research methods in lexical stress and prominence largely involve measurements of syllable duration, intensity, pitch change and vowel weakening, and the effect of re-location of the stressed syllable on intelligibility, i.e., on how well a speaker's message is understood by a listener. A selection of studies using these methods has been cited above. In this section, the focus will be on rhythm metrics.

Rhythm metrics are mathematical formulae used to classify languages into rhythmic classes (Arvaniti, 2012). Ramus et al. (1999) could be said to be the pioneers of rhythm metrics as they were the first to devise mathematical equations that could quantify languages into rhythmic classes. Ramus et al. (1999) and others after them (Arvaniti, 2012; Dellwo, 2006; Grabe & Low, 2002; Low, Grabe & Nolan, 2000; White & Mattys, 2007) were largely influenced by language acquisition research. In particular, they were influenced by the finding that new-born babies were able to discriminate between their mothers' language and a language with a different rhythm such as English and Spanish, but unable to distinguish between their mothers' language and a language with the same rhythm, e.g., English and Dutch (Nazzi, Bertoncini & Mehler, 1998). The conclusion drawn from the finding was that language acquisition is to some extent dependent on speech rhythm. For

this reason, Ramus et al. (1999) set out to find out how languages can be classified based on rhythmic classes.

Ramus and colleagues' approach moved away from isochrony; instead, they concentrated on the acoustic phonetic element of rhythm. Their rhythm metric required segmentation of an utterance into successive vocalic and consonantal intervals, and measurement involved the duration of each of these intervals. An example phrase from Ramus et al. (1999, p. 272) is <next Tuesday on>, phonetically transcribed in the original as /nɛkstjuzdeiɔn/ and with the following vocalic and consonantal intervals: /n/ /ɛ/ /kstj/ /u/ /zd/ /eiɔ/ /n/. Measurements were then subjected to further calculations, as follows (Ramus et al., 1999, p. 270):

- the proportion of vocalic intervals in the utterance, i.e., the sum of vocalic intervals divided by the total duration of the utterance (%V);
- the standard deviation of vocalic intervals within the utterance (ΔV);
- the standard deviation of consonantal intervals within the utterance (ΔC).

These rhythm metrics were applied to short declarative sentences in eight languages, largely drawn from a multi-language corpus, which were considered stress-, syllable- and mora-timed. Results indicated that the so-called stressed-timed languages (English, Polish and Dutch) exhibited a low %V and a high ΔC, whereas languages considered syllable-timed (Spanish, Italian, Catalan and French) displayed a high %V and a low ΔC. Japanese, a 'mora-timed' language, did not cluster with either stress or syllable-timed languages. The findings demonstrated that languages do indeed fall into different rhythm classes, contrary to the findings of Roach (1982) and Dauer (1983). Ramus et al. (1999) concluded that the measurements which successfully distinguish languages into rhythm classes are %V and ΔC.

A possibly much better known metric is that developed by Low, Grabe and Nolan (2000), later expanded in Grabe and Low (2002). Like Ramus et al. (1999), Low and colleagues measured the acoustic phonetic component of rhythm by dividing utterances into vocalic and consonantal intervals, and developed a Pairwise Variability Index (PVI) which reflected the level of variability in consecutive vocalic and consonantal intervals. Unlike Ramus et al., they used constructed sentences such as <Grace works through huge mounds each Friday>, in which each syllable is likely to be pronounced with a full vowel, and <Grace was tired of Matthew Freeman>, in which only <Grace>, <tired>, and the first syllables of <Matthew> and <Freeman> are likely to have full vowels in so-called stress-timed languages. Low and colleagues argued that, unlike the rhythm metric in Ramus et al. (1999), the PVI would not show spurious variability caused by speaker rate variation within and across sentences.

The PVI is expressed as both raw PVI (rPVI) and normalised PVI (nPVI). The rPVI is normalised to nPVI by dividing each absolute difference between successive intervals by their mean. The total is multiplied by 100 to yield values comparable to rPVI. Grabe and Low (2002) asserted that languages that are considered stress- and syllable-timed contrasted in the durational variability of vowels and thus proposed that nPVI should be reserved for vowels and rPVI for consonants. They subjected the languages used in Ramus et al. (1999) to nPVI. Their classification of English, Spanish and French agreed with Ramus et al. (1999). However, whereas Ramus et al. (1999) did not group Japanese with either stress-timed or syllable-timed languages, Grabe and Low (2002) grouped it with syllable-timed languages.

Many studies have been carried out which use this metric or slightly modified versions and, in most cases, use it to classify languages as either stress- or syllable-timed. For example: Gibbon and Gut (1999) show that Nigerian language Ibibio is considered syllable-timed; Mok and Dellwo (2008) use

the PVI to show that Cantonese and Mandarin are syllable-timed. Nolan and Asu (2009) indicate that a foot-based PVI measurement may be more successful in capturing rhythmic differences among languages and, looking at Estonian, English, Mexican Spanish and Castilian Spanish, suggest that a language can be both stress- and syllable-timed. Arvaniti (2012) and Bunta and Ingram (2007) establish that nPVI is more successful in distinguishing the speech rhythm of monolingual English speakers than the rPVI.

Other metrics which have been developed include the 'rate-normalised interval measures of vocalic interval variation' Varco Vowel (VarcoV) and Varco Consonant (VarcoC) (White & Mattys 2007: 501). Among other things, it has been shown that VarcoV is particularly successful at showing rhythmic differences among speakers switching between their first and second languages (White & Mattys, 2007). See White and Mattys (2007) for a review and comparison of all the measures mentioned in this section.

Recommendations for Practice

Lexical stress

Although lexical stress rules for English are somewhat cumbersome, there have been attempts to teach it using simplified rules, by analogy with other words with similar stress patterns, or by focusing on predictable patterns in words containing (usually) stress-attracting suffixes.

Stress-attracting suffixes either attract the primary stress on to a predictable syllable in the suffix or predictably affect the placement of the primary stress elsewhere within a word. For example, the suffix <-ation> is predictably stressed on the first syllable. In a word such as <harmonize> /ˈhɑːmənaɪz/, the rhythmic stress on the first syllable stays in place as a secondary stress once the suffix is added to form <harmonization>, but the primary stress moves to the suffix: /ˌhɑːmənaɪˈzeɪʃn̩/. By comparison, the suffix <-ic> attracts the primary stress to the syllable preceding it. In, e.g., <demon> /ˈdiːmən/, the primary stress is on the first syllable, but in <demonic> it moves to the second syllable, which is the one now preceding the suffix: /dɪˈmɒnɪk/. It is possible to train learners encountering words formed using stress-attracting suffixes to identify them as such and apply the appropriate rule (see, e.g., Hewings, 2004, p. 122-123).

This is used to good effect in Hancock's (1995) pronunciation game 'Happy Families' (pp. 27-31). In the traditional game, players have to collect four members of a number of families by asking each other for specific family members until all cards are collected. In Hancock's variant, the families are four cards comprising a non-complex lexeme which acts as a stem, followed by three words formed using the suffixes <-ity>, <-ize> and <-ization>. Both <-ity> and <-ization> are stress attracting, <-ity> by moving the primary lexical stress to the syllable preceding it, and <-ization> by attracting the stress on to the second syllable of the affix. <-ize> is stress-neutral, i.e., it does not affect the stress in the stem. Players must collect the word families by asking opposing players for cards in a particular family, pronouncing them correctly, by saying: '(Name), can I have (word), please?' For example, if a player had been dealt <civil> and <civilize>, the remaining family members <civility> and <civilization> must be collected by asking, e.g., 'Murat, can I have *civility* please?' (Hancock 1995: 27, italics and underlining in the original). This game is preceded by a class activity which demonstrates how affixes affect the stress placement in words, using the three affixes in the game.

Approaches using simplified rules, such as those in Kelly (2000), Hancock (2003) and Hewings (2004), often focus on the differences between two syllable nouns and verbs – e.g., IMport (n.) and imPORT

(v.) – and 'core vocabulary' (Kelly, 2000, p. 69), before moving on to 'rule of thumb' suggestions for compound and complex words.

Examples of approaches in which learners match words with similar lexical stress patterns include Kelly's 'Find a partner' (2000: 77), using the following commonly-seen convention of a string of letters 'o' to represent syllables, with large 'O' denoting primary stress, as visual reinforcement (6):

(6)   Politician   ooOo
        It's important   ooOo

Learners are given a set of words and phrases, asked to find the stressed syllable, and are then required to match the words and phrases to those with similar patterns. Hancock (1995) has a number of pronunciation games which use the same basic approach but with filled circles and, as Kelly (2000, p. 76) notes, 'some teachers like to use boxes'.

There are also approaches which teach learners how to produce prominent syllables in comparison with weak syllables, most of which focus on syllable length as an indicator. Gilbert (2012) and Hewings (2004), for example, recommend the kinaesthetic reinforcement of stretching an elastic band between one's thumbs simultaneously with producing the prominent syllable in order to make it tangibly longer than surrounding syllables.

Rhythm

One of the most common basic approaches to teaching English speech rhythm involves starting with a sentence in which all words are single syllable stressed words and gradually adding increasing numbers of weak or unstressed syllables. Learners produce the sentences by beating time or clapping a regular rhythm only on the words which appeared in the first sentence. For example (7):

7   DOGS CHASE CATS
        the DOGS CHASE the CATS
        the DOGS are CHASing the CATS
        the DOGS will have CHASED the CATS
        all the DOGS will have CHASED all the CATS

Assuming the regular rhythm is maintained, this encourages the compression of syllables between the stresses, resulting in a stress-timed rhythm.

Chela-Flores (1998) uses a 'Morse code' approach, recommending each dot be spoken as 'ti', representing a weak syllable, and each dash be spoken as 'TA' or 'TAA' to represent a stressed syllable, with 'TAA' indicating the main stress in the utterance, placed on the last stressed syllable.

For example, a three syllable pattern would look like this (8a):

8a   _ . _   TA ti TAA
        . _ .   ti TAA ti

Phrases which match these two patterns are as follows (8b):

8b   What's your name?   _ . _
        She's going.   . _ .
        He couldn't.   . _ .

| Believe it. | . _ . |
| Twice a day. | _ . _ |

This builds up to patterns as long as the following (9):

9   ._.._.._ (ti TA ti ti TA ti ti TAA) – e.g., It's time I was having my lunch.
._..._... (ti TA ti ti ti TAA ti ti ti) – e.g., He took it to the railway station.

Learners are first trained on the patterns, then on matching sentences to the patterns, then on trying to produce sentences with the patterns.

Hewings (2004), in a series of exercises, focuses on the functional aspect of identifying and producing prominent words in sentences. He includes work on contrastive stress in sentences. There is also guidance for teachers in Kelly (2000). Hancock (2003) has a number of exercises on sentence stress, including noticing that function words <a>, <of> and <or> are produced with a much reduced vowel in connected speech.

Cauldwell's teacher-oriented text (2013), focusing on teaching the listening process and the identification of stressed syllables in speech, looks at what happens when words get squeezed together in rapid speech produced by English native speakers, as this can lead to unintelligibility. He highlights areas of speech referred to as 'squeeze zones', i.e., the places where it is difficult to understand a word because the spontaneous speech stream has caused the 'sound shapes' to be compressed so much that the output does not resemble citation forms of that word very closely. If you are familiar with the excellent *Streaming Speech* (Cauldwell, 2001), you will know that these zones were animated so they squashed together to give a visual representation of what happens in rapid speech. Cauldwell's position is that '[s]tress-timing theory is (or should be) dead' (2013, p. 141). He admits, however, that the notion of stress-timing is useful from a pronunciation teaching and learning point of view, as exercises such as (7) above, which aim at the learner trying to squash more and more words in between stressed syllables, encourage learners to practice producing words in squeeze zones.

Finally, Coniam (2002) recommends using speech technology to train teachers to recognise the difference between English of different rhythmic types. Using the waveform element of a speech display (see Figures 1 and 2 above) and recordings of native and non-native speakers, he encourages teachers to notice the difference between English spoken with a stress- and syllable-timed rhythm. Teachers reported that being able to see the visual information 'had given them a clearer perspective of syllable-timed speech' (Coniam, 2002, p. 39).

Future directions

Global Englishes and NVEs do not necessarily demonstrate lexical stress, prominence or rhythmic patterns in the same way as OVEs. However, as we have seen, unexpected word stress can lead to extreme difficulty in lexical retrieval. It is also important that speakers make prominent syllables in words stand out in a stream of speech so a listener can catch the meaning. At discourse level, making certain words prominent is also crucial, and is discussed elsewhere in this volume.

We would like to suggest, therefore, that where classroom work on stress-timed rhythm may fall out of favour – fun though it can be – more focus will be placed on harmonising lexical stress placement across speakers of English in order to support intelligibility in international contexts. This should be practiced in multi-cultural interactive contexts, where possible, preferably between speakers from

different rhythmic language backgrounds in order to experience as much as possible the variance in stress and rhythm in different speaker groups. Work could be on improving accommodation strategies in this respect among learners, not only in interactions between learners of English and speakers of NVEs, but also among OVE speakers, whose language is not always the most easily understood.

In addition, the focus in longer stretches of speech should be in making the prominent, information-bearing syllables stand out more, but not in generating strict stress-timed rhythms. From the research, it is apparent that rhythm of this kind does not occur very often in spontaneous speech anyway.

Further reading

Arnold, G. F. (1957). Stress in English words. Lingua, 6, 221-441.

A thorough and comprehensive account of lexical stress in English.

Cauldwell, R. (2013). Phonology for listening. Birmingham: speechinaction

Teachers of English pronunciation will fully understand the issues described in this text and the frustration that can occur when the patterns in real speech do not match with what the text books tell us.

Roach, P. (2001). Phonetics. Oxford: Oxford University Press, Chapter 5.

This chapter on suprasegmentals includes an excellent short introduction to lexical stress, prominence and rhythm, written in Roach's engaging prose.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. Journal of Phonetics, 35(4), 501-522.

An overview of rhythm metrics currently used in research.

Related topics

Syllable structure; intonation.

References

Abercrombie, D. (1967). *Elements of General Phonetics.* Edinburgh: Edinburgh University Press.

Allen, G. D. (1975). Speech rhythm: its relation to performance universals and articulatory timing. *Journal of Phonetics, 3*, 75-86.

Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System, 20* (1), 51-62.

Anderson-Hsieh, J. & Venkatagiri, H. (1994). Syllable duration and pausing in the speech of Chinese ESL speakers. *TESOL Quarterly*, *28*(4), 807-812.

Arciuli, J., Simpson, B. S., Vogel, A. P., & Ballard, K. J. (2013). Acoustic Changes in the Production of Lexical Stress during Lombard Speech. *Language and speech*, 0023830913495652.

Arnolda, G. F. (1957). Stress in English words. *Lingua*, *6*, 221-441.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*(3), 351-373.

Auer, P., Couper-Kuhlen, E., & Müller, F. (1999). *Language in time. The Rhythm and Tempo of Spoken Interaction*. Oxford: Oxford University Press.

Bloch, B. & Trager, G. L. (1942). *Outline of linguistic analysis.* Baltimore: Linguistic Society of America at the Waverly Press.

Bloomfield, L. (1933). *Language*. New York, NY: Taylor & Francis.

Brazil, D. (1994). *Pronunciation for Advanced Learners of English.* Cambridge: Cambridge University Press.

Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish-and English-speaking 4-and 5-year-old children. *Journal of Speech, Language, and Hearing Research, 50*(4), 999-1014.

Cauldwell, R. (2001). *Streaming speech*. Birmingham: speechinaction

Cauldwell, R. (2002). The functional irrhythmicality of spontaneous speech: A discourse view of speech rhythms. *Apples – Journal of Applied Language Studies*, *2*(1), 1-24. Available from http://apples.jyu.fi. [Accessed: 10th June 2015].

Cauldwell, R. (2013). *Phonology for listening*. Birmingham: speechinaction.

Chela-Flores, B. (1998). *Teaching English rhythm: from theory to practice.* Caracas, Venezuela: Fondo Editorial Tropykos.

Classe, A. (1939). *The Rhythm of English Prose.* Oxford: Basil Blackwell.

Coniam, D. (2002).Technology as an awareness-raising tool for sensitising teachers to features of stress and rhythm in English. *Language Awareness*, *11*(1), 30-42.

Couper-Kuhlen, E. (1991). A rhythm-based metric for turn-taking. In *Proceedings of the 12th International Congress of Phonetic Sciences* (Vol. 1, pp. 275-278).

Couper-Kuhlen, E. (1993). *English speech rhythm. Form and function in everyday verbal interaction*. Amsterdam: Benjamins.

Cruttenden, A. (2014). *Gimson's pronunciation of English*. (8th ed) New York: Routledge

Crystal, D. (1996). The past, present and future of English rhythm. *Speak Out*,*18*, 8-13.

Cutler, A. (1984). Stress and accent in language production and understanding. In D. Gibbon & H. Richter (Eds.), *Intonation, accent and rhythm: Studies in discourse phonology* (pp.77-90). Germany: Library of Congress Cataloging.

Cutler, A. & Isard, S. D. (1980). The production of prosody. In B. Butterworth (Ed.), *Language Production Volume 1: Speech and Talk (*pp.245-269). London, New York, Toronto, Sydney & San Francisco: Academic Press.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human perception and performance*, *14*(1), 113-121.

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

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient forΔ C. *Language and language-processing*, 231-241

Faber, D. (1986). Teaching the rhythms of English: a new theoretical base. *IRAL-International Review of Applied Linguistics in Language Teaching*, *24*(3), 205-216.

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *The Journal of the Acoustical Society of America*, *97*(3), 1893-1904.

Gibbon, D., & Gut, U. (2001). Measuring speech rhythm. In *INTERSPEECH* (pp. 95-98).

Gilbert, J. (2012). *Clear speech from the start* (2nd ed.). Cambridge: Cambridge University Press.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology, 7*, 515-546.

Grosjean, F. & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition*, *25*, 135-155

Halliday, M. A. K. (1989). *Spoken and written language* (2nd ed). Oxford: Oxford University Press.

Hancock, M. (1995). *Pronunciation games*. Cambridge: Cambridge University Press.

Hancock. M. (2003). *English pronunciation in use*. Cambridge: Cambridge University Press.

Hewings, M. (2004). *Pronunciation practice activities*. Cambridge: Cambridge University Press.

Jenkins, J. (2000). *The Phonology of English as an International Language.* Oxford: Oxford University Press.

Jones, D. (1949). *An outline of English phonetics*. Cambridge: Heffer.

Jones, D. (1956). *The pronunciation of English*. Cambridge: Cambridge University Press.

Jones, D. (2011). *English Pronouncing Dictionary* (18th ed.). Edited by P. Roach, J. Setter and J. Esling. Cambridge: Cambridge University Press.

Kelly, G. (2000). *How to teach pronunciation*. Harlow, Essex: Pearson Education.

Kelso, J. A. S. & Munhall, K. G. (1988). *RH Stetson's Motor Phonetics: a retrospective edition*. Boston, Toronto & San Diego: College-Hill Press. (Originally published as Stetson, R. H. 1928. *Motor Phonetics, a Study of Speech Movements in Action*. The Hague: Archives Néerlandaises de Phonétique Experimentale (Vol. III) and: 1951. *Motor Phonetics* (2nd Edn). Amsterdam: North Holland.)

Kenyon, J. S. & Knott, T. A. (1953). *A pronouncing dictionary of American English*. Springfield, MA: Merriam-Webster.

Knowles, G. (1987). *Patterns of Spoken English*. Harlow, Essex: Longman.

Ladefoged, P. (1967). Stress and respiratory activity. *Three Areas of Experimental Phonetics*. London: Oxford University Press.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5*, 253-263.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English. *Language and speech*, *43*(4), 377-401.

Marks, J. (1999). Is stress-timing real? *ELT Journal, 53*(3), 191-199.

Miller, M. (1984). On the perception of rhythm. *Journal of Phonetics* , *12*, 75-83.

Mok, P. P., & Dellwo, V. (2008, May). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. In *Proc. Speech Prosody* (pp. 423-426).

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance, 24*(3), 756-766.

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

O'Connor, J. D. (1973). *Phonetics*. Harmondsworth, Middlesex: Penguin.

O'Malley, M. H., Kloker, D. R. & Dara-Abrams, B. (1973). Recovering parentheses from spoken algebraic expressions. *IEEE Transactions on Audio and Electro-acoustics AU, 21*, 217-220.

Peperkamp, S., & Dupoux, E. (1992). A typological study of stress 'deafness.' In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 203–240). Berlin: Mouton de Gruyter.

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

Pike, K. L. (1947). *Phonemics*. Michigan: Ann Arbor, University of Michigan Press.

Pisoni, D. B. (1981). Some current theoretical issues in speech perception. *Cognition*, *10*, 249-259.

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

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic Controversies: essays in linguistic theory and practice in honour of F R Palmer* (pp. 73-79)*.* London: Edward Arnold,

Roach, P. (2001). *Phonetics*. Oxford: Oxford University Press.

Roach, P. (2009). *English Phonetics and Phonology* (4th ed.). Cambridge: Cambridge University Press.

Szczepek Reed, B. (2010). Speech rhythm across turn transitions in cross-cultural talk-in-interaction. *Journal of Pragmatics, 42*(4), 1037-1059.

Szczepek Reed, B. (2012). A conversation analytic perspective on teaching English pronunciation: The case of speech rhythm. *International Journal of Applied Linguistics*, *22*(1), 67-87.

Wells, J. C. (1990). *Longman pronunciation dictionary*. Harlow, Essex: Pearson Education.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501-522.