# Current Debates in the Theory and Teaching of English L2 Pronunciation

*Sandor Danka*
Institute for English Language Education (IELE)
Assumption University of Thailand
Email: s.danka.tl@gmail.com

## Abstract

Ironically, the single concept that appears to be universal in the field of English pronunciation research and instruction, its common denominator as it were, is diversity. Research theory and classroom practice have both convincingly proven that explicit training may indeed lead to improvements in a learner's clarity of speech, but it seems that everything else is open for debate. Variability in opinions begins with different interpretations of basic concepts, of individual speech sounds, syllables, phrases and utterances. Correctly identifying research foci, and by extension, educational priorities for classroom instruction also divides English L2 pronunciation professionals. Models are yet another area of contention – whether to focus on traditional pronunciation points of reference, e.g. features of Received Pronunciation or General American, or to concentrate instead on interactions where no native speaker is present, as proposed by the English as an International Language (EIL) framework. Next, dispelling doubts about its effectiveness can be a challenging endeavour when progress often manifests in small increments which require a significant investment of time and effort. Finally, the decision to incorporate digital technology and the Internet into the pronunciation classroom remains a dividing line between enthusiasts and those that call CALL (Computer-Assisted Language Learning) a fad that will soon pass. The purpose of this paper is to examine these hotly debated issues, while acknowledging that its emphasis on depth may be at the expense of breadth. Its scope will allow it to touch upon but the most significant disputes, those that bridge research theory with English L2 pronunciation classroom practice.

**Keywords:** English L2 pronunciation instruction; curriculum and materials design; pronunciation teaching effectiveness

"Two roads diverged in a wood, and I –

I took the one less travelled by,

And that has made all the difference."

(*The Road Not Taken* by Robert Frost – excerpt)

## Conceptual Qquandary

Discussions about the 'units' of human communication often begin with what is commonly considered the smallest meaningful "building block" (Pennington, 1996:25) of spoken language, the *phoneme*. Also called a *segment*, it refers to individual speech sounds, i.e., vowels and consonants (Dalton & Seidlhofer, 1994; Ladefoged, 2000). When replacing one with another in a word results in a difference in meaning, these sounds are considered different phonemes. The clearest such contrast is a *minimal pair*, i.e. two words with only one phoneme difference between them, such as *pet/bet* (word initially), *pet/pot* (word-medially) and *pet/pen* in a word-final position. Theoretical phoneticians argue that a phoneme is not a single speech sound but rather a group of variants that by themselves do not change meaning (Ladefoged, 2000; Davenport & Hannahs, 2010). They propose two distinct levels: an abstract, underlying concept of 'phoneme' and its concrete, surface representation, called 'allophone,' the form that actually gets produced in speech. Examples for /t/ allophones, for example, would be (i) an aspirated *t* (pronounced with a small puff of air) word-initially as in *tea*, and (ii) a glottal stop in a syllable-final position, e.g. *got*. In practice, however, especially in a lower-level pronunciation classroom, this distinction is rarely considered significant (Underhill, 2005). Beyond simple precision in articulation, extreme clarity when speaking fluently can sometimes add additional, non-linguistic layers to communication, e.g. it can suggest a careful weighing of the impact of the intended message, but could also signal speaker anxiety (Pennington, 1996).

The next focus of phonological inquiry, one that operates on longer stretches of speech (Gimson & Cruttenden, 2008), i.e., beyond individual phonemes, regardless of "whether they are long texts or just one word" (Dalton & Seidlhofer, 1994:42) is called the transsegmental (Pennington, 1996:19), or, to use more widely accepted terminology, the suprasegmental or prosodic unit of spoken language. These features extend from single syllables to whole utterances. Their internal structure and function will be described below.

Although there appears to be no consensus as to its precise phonetic definition, it is generally taken for granted that one or more phonemes together form a syllable. It could be interpreted "in terms of the inherent sonority of each sound" (Ladefoged, 2000:227), where sonority means loudness relative to other sounds with the same length, stress and pitch. In this theory, syllables are marked "by peaks of prominence" (ibid), i.e. how much a sound stands out from its surroundings because of its sonority, length, stress and pitch. This method is believed to be the most reliable of all the competing interpretations (Brown, 2015), although its critics have pointed out that assigning prominence to a syllable is a subjective speaker choice. The next interpretation regards syllables as "abstract units that exist at some higher level in the mental lexicon of the speaker [as] units in the organization of [speech] (Brown, 2015:229). In addition, Davenport and Hannahs (2010:74) offer two

further constructs. Firstly, reiterating Crystal (1986:246), they propose an analytical approach that looks at constituent parts: a vowel segment (or vowel-like nucleus) "preceded and/or followed by zero or more consonantal segments." An inherent shortcoming of this method, however, is that it does not account for syllabic consonants, i.e., /l/, /n/ or /m/, which can form syllables on their own, as in *bottle*, *button*, or *bottom*. One could argue that during careful enunciation, there is a schwa inserted, thus providing the required vowel. Nevertheless, in faster or more colloquial speech, the definition above is incorrect. Secondly, if one adopts an articulatory point of view, producing a syllable may involve "a chest pulse" or "initiator burst," i.e., "a muscular contraction in the chest involving the lungs [therefore] each syllable is produced with one burst of muscular energy." The authors themselves are quick to disprove this latter interpretation, saying that if this were true, every syllable would be of the same length which, especially in connected speech, is hardly ever the case.

The next structural level is called an utterance, a continuous stretch of speech with a clear pause on either side (Brown, 2017; Roach, 2010). It may be as short as one syllable, as in *yes*, or as long as a clause or sentence. Pennington (1996:139) proposes an additional term, 'pause group' for a unit of speech between two pauses, but it is less common in the literature. Ladefoged (2000) points out an important distinction that is worth mentioning here: components under investigation in phonology are most often chunks of information "rather than syntactically defined units" (page 100), i.e., they may coincide with grammatical structures, with phrases, clauses or sentences, but this is not always the case. They are more likely to follow the dynamic variation of pitch, the frequency of vocal cord vibration which manifests in a speaker's voice going up or down.

Pitch 'highs and lows,' together with pauses before and after them, form tone groups. Pausing for breath or to think usually occurs at tone group boundaries, but pauses may also happen at other places to show hesitation, to look for a word, or as Davenport and Hannahs (2010:87) point out, "to keep someone waiting or to create suspense (or because of poor memory)." Roach (2010:155) brings up as examples politicians and philosophers, who use the technique of controlled pauses, of stopping in unlikely places during debates because then "they are less likely to be interrupted." The choice of where to place a tone group boundary, where to drop one's voice, may also provide two different interpretations of a pair of sentences which are ambiguous when written down (Roach, 2010). For example, in the sentence *The lady hit the man with an umbrella*, was the umbrella a weapon the lady wielded, or did she hit a man who was holding an umbrella in his hand?

Tone groups have many different names. Roach (2010:134) enumerates several alternative terms, including "tone-unit," "intonation group," "intonation unit," or "intonation phrase" as synonyms. Ladefoged (2000) and Ashby (2011) propose "intonational phrase,"

Büring (2015) calls it "prosodic phrase" or "prosodic domain," and Gilbert (2008) adds "thought group" to the list. These units have varying pitch patterns depending on the prominence of certain syllables. This prominence (the extent to which a syllable stands out from others) may occur because of some inherent property of words (word stress) or because of speaker intent (highlighting). These two concepts will be discussed next.

A syllable which is especially prominent, i.e. carries "the major pitch change in a tone group" (Ladefoged, 2000:96) is called a tonic syllable. These syllables are said to have tonic stress, also known as intonation peak or tonic accent (ibid). There seems to be no agreement as to the precise definition of the terms *stress* and *accent*, with linguists proposing a conflicting variety of interpretations. Roach (2010) uses *stress* to refer only to syllables (as in "word stress"), as does Cutler (2015), but she calls it "salience." Dalton and Seidlhofer (1994), meanwhile, reserve it for the discourse level ("sentence stress"). Brown (2017), Ladefoged (2000) and Davenport and Hannahs (2010), on the other hand, combine these approaches, calling *stress* a variation in prominence at both word and sentence levels. Jenkins (2000) and Levis (1999) call emphasis on a word at sentence level "nuclear stress." Levis (1999) presents a comprehensive list of authors who use different names for stressed syllables in an utterance:

> "*focus* (Gilbert, 1993; Grant, 1993), *emphasis* (Smith, Meyers, & Burkhalter, 1992), *prominence* (Dalton & Seidlhofer, 1994), *sentence stress* (Dauer, 1993), *major sentence stress* (Avery & Ehrlich, 1992), *phrase stress* (Chan, 1987), *primary phrase stress* (Dickerson, 1989a), and *main stress* or *tonic accent* (Pennington, 1996)" (cited in Levis, 1999:39).

As regards *accent*, Gimson and Cruttenden (2008) use it to mean extra emphasis on syllables, Wells (2008:783) suggests "pitch prominence on a word," i.e. synonyms for *stress* as described above, while Brown (2017), Crystal (1986), Derwing and Munro (2005), Jenkins (2000), and Roach (2010) set *accent* apart for varieties in the pronunciation in English, e.g. a cover term for phonological variety spoken in an often geographically defined L1 community, e.g. what is colloquially termed "American accent," "Australian accent," and so on.

From the level of syllable stress upwards, every suprasegmental feature involves a degree of contrast. They are not prominent *per se*; they stand out only in comparison with other items (phonemes, syllables or words) in the same utterance. Roach (2010) offers a two-pronged analysis of stress, from the points of view of production and that of perception. In multisyllabic English words, one syllable is always more prominent than others. On the production side, to make it more prominent requires more effort (Pennington, 1996), more strength (Roach, 2010) or greater respiratory energy (Ladefoged, 2000). From a listener's point of view, length, loudness and pitch are significant. Correctly stressing the most prominent syllable in a word is governed in part by tradition and in part by context, and

lexical stress anomalies often cause intelligibility problems (Roach, 2010). Recent research by Brown (2017) seems to indicate that L1 listeners (native speakers of English) are at a disadvantage at understanding words with misplaced stress. Vocabulary is 'sorted' in their mental lexicon into stress pattern categories; therefore, when processing auditory input, they will first search for a corresponding stress pattern, and only then consider lexical meaning. Field (2005) adds to this another dimension, i.e., which direction, forward or backwards, the misplaced word stress is shifted. He found that in the case of L1 (native English) speaking listeners, the impact on intelligibility of a shift to the left (= forwards) was statistically less significant than when it shifted to the right (towards the end of the word). His results may be due to the way many L1 listeners 'chunk' English utterances into words, with stressed syllables often being perceived as word-initial (Cutler, 2015).

The most common position for a tonic syllable (marked here with capital letters) in unmarked, i.e., normal situations, is on the last content word (noun, verb, adjective or adverb) of a tone unit, as in *white HOUSE* (a residential building which is white in colour), unless a change in meaning requires it to shift, e.g. (the) *WHITE House* (the office of the American president). Tonic syllable placement can also indicate the focus of information, a process also known as highlighting, tonic prominence (Roach, 2010:153) or foregrounding (Dalton & Seidlhofer, 1994:81). This has two separate functions. First, by making subjective choices about what to emphasize, the speaker draws the listener's attention to what is considered the most important part of their message. In this sense, any syllable within the utterance may bear tonic stress. For example, *it was my WIFE*... (not my sister) is different from *it was MY wife*... (not yours). This is generally called contrastive stress (Dalton & Seidlhofer, 1994; Ladefoged, 2000; Pennington, 1996). Second, simply through the placement of the tonic syllable, a speaker can signal words of high information content with a falling tone (Roach, 2010:157), thus backgrounding what is common or shared (rising/rising-falling tone). Perhaps the two most problematic characteristics of these less important tone units are increased speed and reduced loudness (for a complete list, see Roach, 2010:158), which often cause difficulties. Jenkins (2000, citing Nash, 1969 and Van Els & De Bot, 1997), adds lack of obvious 'chunking' of the language, i.e. too short or non-existent pauses, which make a (false) impression of speed, not giving listeners enough time to process what they have heard. Consequently, it is a common complaint among learners of English as a second or foreign language, especially at lower levels, that native speakers speak too fast. It is a teacher's job then to provide adequate "ear training" (Gimson & Cruttenden, 2008:334) if students cannot make sense of this "undifferentiated babble" (Pennington, 1996), this "acoustic blur" (Brown, 2017; Cauldwell, 2013, 2014).

To sum up, with higher information content than sounds, suprasegmental features of speech – a cover term for intonation, voice quality, rhythm, stress, tone (Pennington,

1996), length (Ladefoged, 2000) and tempo or speech rate (Gimson & Cruttenden, 2008) – are not only longer than individual phonemes, but are also far more than basic tools that simply deliver a message; they can also express how a speaker feels about it. Their relative importance in achieving intelligibility has led many influential academics to suggest that prosody is much more important than individual vowel clarity, resulting in perhaps the most controversial debate of all in the field of phonetics and phonology (and by extension, of pronunciation teaching): which of the two comes first?

## Starting Small or Thinking Big? The Segmental/Suprasegmental Debate

In Dalton and Seidlhofer's (1994) conceptual model, the *bottom-up approach* gives precedence to separate segments of sound, proceeding from vowels/consonants to suprasegmentals. The other dimension, called *top-down approach*, starts with prosody and works its way down to phonemes. The authors acknowledge the difficulty of choosing between these two approaches, but eventually conclude that

"[intonation is] particularly important in discourse, [at] the same time [it is] particularly difficult to teach. With individual sound segments, it's the other way round: they are relatively easy to teach, but also less important for communication" (page 73).

A modern representative of the 'phoneme first' school of thought, Adrian Underhill's *Sound Foundations* approach (2005) makes extensive use the International Phonetic Association (IPA) phonemic chart, as well as a concept called *proprioception*, "our internal kinesthetic awareness of the position and movement of our muscles and parts of the body" (Underhill, 2012). A valid argument in favour of this approach could be that by paying attention to sensory-motor skills under our conscious control, i.e., the position of tongue, jaw and lips, we can become comfortable more quickly with the 'new' sounds of the target phonological system. Catford (1987), cited in Hagen and Grogan (1992), promotes teaching students "precisely what to with their vocal organs." By closing the eyes, thus blocking out outside distractions, the authors advocate "intensive silent introspection" to achieve this. Kenworthy (1987), cited in Dalton and Seidlhofer (1994:129), however, does not seem to give much credit to this assumption, arguing rather strongly that "receiving directions about what to do with their vocal organs is completely alien to people."

At the other end of the spectrum, many key theorists advocate beginning with suprasegmental aspects of English. "Pronunciation teaching works better if the focus is on larger chunks of speech," says Fraser (2001:17). In a highly regarded publication, *Teaching Pronunciation Using the Prosody Pyramid*, Gilbert (2008) goes as far as saying that teaching sound contrasts, i.e. minimal pair work, is not only tedious, but also discouraging both for the students and their teacher. Emotions aside, starting with segments may also seriously slow down learner progress. As Pennington (1996:19) repeatedly emphasizes,

concentrating on segmentals "can make only piecemeal improvements" but "attention to the prosodic aspects can make [...] improvements to the whole stream of speech" (ibid). Brown (2017:56) strongly agrees, saying that correctly pronouncing individual phonemes "fades into insignificance" compared to the importance of correctly stressing words. The consensus at present within the Communicative Language Teaching framework seems to be the provision of a more balanced teaching methodology, with a strong focus on suprasegmentals, but still incorporating individual sounds. In an apparent attempt to find common ground, Celce-Murcia and colleagues, (2010), cited in Derwing and Munro (2015:9) call this whole debate an

> "artificial [and unproductive] instructional dichotomy [because] in any group of learners there are going to be features from both domains that are problematic for communication and thus should be taught."

## The Chicken or the Egg: Listening or Speaking?

An alternative learning trajectory, not just of pronunciation but of foreign languages in general, sets the problems of production aside and places primary emphasis on perception. Comprehensive four-skill language teaching solutions that are often mentioned as examples are Pimsleur (www.pimsleur.com, audio input in target language which is then translated into the learner's mother tongue) and Rosetta Stone (www.rosettastone.com, audio input with pictures; no translation, full immersion in the target language). Pronunciation-specific approaches often pay homage to Flege's (2003) influential Speech Learning Model (SLM), proposing that a language learner's ability to perceive L2 phonological features directly correlates with their ability to accurately (re)produce them (Martins, Levis & Borges, 2016). Setter and Jenkins (2004:6) concur, saying that one must first be "able to hear a phonemic contrast before one can successfully produce it."

It may be true that many language learners need extensive training until they start noticing features of fluent colloquial speech, but once they realize that certain simplifications and 'distortions,' for example, the dropping of sounds or even of syllables, are the norm rather than the exception, they might decide to imitate these models when they themselves speak. Explicit instruction of what happens in fluent, connected speech – of forms which could very well be considered incorrect in slow, careful English – can aid not just perception but production as well. Noticing, then becoming accustomed to what they hear, learners are more likely to use this information during their own speech production. It is therefore possible that familiarity with fast-paced conversational speech patterns can help a learner improve their production as well, in a process called spill-over, i.e., learning gains in one pronunciation skill leading to improvement in other areas.

Varied input may also be a factor in improving perception. Exposure to different accents, age groups, male and female speakers can potentially open the eyes (and ears) of L2 learners, who soon realize that their teacher's accent is not the only way to pronounce English. Though computer-assisted pronunciation training (CAPT) is a topic that will be discussed in more detail momentarily, the flexibility and effectiveness of High Variability Perception Training (HVPT) merits a brief introduction here. HVPT provides not just an assortment of native speaker models, but also exposes L2 listeners to the allophonic variations of English speech sounds. As Thomson (2011) explains, HVPT was originally designed in 1991 by Logan, Lively and Pisoni, who found that diverse input results in more significant gains in L2 perception. Even though obtaining a multitude of speech samples for listening comprehension and pronunciation training used to be rather problematic in a traditional English L2 classroom, text-to-speech computer technology – as demonstrated, among others, by Qian, Chukharev-Hudilainen and Levis (2018) – as well as the world wide web, provide access to free or low-cost perceptual training resources.

Furthermore, L2 learners also need to be aware, either by inference or through explicit pronunciation instruction, of the difference in intentions of speakers and listeners. Ladefoged (2000:251) points out that for a speaker, the goal is "ease of articulation," "the least possible effort," which entails less attention to phonological detail. A listener, on the other hand, needs to be able to distinguish between sounds that can potentially change message content; therefore, he argues, "a language must always maintain *sufficient perceptual separation*" of its phonemes (italics in the original).

A notable dissenter to this approach is Pennington (1996) who seems to doubt the efficiency of using solely listening comprehension activities to improve pronunciation. Supporters of Pennington's claim include Reed and Michaud (2005:viii) who, under the heading *A Challenge to Conventional Wisdom*, propose that "speech production precedes and facilitates speech perception." In a follow-up study, the authors proposed a closed-circuit loop in which language learners use their own output (production) as input for themselves (perception) to better conceptualize, then pronounce different target language features more clearly (Reed & Michaud, 2011).

## The Nativeness Principle

Levis (2005) identifies two conflicting foci in English pronunciation instruction: accuracy versus intelligibility. Accuracy involves trying to ensure that a language learner's every single speech sound approximates a native-speaker model. With intelligibility a priority, training narrows its focus to elements of spoken English which are critical for successful interpersonal communication. Consequently, a pronunciation teaching curriculum, at its core, most likely adopts one of two positions: that of the *Nativeness* or the

*Intelligibility* principles. The former means aiming at native-speaker competence, while the goal of intelligibility refers to speaking clearly enough to be understood by others.

Perhaps the most important problem with L2 learner goals that strive for native-like pronunciation in English is that it appears to be unrealistic, time-consuming, and ultimately unattainable for the majority of L2 learners (Derwing & Munro, 2005; Field, 2005; Gimson & Cruttenden, 2008; Pennington, 1996). Another just as salient issue is that the precise interpretation of the term 'native-like' is unclear at best. Variability in models manifests itself in several ways. The first and most commonly cited issue is accent based on geographical differences. In English as a foreign language (EFL) contexts, is it British, American, Australian or some other traditional model that is to be adopted? Once a decision has been made, it further complicates matters to decide, in case of British English for example, if it should be Received Pronunciation (RP), or a regional dialect. An advantage of prestige forms, argue Dalton and Seidlhofer (1994), is wider acceptance. In its favour are also the facts that RP enjoys widespread popularity in textbooks and has been extensively researched and described (Roach, 2010:4; Rogerson-Revell, 2011:7). Second, foreign accent and intelligibility are not necessarily interlinked. Derwing and Munro (2015:5) maintain that "a particular utterance could be heavily accented and yet be fully intelligible." This is not necessarily restricted to L2 accents: Roach (2010:6) suggests, quite correctly, that a Scotsman or American talking to an Englishman "may speak with sounds very different from those of his [audience] and yet be clearly intelligible." The third problem with the Nativeness principle is the multitude of pronunciation variables within the target model, including a speaker's age, gender, social class, educational background, occupation, personality, context, purpose (Roach, 2010), as well as a host of idiosyncratic features which, at least from a pronunciation teacher's point of view, are not significant. A potential solution could be to align model boundaries along lines that reflect contemporary, colloquial, yet clear speech phenomena, as well as respond to learner needs. Received Pronunciation, alternatively termed "educated Southern [British] English" by Brown (2017:13), or "BBC pronunciation" by Rogerson-Revell (2011:8) may be adopted, or "North American English" (Celce-Murcia, Brinton & Goodwin, 1996), formerly known as "General American." Options of course include other established varieties, depending on context or the teacher's own accent, as long as this model is consistently observed and presented throughout the training.

## The Intelligibility Principle

The Intelligibility principle neither assumes nor prescribes a native-speaking reference point. A simplistic and very narrow interpretation is through its primary purpose: to enunciate clearly enough to get one's message across. Theoretical linguistics offers

several abstract, sometimes conflicting definitions, some of which will be presented below. Jenkins (2000:78) calls intelligibility in English as an International Language (EIL) "the ability to produce and receive phonological form." Derwing and Munro (2005:385), understand it in terms of "the extent to which the speaker's intended utterance is actually understood by the listener." Furthermore, Underhill (2005:viii) distinguishes between "intelligibility within a local variety of English" – which can mean (i) an L2 learner's clear speech in an L1 community, as in the case of immigration to an English-speaking country, or (ii) the ability to get one's message across with other L2 speakers speaking with the same accent – as opposed to "international mutual intelligibility," which may entail being understood by speakers with different linguistic or cultural backgrounds, including native speakers of English.

His interpretation echoes a distinction made by Gimson and Cruttenden (2008) between what they termed "amalgam English," a mixture of English L1 varieties flavoured by features of a speaker's local language when talking to native speakers, and "international English," used in international contexts between speakers with first languages other than English. These original assumptions about international English, however, no longer hold true. The authors imagined as its settings international business meetings or academic conferences where some variation from English L1 norms was tolerated if these irregular forms did not cause misunderstanding of the relatively predictable message content. English, however, has quickly outgrown these restricted contexts and has become a truly global language. English as International Language, also known as English as a Lingua Franca (ELF), is a framework that recognizes the fact that most interactions in English today take place between L2 speakers; therefore, to ensure mutual intelligibility, certain standardized changes from native-speaking English phonology are to be introduced into English L2 classrooms.

## English as a Lingua Franca

The term English as a Lingua Franca (ELF) was first introduced and most comprehensively presented in Jennifer Jenkins' seminal book, *The Phonology of English as an International Language*, which was first published in 2000. In the introductory chapter, Jenkins argues that misunderstandings are most likely to occur between L2 speakers of English not for grammatical or lexical reasons, but because of differences in their pronunciation. To facilitate "mutual international intelligibility" (page 2) in interactions where no L1 speaker of English is present, she sought to develop a pedagogical framework that re-classified pronunciation variations which were previously considered negatively as deviations from the L1 English norm, i.e., were called incorrect.

Standardization of acceptable forms is critical in ELF. An L2 speaker of English inevitably brings with them traces of their mother tongue (L1); in other words, it influences their accent. Aspects of this L1 transfer can vary significantly depending on a speaker's first language. A conversation between two non-native speakers of English with different linguistic backgrounds can quickly become unintelligible if they are not familiar with each other's accent. The task gets exponentially more difficult as more participants bring in their own 'flavour' of English. Consequently, a set of clearly defined guidelines and teaching practices regarding acceptable variation, called the Lingua Franca Core (LFC), have been proposed to facilitate communicative success. Decisions for inclusion or exclusion of items appear to closely correlate with the concept of the functional load principle. This hypothesis was originally proposed by Catford in 1987, as "a principle to assess the amount of 'work' that phonemic contrasts perform in a language" (cited in Derwing & Munro, 2014). Dalton and Seidlhofer (1994:143) define functional load as "the 'work' individual sounds have to do in the target language in terms of distinctions in lexical meaning and grammar." Sounds or sound pairs which clearly distinguish between a lot of lexical units are considered important, i.e., are said to carry a higher functional load than others. In the ELF initiative, for example, the replacement of the initial phonemes in *think* (voiceless interdental fricative) and *sink* (voiceless sibilant) is not considered a source of error because it does not cause misunderstanding. Even though "I sink you are wrong" may sound a little odd, it is unlikely that a listener will think of a marine catastrophe or kitchen furniture upon hearing it.

Applied linguists and ELT professionals who criticize ELF often agree with its practical use, with many features concentrating on ease of articulation, including a smaller segmental repertoire, but call it "a 'simplified' or 'reduced' form of English" (Jenkins, Cogo & Dewey, 2011:288). This argument is countered by referring to underlying principles of ELF, in which L2 speakers adjust their pronunciation in interactions with other non-native speakers of English (called accommodation), but they do so primarily for maximum efficiency, i.e., mutual intelligibility (ibid). In this light, ELF is seen as a unifying, rather than dividing endeavour.

Perceptions of ELF, its practical use and wider acceptance, or lack thereof, must also take into consideration student goals and motivation. Theories about the distinction between *instrumental* and *integrative* L2 learning motivation may potentially apply to pronunciation instruction as well, propose Dalton and Seidlhofer (1994:11, citing Gardner & Lambert, 1972), The authors draw a parallel between learner drives and ultimate goals. Those for whom English is a tool, a medium (= are instrumentally motivated), will be more interested in getting their message across, without worrying about the minutiae of clarity. Common examples are students studying at international faculties in L2 contexts, where the common language of their multilingual classrooms is English. Very advanced and/or

more ambitious learners on the other hand, will often be more concerned with clarity to make their speech "more acceptable to the [L1] language community [they are] aspiring to be a member of" (ibid). For learners who may wish to sound like a native speaker to communicate with them on an equal footing, the ELF framework may ultimately prove to be too modest.

## At the Chalkface or in Cyberspace?

In addition to debates revolving around the *focus* of pronunciation teaching (segmental/prosodic, production/perception, discrete/integrated, etc.), the question of *locus* is yet another controversial issue. Before the advent of the World Wide Web era with its enormous potential for language education, pronunciation instruction was restricted to the classroom, especially in EFL (English as a foreign language), as opposed to ESL (English as a second language) contexts. Teachers of the former were unable to rely on the L1 environment and community outside the classroom; therefore, they were forced to work in isolation, often under artificial conditions. Despite efforts to implement out-of-class pronunciation instruction, native-speaking input and useful resources were limited.

The Internet, however, offers unlimited access to language-learning opportunities, bringing with it innovative, computer-based pronunciation teaching methodologies. High Variability Perception Training (HVPT), online resources like the Speech Accent Archive (www.accent.gmu.edu), or English Accent Coach (www.englishaccentcoach.com), an interactive, game-based website are all pedagogically sound resources that focus on improving pronunciation through listening. Efforts continue to further develop Automatic Speech Recognition (ASR) software, not only for education, but a host of other, everyday purposes. *Siri*, Apple Inc's personal assistant or *Google Now* are two ubiquitous examples that offer instant, constant feedback to English L2 learners about their pronunciation.

## Does Pronunciation Instruction Work?

Few people would argue that achieving clear, intelligible English pronunciation is a major endeavour. It requires patience, constant practice and unwavering dedication, as progress tends to be slow and learners must be constantly encouraged not to lose heart when improvements take time to manifest. This inherent difficulty has led to the often criticised assumption that pronunciation teaching is not very effective (Ferreiro & Luchini, 2015), or that it would, with time and sufficient exposure, improve by itself (Thomson & Derwing, 2014).

There is, however, a growing body of research that has consistently proven that explicit pronunciation instruction does in fact lead to improvements in clarity of speech.

Focusing on segmentals (Saito, 2015; Thomson, 2012) and suprasegmentals (Pearson, Pickering & Da Silva, 2011), with time periods ranging from a total of 4 hours (Muller Levis & Levis, 2012; Silveira, 2013) to complete college semesters (Henrichsen & Stephens, 2015), incorporating digital technology (Mompean & Fouz-Gonzáles, 2016; Thomson & Derwing, 2016), it has been found that both classroom and laboratory-based instruction leads to not only long-term retention (Young & Wang, 2014), but can also successfully instil learning strategies for future self-study (Henrichsen & Stephens, 2015; Pearson, Pickering & Da Silva, 2011).

## Conclusion

The primary purpose of this paper is to offer readers a review and critical analysis of current differences of opinion in the field of English L2 pronunciation research and instruction. It attempts to provide an in-depth investigation of a limited number of issues, rather than a comprehensive survey of pronunciation teaching *per se*, and admits sacrificing breadth for the sake of depth. It is hoped, however, that this selection of ongoing disputes, including variability in how to interpret terminology, which model or teaching methodology to follow, different contexts and delivery methods, and finally, recurring doubts about effectiveness, has touched upon the most salient issues of disagreement among English L2 pronunciation professionals.

Further unintended consequences of this apparent diversity of opinion include a lack of guidance for classroom instructors, who may be overwhelmed or confused by the abundance of strongly held, and convincingly argued beliefs discussed here. Judy Gilbert, who is considered by many to be not only the pioneer, but also *the* iconic figure of the 'suprasegmentals first' approach, found during field tests for a new edition of her book, *Clear Speech* that students specifically requested individual sound segments first (cited in Dalton & Seidlhofer, 1994:143). A defining characteristic of a good educator, when faced with student requests that at first seem incompatible with their teaching philosophy is to adapt, to adjust their curriculum and methodology. There is a proverb, "All roads lead to Rome," a reminder that one can achieve the same end result regardless of the methods used. Perhaps by experimenting with a less familiar approach that is outside one's comfort zone, a teacher not only follows learner-centred principles, but also contributes to their own professional development. To paraphrase Robert Frost, taking the road less travelled by might eventually turn out to have made all the difference.

## Acknowledgments

The author would like to gratefully acknowledge the contributions of two anonymous reviewers and Dr. Pamela Rogerson-Revell, and to thank them for helpful critical insights and constructive criticism on an earlier draft of this article.

## References

Ashby, P. (2011). *Understanding Phonetics*. New York: Routledge.

Brown, A. (2015). Syllable structure. In Reed, M., & Levis, J. M. (Eds.), *The handbook of English pronunciation*, pp. 85-105. Chichester: John Wiley and Sons.

Brown, G. (2017). *Listening to Spoken English* (2nd ed.). Harlow: Longman.

Büring, D. (2015). *Intonation and meaning*. Oxford: Oxford University Press.

Cauldwell, R. (2013). *Phonology for listening: Relishing the messy*. Downloaded from https://www.academia.edu/568610/Phonology_for_listening_relishing_the_messy. Last accessed 08 March 2018.

Cauldwell, R. (2014). Listening and pronunciation need separate models of speech. In J. Levis & S. McCrocklin (Eds.), *Proceedings of the 5th Pronunciation in Second Language Learning Teaching Conference* (pp. 40-44). Ames, IA: Iowa State University.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.

Crystal, D. (1986). *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.

Cutler, A. (2015). Lexical stress in English pronunciation. In Reed, M., & Levis, J. M. (Eds.), *The handbook of English pronunciation*, pp. 106-124. Chichester: John Wiley and Sons.

Dalton, C., & Seidlhofer, B. (1994). *Pronunciation*. Oxford: Oxford University Press.

Davenport, M., & Hannahs, S. J. (2010). *Introducing phonetics and phonology* (3rd ed.). London: Hodder Education.

Derwing, T. M. & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly 39*(3), 379-397.

Derwing, T. M. & Munro, M. J. (2014). Once you have been speaking a second language for years, it's too late to change your pronunciation. In Grant, L. (Ed.), *Pronunciation myths: Applying second language research to classroom teaching*. Ann Arbor: University of Michigan Press.

Derwing, T. M. & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins.

Ferreiro, G. M., & Luchini P. L. (2015). Redirecting goals for pronunciation teaching: A new proposal for adult Spanish-L1 learners of English. *International Journal of Language Studies 9*(2), 49-68.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*(3), 399-423.

Flege, J. E. (2003). A method for assessing the perception of vowels in a second language. In Fava, E., & Mioni, A. (Eds.), *Issues in clinical linguistics*, pp. 19-44. Padova: Unipress.

Fraser, H. (2001). *Teaching pronunciation: A handbook for teachers and trainers. Three frameworks for an integrated approach*. Sydney, Australia: Department of Education Training and Youth Affairs (DETYA).

Gilbert, J. B. (2008). *Teaching pronunciation using the prosody pyramid*. Cambridge: Cambridge University Press.

Gimson, A. C., & Cruttenden, A. (2008). *Gimson's pronunciation of English* (7th ed.). London: Hodder Education.

Hagen, S. A., & Grogan, P. E. (1992). *Sound advantage: A pronunciation book*. Upper Saddle River: Prentice Hall Regents.

Henrichsen, L. & Stephens, C. (2015). Advanced adult ELS students' perspectives on the benefits of pronunciation instruction. In J. Levis, R. Mohammed, M. Qian, & Z. Zhou (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning Teaching Conference* (pp. 197-205). Ames, IA: Iowa State University.

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

Jenkins, J., Cogo, A., & Dewey, M. (2011). Review of developments in research into English as a lingua franca. *Language Teaching, 44*(3), 281-315.

Ladefoged, P. (2000). *A course in phonetics* (4th ed.). Boston: Thomson Wadsworth.

Levis, J. M. (1999). Intonation in theory and practice: Revisited. *TESOL Quarterly, 33*(1), 37–63.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3), 369-377.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874-886.

Martins, C. G. F. M., Levis, J. M., & Borges, V. M. C. (2016). The design of an instrument to evaluate software for EFL/ESL pronunciation teaching. Available at

https://www.academia.edu/21086736/The_design_of_an_instrument_to_evaluate_software_for_EFL_ESL_pronunciation_teaching. Last accessed on 9 March 2018.

Mompean, J. A., & Fouz-Gonzáles, J. (2016). Twitter-based EFL pronunciation instruction. *Language Learning & Technology 20*(1), 166-190.

Muller Levis, G., & Levis, J. (2012). Learning to produce contrastive focus: A study of advanced learners of English. In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference,* (pp. 124-133). Ames, IA: Iowa State University.

Pearson, P., Pickering, L., & Da Silva, R. (2011). The impact of computer assisted pronunciation training on the improvement of Vietnamese learner production of English syllable margins. In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference* (pp. 169-180), Ames, IA: Iowa State University.

Pennington, M. C. (1996). *Phonology in English language teaching: An international approach*. Harlow: Pearson Education Ltd.

Qian, M., Chukharev-Hudilainen, E., & Levis, J. M. (2018). A system for adaptive high variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning & Technology 22*(1), 69-96.

Reed, M., & Michaud, C. (2005). *Sound concepts: An integrated pronunciation course*. New York: McGraw-Hill.

Reed, M., & Michaud, C. (2011). An integrated approach to pronunciation: Listening comprehension and intelligibility in theory and practice. In Levis, J. & LeVelle, K. (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning Teaching Conference* (pp. 95-104). Ames, IA: Iowa State University.

Roach, P. (2010). *English phonetics and phonology: A practical course* (4th ed.). Cambridge: Cambridge University Press.

Rogerson-Revell, P. M. (2011). *English phonology and pronunciation teaching*. London: Continuum.

Saito, K. (2015). Communicative focus on second language phonetic form: Teaching Japanese learners to perceive and produce English /r/ without explicit instruction. *Applied Psycholinguistics 36*, 377-409.

Setter, J., & Jenkins, J. (2004). State-of-the-Art Review Article. *Language Teaching*, *38*. 1-17.

Silveira, R. (2013). Pronunciation instruction and syllabic-pattern discrimination. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 147-156). Ames, IA: Iowa State University.

Thomson, R. I. (2011) Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal 28*(3), 744-765.

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning 62*(4), 1231-1258.

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics* 2014:1-20. Available at https://www.researchgate.net/publication/271827649_The_Effectiveness_of_L2_Pronunciation_Instruction_A_Narrative_Review. Last accessed on 9 March 2018.

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le., I. Lucic, E. Simpson, & S. Vo (Eds.). *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference* (pp. 88-97). Ames, IA: Iowa State University.

Underhill, A. (2005). *Sound foundations: Learning and teaching pronunciation* (2nd ed.). London: Macmillan Education.

Underhill, A. (2012). *Proprioception and pronunciation*. Downloaded from http://www.adrianunderhill.com/2012/08/28/proprioception-and-pronunciation/. Last accessed 8 March 2018.

Wells, J. C. (2008). *Longman pronunciation dictionary* (3rd ed.). Harlow: Pearson Education Ltd.

Young, S. S., & Wang, Y. (2014). The game embedded CALL system to facilitate English vocabulary acquisition and pronunciation. *Journal of Educational Technology & Society 17*(3), 239-251