



Evaluating English Teachers' Pronunciation Skills (E-TPS) Using Praat Technology and Human Raters

Mas Darul Ihsan^{1,2} & Ahmad Munir¹

¹Universitas Negeri Surabaya, Indonesia

²Universitas Gresik, Indonesia

Abstract: *When the goals in teaching pronunciation are that students attain “near-native” and “socially acceptable” pronunciation, and are assessed with a valid and reliable instrument using the Praat device and human interpretation raters, a new way to assess pronunciation skills is needed. This article proposes such a new way. Through suprasegmental identity (rhythm), English teachers’ pronunciation skill (E-TPS) will be correctly assessed. Using Praat as a device and human raters for interpretation to assess E-TPS stands on several premises. First, the measures of pronunciation constructs should be valid and reliable. Second, the role of raters in English pronunciation assessment presents unique challenges in drawing valid inferences from performance, in scoring assignments, in making decisions for which a pronunciation assessment was intended, to the pedagogical and social consequences beyond, and to connect concerns in giving judgments. Third, the raters should have familiarity with both accent and content. Fourth, a suprasegmental (stress, rhythm, and intonation) approach is an important complement to segmental teaching. Fifth, rhythm plays a big role and is defined as a “continuum” of functions and effects in pronunciation assessment. Sixth, the trend among the technology-minded for testing using automated assessment of pronunciation has as its goal accuracy in imitating native speakers, as is reflected in the use of Praat as a device. Therefore, the blending approach (Praat and raters) represents the future in assessing E-TPS.*

Keywords: *Praat; human interpretation; rater; pronunciation assessment; English teachers’ pronunciation skill (E-TPS)*

INTRODUCTION

Assessing pronunciation among English language learners is commonly part of the assessment of speaking skills. This is evident in Brown and Abeywickrama’s (2019) imitative speaking assessment which includes word and sentence repetition. In the rubric of such assessments, you can find “acceptable pronunciation; comprehensible, partially correct pronunciation; and silence, seriously incorrect pronunciation” (p.161). In practice, however, assessing pronunciation is based on a simple series of speaking tasks to score pronunciation of L2 speakers against L1 speaker norms. These include sounds produced in segmental identities, such as simple sounds – i.e., consonants and vowels – and suprasegmental identities such as intonation, pitch, tempo, and rhythm, which

extend beyond a single vowel or consonant to syllables, words, and entire sentences (Chun, 2002).

English teacher education students in non-English speaking countries are also English language learners. Their education curriculum usually contains speaking skills as well as pronunciation skills. In the context of English teachers' pronunciation skill (E-TPS) in these countries, they will also be assessed for their ability to produce segmental identity which reflects the phonetics – the “sounds” of the language – and how to recognise and reproduce the consonant sounds and vowel sounds of English. In addition, they will also be assessed on their skills in producing suprasegmental identity and intonation. What the E-TPS produce may not be the same as what they think they are saying. The rhythm of speech is combined in turn with stretches of silence (pauses) to break up the flow of speech (Wells, 2009).

To help English teachers achieve these goals, practising rhythmic patterning in pronouncing English at all levels is important. Moreover, suprasegmental identity typically extends over more than one sound segment in an utterance, over longer stretches of speech. Rhythm plays a big role and is defined as a “continuum” of functions and effects, ranging from the non-linguistic or extralinguistic at one end, through the paralinguistic, to the essentially linguistic (Chun, 2002). Furthermore, pronunciation assessment tends to focus on standard varieties of English. This is because varieties tend to lack the sort of accepted codification typically required by applied linguists as a foundation, whether in building resources in language learning, language teaching, or speech technology, or in a particular area of interest, such as clinical and development assessments of typical phonological acquisition (Thomas & Scobbie, 2015). In the context of speech technology, Praat is the proper device to be applied (Boersma et al., 2007).

For the last five to ten years, assessing E-TPS has been limited to only conventional methods (Mairano & Santiago, 2020; Monfared, 2018; Zhang et al., 2017). These mainly focused on human interpretation and possibly a lack of truthfulness in assessing a certain skill, especially the level of rhythm for suprasegmental identity. Furthermore, from a different direction, the device used to assess pronunciation skills has also been used as only a single way to assess (Behr, 2022; Nagle et al., 2020; Rozaimie, 2018; Ryan & Ryuji, 2021; Wang, 2019; Young & Shishido, 2022), which is full of truthfulness, but the substance is not portrayed because the device is incapable of changing the perspective that assesses the E-TPS. Therefore, this paper proposes a new way of assessing E-TPS by using Praat as a device, with human interpretation as a rater. Such a combination is expected to reorient the pedagogical goal in pronunciation teaching from the traditional focus on accent reduction, to investigating global measures of prosody and fluency, reliable automated comprehensibility assessments, and the effect of mobile-based aural comprehension lessons (Isaacs, 2018; Kallio et al., 2023; Saito et al., 2022; Woldetsadik et al., 2022).

THE URGENT NEED FOR ASSESSING ENGLISH TEACHERS' PRONUNCIATION SKILL (E-TPS)

It is vital that speakers of English, whether they are native or non-native speakers, are able to exchange meaning effectively. This is reasonable; as Burns and Claire (2003) observed, English has increasingly become the language used for international communication (Baratta, 2019; Nunan, 2023; Seargeant, 2016; Starfield, 2016). In fact, in the most recent discussions of English-language teaching, the unrealistic idea that learners should sound and speak like native speakers is fast disappearing. The recent development in E-TPS suggests that it is more important that speakers of English can achieve intelligibility (the speaker produces sound patterns that are recognisable as English) and comprehensibility (the listener is able to understand the meaning of what is said), and interpretability (the listener is able to understand the purpose of what is said). For example, a speaker might say It's hot today as IS ho day. This is unlikely to be intelligible because of inaccurate sound, stress and intonation patterns. For a temporary reason, a listener would not find the speaker comprehensible, because meaning is not available. Because the speaker is incomprehensible, the listener would also not be able to interpret the utterance as an indirect request to open the window.

In the context of suprasegmental identity, if the goals in teaching pronunciation are that students attain "near-native" and "socially acceptable" pronunciation, however, then teaching suprasegmental (stress, rhythm, and intonation) will be an important complement to the teaching of segmental – the consonants and vowels of old-style pronunciation teaching (Chun, 2002). Figure 1 shows features of English pronunciation which include segmental and suprasegmental features.

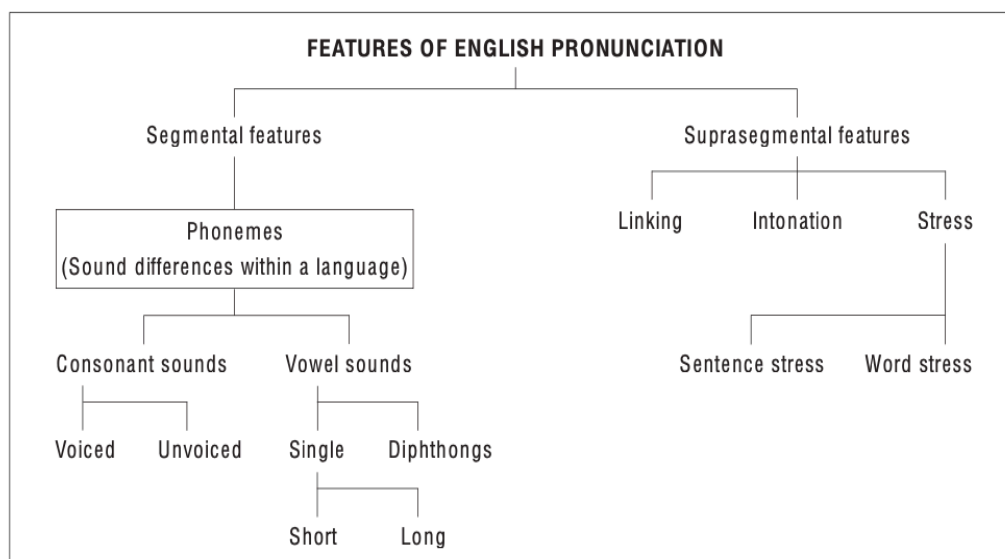


Figure 1. The illustration of pronunciation skill features, as adapted from Burns and Claire (2003)

Segmental identities were the major focus for pronunciation teaching (for example, minimal pairs such as ship/sheep). While these identities are important, more recent research has shown that when teaching focuses on suprasegmental identities, learners' intelligibility is greatly enhanced. It is important, therefore, to provide activities at both levels. Burns and Claire (2003) explained the aspects of pronunciation skill, where suprasegmental identities relate to sounds at the macro level. Advances in research have developed descriptions of the suprasegmental identities of speech extending across whole stretches of language (prosody). Unlike languages such as Vietnamese or Mandarin which are tonal, English is stress-timed and syllable-timed (for example, WHAT's his address?).

Linking, intonation and stress are important features for effective pronunciation at the suprasegmental level. (1) Linking refers to the way the last sound of one word is joined to the first sound of the next word. To produce connected speech, we run words together to link consonant to vowel, consonant to consonant, and vowel to vowel. We also shorten some sounds and leave others out altogether. (2) Intonation can be thought of as the melody of the language – the way the voice goes up and down according to the context and meanings of the communication. (3) Word stress relates to the prominence given to certain words in an utterance. These focus words are stressed (made long and loud) to convey: (a) the overall rhythm of the utterance, and (b) the most meaningful part of the utterance. At the meaning level, some words are given more prominence than others to foreground which meaning is important. Segmental features relate to sounds at the micro level. They include specific sounds within words (for example, l as in lamp, r as in ramp, a as in hat). The sound systems of consonants, vowels or their combinations are called phonemes. Phonemes are sounds that, when pronounced incorrectly, can change the meaning of the word.

To make the application of suprasegmental in assessing pronunciation some adaptation examples in the classroom context can be taken. Norton (2003) believed that focusing on suprasegmental identities of pronunciation enables students to improve their overall intelligibility from an early stage. Consequently, Norton did not see pronunciation activities as stand-alone, but rather as an integral part of lessons. Norton (2003) offered some steps in practising pronunciation assessment, as shown below:

- Step 1. Referring to the board, Norton reminded the learners of the terms pronunciation and stress. Norton (2003) confirmed that stress refers to syllable sounds that are “long and loud” or “the important words”. Norton wrote the sentence What's his address? on the board, and the class identified the number of syllables (four) and which ones were stressed. The learners practised this pattern by repeating it.
- Step 2. Norton (2003) showed the learners green cards with short questions written on them and orange cards marked with dotted stress patterns. Norton first demonstrated the activity, using the question How old is your mother? Norton asked three learners holding orange cards to pronounce their stress patterns to see if they corresponded with his question card: How old is your mother? The third learner Norton asked had a stress pattern matching his question.

- Step 3. Norton (2003) divided the class into two and distributed green or orange cards to the two groups. After individually practising their questions (green cards) or stress patterns (orange cards), the learners circulated to find partners.
- Step 4. When most learners appeared to have found partners, Norton (2003) assembled them in a circle. They placed their cards on the floor and together checked which ones matched. Where learners could not find a match or their match was incorrect, Norton (2003) helped them to find their partners.
- Step 5. Norton concluded the sequence with a class drill of the sentences used in the activity.

The above ideas on applying suprasegmental pronunciation show the urgent need for assessing E-TPS. The first point was for learners to be able to acquire native or near-native pronunciation proficiency in the sound patterns of new languages, and they had not entirely lost their ability to perceive and produce novel sounds. However, second, subjects were given neither rules of the language nor the meaning of the target utterances. It is thus quite possible that the subjects were processing only the phonetic material and were therefore imitating sounds without attaching any meaning to them or using them to communicate (Chun, 2002). E-TPS assessment was seen as an urgent need and emerged from its time wrap.

Isaacs and Trofimovich (2017) summarised the chronological development of E-TPS, as presented in Table 1.

Table 1. *Time wrap of pronunciation assessment urgency*

Pronunciation-focused articles were published in the longest-standing language assessment journal, *Language Testing*. During its first 25 years of publication, seven articles appeared in the five years from 2010-2015.

Pronunciation assessment has also been featured in major events targeting the L2 speaking construct (e.g., the 2013 Cambridge Centenary Speaking Symposium).

In at least four externally funded TOEFL and IELTS research projects since 2010, pronunciation was a topic hitherto rarely focused on in the validation of high-stakes tests. This implies that pronunciation is increasingly being viewed as integral to the L2 speaking construct.

Pronunciation from the language assessment community at large has been seen in the introduction of fully automated standardised L2 speaking tests (e.g., Pearson's Versant test and Educational Testing Services' SpeechRater).

In some cases, pronunciation assessment has focused on the accuracy of segmental; in others, on the approximation or the mastery of suprasegmental. Focused on suprasegmental identities are those that occur beyond the production of the phonic itself (e.g., pauses, intonation, and stress). These identities have been examined in recent studies on L2 oral production and proficiency (Ghanem & Kang, 2018). In the same sense, Moere and Suzuki; Mompean and Gonzales (2018; 2015) developed the notion that suprasegmental identities included aspects

of speech such as intonation, word and syllable stress, rhythm and sentence-level stress, and speed. To focus on rhythm, the most general definition of rhythm is the perceived regularity of prominent units in speech. Although these regularities are sometimes described in terms of patterns of syllable length (long vs. short) or pitch (high vs. low), rhythm is most commonly discussed in terms of patterns of stressed vs. unstressed syllables (Chun, 2002). Moreover, Cummins and Port (1998) define rhythm in speech as the hierarchical organisation of temporally coordinated prosodic units.

The applied linguistics community has been assessing E-TPS by bringing together insights and highlighting pedagogical and assessment implications. This is of relevance to language assessment researchers and practitioners on a common platform, responding to the urgent need. A pronunciation assessment is often high-stakes. This is particularly the case for situations where the assessment might determine the legitimacy of the claims of an asylum seeker or in cases where the assessment determines whether an international teaching assistant is allowed to work while undertaking graduate studies. In assessing and teaching a second language (L2), pronunciation is therefore clearly important, but topic discussion in this area is still limited and has had only a recent awakening (Isaacs & Trofimovich, 2017).

THE IDEAL PERSPECTIVE OF PRAAT AS A DEVICE AND HUMAN INTERPRETATION AS RATERS: METHODOLOGY

One of the critical issues involved in assessing pronunciation is determining to what extent the measures of pronunciation constructs are valid and reliable (Kang & Ginther, 2018). The rater operates in a context where English pronunciation assessment presents unique challenges in drawing valid inferences from performance, in scoring assignments, and in the ultimate decisions for which a pronunciation assessment was intended, while conscious of the pedagogical and social consequences beyond. The question will be: is the assessment fair? While such inquiries can be generally perceived across all types of language assessments, some of them are central to the concept of validity in English pronunciation (Harding, 2018). Moreover, Yan and Ginther (2018) proposed that in operational assessment contexts, it is important to consider how rater interactions with English accents may introduce construct irrelevant variance into the assessment domain. Common research methods into rater interpretation – such as examined reliability, rater bias, and the effect of training on rater consistency – further explored the usability of scale criteria and rater experience with rubrics (Isaacs & Harding, 2017). According to the rater certification manual, raters are not required to be native speakers (Dimova, 2018).

Isaacs and Harding (2017) proposed that human raters could now be supplanted through the use of modern technology, which addresses the issue of human behavioural variability. However, machine scoring of speech is not without limitations, with automated scoring systems, as yet, able only to robustly approximate human judgments on highly controlled L2 speaking tasks that yield predictable learner output (e.g. sentence read-aloud, construction, or repetition tasks). Regarding the role of the rater in assessing E-TPS, human interpretation

should provide the link between concerns in giving judgments and scoring, at least, as adapted from (Dimova, 2018):

Table 2. *Characteristics of English speech ratings*

Familiarity with the accent	<ol style="list-style-type: none"> 1. Shared or similar L1/L2 backgrounds, exposure to a particular accent 2. When a listener hears something that does not meet his or her expectations (e.g., the speaker speaking in a foreign accent), s/he may notice the phonological deviation and thus perceive a speaker as less intelligible. 3. If a listener shares the same or similar language background with the speaker, he or she will be familiar with the speaker's accent and thus may perceive the speaker as more intelligible than speakers of other accents.
Familiarity with content	<ol style="list-style-type: none"> 1. Shared academic interests or background knowledge 2. In English content classrooms, familiarity with the subject content tends to help students comprehend unintelligible speech. 3. Familiarity with speech topics, based on prior interactions or world knowledge, facilitated native speakers' comprehension of non-native speech more than familiarity with the accent or the speaker.

To refine measures of pronunciation assessment test development is very important (Pennington & Rogerson-Revell, 2019), because if their ratings of pronunciation do not match those of other accepted ways of assessing it, then their consequential validity is at issue in the important sense that the testing of pronunciation has become big business with significant consequences for people's lives and employment. This puts the role of the human rater into the working sphere. Assessment of pronunciation by humans may be carried out by naive raters who have no relevant specialised knowledge or experience, other than their native speaker status, or by expert raters who have relevant specialised knowledge and experience. The logic behind using naive native speaker raters is that they would be expected to assess pronunciation in the same intuitive way any native speaker would. The logic behind using expert raters is that they would have skills relevant to pronunciation assessment and so be more likely to assess in a detailed way, according to multiple criteria. Expert raters may be selected for expertise in linguistic phonetics and phonology and/or for their knowledge and experience of non-native pronunciation or the specific second language(s) of test takers. They are sometimes selected specifically for their experience in rating pronunciation by a specific measurement instrument or type of rating scale or task.

Proposing high inter-rater reliability (Pennington & Rogerson-Revell, 2019) is intended to ensure high reliability of individual raters' scores, and also high reliability across raters. Inter-rater reliability itself suggests the extent of the arrangement of interval subjective ratings by multiple raters, inspectors, judges, or appraisers. Moreover, it refers to the degree of agreement when a measurement is repeated under identical conditions by different raters. In systematic review, it can

be used to evaluate agreement between authors in the process of extracting data. At the same time, there has been a variety of methods to measure inter-rater reliability (Park & Kim, 2015).

Furthermore, there is a need for pronunciation assessment development among software developers as well as language teachers to expand their horizons beyond teaching phonology as full mastery of a target language. Yet the teaching of pronunciation, including that with pronunciation technology, is still often based on minimal pair activities or articulation work, with no curriculum to move the learner beyond just an accurate goal of imitating a native speaker. In technology, there is room for improvements in speech recognition systems and especially in how these are used in teaching. Therefore, Pennington (2015) proposed the idea in Table 3.

Table 3. *Trends in pronunciation assessment development*

Trends in theory	Enriched models of speaking proficiency; changing views of nativeness; multilingual/plurilingual models of language
Trends in research	More data-driven research and theory; more attention to phonetic levels; more attention to variable phonology in multilingualism/plurilingualism and specific contexts
Trends in teaching	Focus on form; strategic competence; computer-assisted pronunciation (CAP) and pronunciation for specific purposes (PSP)
Trends in testing	Automated assessment of pronunciation; complex multifactor models; testing in PSP

Therefore, automated speaking or pronunciation assessment systems are continually being improved in order to increase the correlation between their ratings and those of human raters, an important consideration for both their criterion-related and face validity that is also central to building an argument about whether automated assessment matches the reality of human judgment of L2 speaking competence (Pennington & Rogerson-Revell, 2019). Then, to standardise the test, a wide variety of oral proficiency tests and scales has been used, such as (1) American Council on the Teaching of Foreign Languages (ACTFL) oral proficiency interview and proficiency guidelines, (2) the Interagency Language Roundtable (ILR) scale, (3) the Common European Framework of Reference (CEFR) language proficiency scale, (4) the International English Language Testing System (IELTS) speaking exam, and (5) Cambridge English: Advanced (CAE) exam. Focusing on CEFR in detail, below are the criteria.

Table 4. *The CEFR phonological control scale, and extract from the CEFR phonological control descriptor scale*

The CEFR Phonological Control Scale	Explanation
C1	Can vary intonation and place sentence stress correctly in order to express finer shades of meaning.
B2	Has acquired a clear, natural, pronunciation and intonation.
B1	Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur.
A2	Pronunciation is generally clear enough to be understood despite a noticeable foreign accent, but conversational partners will need to ask for repetition from time to time.
A1	Pronunciation of a very limited repertoire of learnt words and phrases; can be understood with some effort by native speakers used to dealing with speakers of his/her language group.

To explain, as adapted from Pennington and Rogerson-Revell (2019), mainly on C2 and A1 in a context of overall phonological control, sound articulation and prosodic features: for C2, overall phonological control reflects an employment of the full range of phonological features in the target language with a high level of control, including prosodic features such as word and sentence stress, rhythm and intonation, so that the finer points of his/her message are clear and precise. Intelligibility and effective conveyance of and enhancement of meaning are not affected in any way by features of accent that may be retained from other language(s). Then, articulation of virtually all the sounds of the target language with clarity and precision. While for prosodic features, subject exploits prosodic features (e.g. stress, rhythm and intonation) appropriately and effectively in order to convey finer shades of meaning (e.g. to differentiate and emphasise).

For A1, in a context of overall phonological control, sound articulation and prosodic features: overall phonological control reflects pronunciation of a very limited repertoire of learnt words and phrases which can be understood with some effort by interlocutors used to dealing with speakers of the language group concerned. Can reproduce correctly a limited range of sounds as well as the stress on simple, familiar words and phrases. Then, sound articulation reproduces sounds in the target language if carefully guided. The subject can articulate a limited number of sounds, so that speech is intelligible only if the interlocutor provides support (e.g. by repeating correctly and by eliciting repetition of new sounds). Where prosodic features are used with a limited repertoire of simple words and phrases intelligibly, in spite of a very strong influence on stress, rhythm, and/or intonation from other language(s) he/she speaks; his/her interlocutor needs to be collaborative.

More specifically, automated assessment of pronunciation is technology-inclined and can be used to: (1) provide learners with visualisations of their intonational patterns, with specific feedback to help them perceive the meaningful contrasts between L1 and L2, so that they can improve their speech production; (2) provide learners with authentic and extensive speech and cultural input and in turn to hone learners' perceptual abilities; (3) facilitate, record, and analyse interactions between and among speakers; (4) build devices for research purposes, e.g., data collection devices to record student performance, progress, and steps toward self-correction (Chun, 2002). Trends in automated assessment of pronunciation are based on complex multi-factor models such as those in development by De Jong (2018) and Isaacs (2018).

Praat (Dutch for “talk”) is a free computer software package for the scientific analysis of speech and phonetics (Stanley & Lipani, 2019). The ideal perspective on which to base Praat as a practical device is to assess suprasegmental identity which is then rated using human interpretation of the score generated by the device. Praat itself is a device for acoustic analysis and transcription of speech (Boersma et al., 2007).

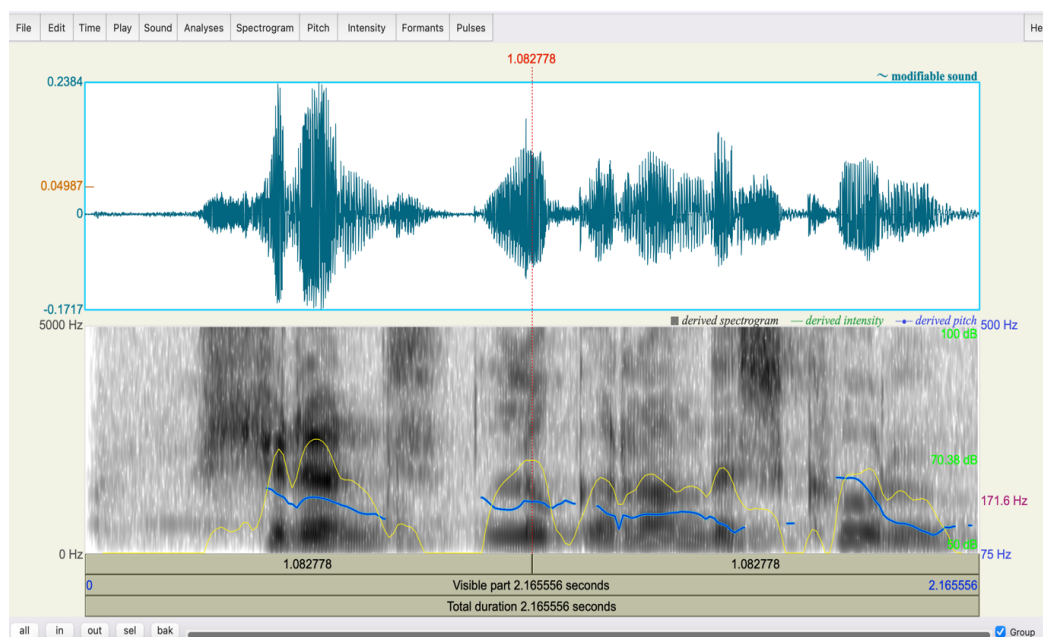


Figure 2. Images from Praat

Stanley and Lipani (2019) offered more specifics on how to use Praat. It is a very flexible tool for speech analysis. It offers a wide range of standard and non-standard procedures, including spectrographic analysis, articulatory synthesis, and neural networks. To be more specific, it can do speech synthesis such as pitch, formant, intensity, and articulatory synthesis (Boersma et al., 2007; Lieshout, 2017; Styler, 2023). Praat measures not only an acoustic speech signal that can be seen as a source signal (the glottal source, or noise generated at a constriction in the vocal tract), filtered with the resonances in the cavities of the vocal tract downstream from the glottis or the constriction, but also can create a source signal from scratch or from an existing speech signal, create a filter from scratch, or

extract it from an existing speech signal, and manipulate (change, adapt) the signal by doing the actual synthesis, which combines the two.

For its functionality, as part of an acoustic evaluation of speech and voice samples, it covers elements such as: finding information in the manual, creating a speech object, processing a signal, labelling a waveform, general analysis (waveform, intensity, sonogram, pitch, duration), spectrographic analysis, intensity analysis, pitch analysis and using long sound files (Lieshout, 2017). Praat is a highly sophisticated piece of software under the hood. There are some definite advantages in using Praat (Stanley and Lipani, 2019): (1) it is the standard; nearly every linguist has used Praat at some point. This is especially true of phoneticians, phonologists, and sociolinguists; (2) while the software itself may not be pretty, the visualisations it can produce are of professional quality; (3) it actually has its own scripting language, to help automate any of its functions. This is especially useful if you have a task that you need to do over and over. Say you have a recording half an hour long and you want to extract the duration of all the vowels, a script will take care of that in seconds. But Praat scripting is even more of a black box than with regular software, with even fewer tutorials online; (4) It is free of charge; (5) the operation is more convenient, and the graphic information feedback is clear (Yang & Zhao, 2021); in the same sense, (6) (Osatananda & Thinchana, 2021) the highlight of Praat is that it is a freeware program, which can be downloaded instantly. Users have freedom in selecting authentic voice models, unlike with most commercial programs which are designed to equip users only with built-in speech files; (7) Praat shows clear graphic movement; the lines are quite straightforward and easy to understand. The pitch movement, rising and falling, is easy to interpret and practice accordingly; (8) it is easy to find within the screen display where the difference in stress, pitch, and intonation are shown, thus learners can repeatedly practice isolated words.

The next benefit to use Praat (Maryn, 2017): (9) many relevant markers (i.e., acoustic voice measures related to fundamental frequency, sound level, formant, perturbation, spectral configuration, cepstral configuration, etc.) are readily available in Praat; (10) there are numerous Praat scripts for automated analysis of the acoustic voice signal, providing single-button prompts for sometimes particularly complex procedures that otherwise would be far too time- and labour-consuming to apply; (11) applying a script increases Praat's user-friendliness and induces standardisation and consistency in analysis methods.

From the intensive discussion above, the ideal perspective on Praat as a device with human interpretation from raters in a focused-on methodology is applicable to assessment pronunciation skill for English teachers. In a classroom context, Praat should be as a device to guide the level of acceptance for teachers and students. Its urgency is to avoid losing meaning in an interaction context – about what is said through what is implied. As imagined, any meaning from the utterance sounds delivered by the teachers, then, carries a misperception accepted by the students. The lack of a concept of perception and production makes Praat an applicable device to be engaged. The speech recognition will function on clearly applicable messages, or at least attach meaning to communication. The expression or utterances may vary in the classrooms; however, these could be in

the form of statements, wh-questions, and commands. Such forms have the default tone level or accepted tone implied. If the utterances or sentences pronounced by the teachers or the students stray far from the default, as suggested, then the meaning will be unpredicted, and it is the role of Praat to measure them.

For example, because of the mis-intonation of a certain utterances pronounced, like the command below.

- Teacher (Mr): *Students, open your book on page 23.*
 (Pronounced by using a falling tone)
- Student 1: *Yes, sir*
 (Pronounced by using a falling tone)
- Student 2: *Yes, sir*
 (Pronounced by using a rising tone)

Student 1 implies the meaning of the terminality of the conversation proposed, no more additional conversation. However, student 2 implies the meaning that s/he is attentive to the teacher continuing to speak.

ENGLISH TEACHERS' PRONUNCIATION SKILL (E-TPS): APPROACHES TO TEACHING SPEAKING

Setting up contexts for students to speak in the English classroom is not the same as teaching students in a second language. Teaching is in principle a systematic activity. To teach speaking effectively, teachers require a set of theoretical and pedagogical principles that can be applied to the planning and delivery of lessons on speaking (C. C. M. Goh & Burns, 2012). Pedagogical principles which shape such approaches are fully understood as judging how far the pedagogy can be accommodated in a different cultural context. Out of such an accommodation will come new teaching approaches with a greater chance of them being implemented in the classroom – focused on the pedagogical function of the target language in speaking, in a pedagogically and theoretically principled approach to target and first-language use. Further, that second-language learning recognises pedagogical principles from the first-language use to benefit second-language learning.

As an approach to teaching speaking, an approach itself refers to language assessment in which test designers seized the tools of the day to focus on issues of validity, reliability, and objectivity (D. Brown & Abeywickrama, 2015). Furthermore, H. D. Brown and Lee (2015) proposed the idea of an approach that is a set of assumptions dealing with the nature of language, learning, and teaching, or theoretical positions and beliefs about teaching, language, language learning, learners, institutional and societal factors, purposes of a course, and the applicability of all to a specific educational context. For a better understanding on the application of such a certain approach, then, an approach must have criteria for: (1) providing teachers with a view of how the field of language teaching has evolved and forms part of the disciplinary knowledge expected of language teachers today; (2) introducing teachers to the issues and options that are involved in planning and developing a language course; and (3) introducing a variety of principles and procedures that teachers can review and evaluate in relation to their

own knowledge, beliefs, and practice in teaching speaking (H. D. Brown & Lee, 2015; Richards & Rodgers, 2014).

The importance of better understanding by English teachers will bring a bright future and improvement on the application of E-TPS. At least, they will realise that language teaching has evolved and forms part of the disciplinary knowledge expected of language teachers, especially in the classroom for teaching speaking skills as a part of pronunciation skills.

Teaching speaking is about extreme importance of the students' need to practice freely and openly without fear of being corrected for every minor flaw. On the other hand, there is a need to correct some selected grammatical and phonological errors so that students do not fall into the trap of assuming that "no news is good news" (no correction implies perfection). Pronunciation work (on phonemes, phonemic patterns, intonation, rhythm, and stress) is very important at this stage. Neglecting phonological practice now may be at the expense of later fluency (H. D. Brown & Lee, 2015). In a similar sense for teaching pronunciation, in detail, it would be acceptable to say that a language programme should emphasise whole language, meaningful contexts, and automaticity of production, focusing on tiny phonological details of language. Deep down, teaching pronunciation will allow an attempt to build a learner's articulatory competence from the bottom up, and simply as the mastery of a list of phonemes and allophones; top-down approaches – rhythm and intonation – are then given high priority. Interestingly, the ideal goal is focused on clear, comprehensible pronunciation. At the beginning levels, we want learners to surpass that threshold beneath which pronunciation detracts from their ability to communicate. At the advanced levels, pronunciation goals can focus on elements that enhance communication: intonation features that go beyond basic patterns, voice quality, phonetic distinctions between registers, and other refinements that are far more important in the overall stream of clear communication than rolling the English /r/ or getting a vowel to perfectly imitate that of a native speaker.

In speaking English classrooms, there is much the teachers can do to help students improve their speaking competence. This is not only students practising speaking through engaging classroom activities, but also their learning about the nature of speaking in a second language and ways to manage speaking development (Chong et al., 2022; Darmi et al., 2018; P. S. C. Goh, 2019; Huang & Hashim, 2020; Quinto & Macayan, 2019). In the same sense, it involves principles and experience in creating and implementing approaches such as direct/controlled and indirect/transfer approaches respectively (Burns, 1998; C. C. M. Goh & Burns, 2012; Richards, 2004; Thornbury & Slade, 2006). The direct/controlled approach is concerned with structural accuracy and emphasises the practice of language form, such as the pronunciation of sounds in English, and aims to raise learners' awareness about the grammar of the target language (Larsen-Freeman & Anderson, 2011; Nagy & Robertson, 2009; Richards & Rodgers, 1999; Turnbull & Dailey-O'Cain, 2009), as well as discourse structures and routines. On the other hand, the indirect/transfer approach is concerned with the fluency of speech. It engages in functional language use by getting students to talk in the classroom.

Below is the table on approaches and activities for teaching speaking as proposed by Burns (1998), as cited in C. C. M. Goh and Burns (2012).

Table 5. *Approaches and activities for teaching speaking*

	Direct/Controlled	Indirect/Transfer
Aim	Develop enabling skills	Develop interaction strategies
Focus	Accuracy	Fluency
	Language analysis	Language for communication
Characteristics	Controlled language use	“Authentic”/functional language use
	Skill getting	Skill using
	Pedagogic	Real life
	Pre-communicative	Communicative
	Pre-task practice	Whole-task practice
Activities	Drills	Discussions
	Pattern practice	Information gaps
	Structure manipulation	Project work
	Language awareness	Role plays
	Consciousness raising	Simulation; taking circles
Interaction	Teacher-led	Learner-centred

In the same sense, the approaches should be in balance between language knowledge and language use (McMillan & Turnbull, 2009; Turnbull & Dailey-O’Cain, 2009); the framework is language as a structural system, but one whose primary function is to enable communication to take place. The activities offered are part-skill as a direct practice, while whole tasks are indirect practice or transfer practice activities, as explained in the table below:

Table 6. *Approaches; part-skill and whole-task practices*

	Learning strategies	Characteristics
Part-skill practice	Pre-communication task	Controlled, predictable
	Language work	
Part-skill practice/ whole-task practice	Communication language practice.	Variation in degree of control and predictability
	Structured communication task	
whole-task practice	Authentic communication task	Flexible, less flexible

More recently, Thornbury (2005) advocated a general approach to skill development for the teaching of second language speaking. The approach consists of three stages: (1) awareness raising, which aims at helping learners uncover gaps in their knowledge about speaking; (2) appropriation, which goes beyond controlled practice or restructuring of knowledge for speaking, and develops “practice control” to demonstrate “progressive control” or “self-regulation of a skill; and (3) autonomy, which refers to engagement in activities that demonstrate a degree of autonomy inside and outside the classroom. There is merit in integrating features of direct and indirect approaches; as will show up later, exposure tends to dominate teaching speaking skills in the classroom.

Then, as a part of speaking skill, E-TPS offers closely related mutual benefits. The meaning is clear that there is much teachers can do to help and engage students in improving their speaking competence, while this process is improving teachers' pronunciation skill through valid raters and a provable device.

CONCLUSION

Future pronunciation skills assessment for E-TPS can be carried out using Praat as a device and human interpretation by raters with a goal in teaching pronunciation of being “near-native” and “socially acceptable”. To conclude, “near-native” is an ability possessed by English teachers which coherently exists within their knowledge, assuming that the English teachers have such a certain knowledge of English as a language before they teach in the classroom, and one of their skills is the ability to pronounce English to “near-native” standard – in other words, at an achieved level of proficiency. While “socially accepted” proposes a certain behaviour appropriate to the nature of English as a language, engaging interaction where people using English as a language communicate with one another, with the ability to accept any cultural lag among society. Furthermore, the ideal perspective of Praat as a device with human interpretation by raters, concludes on the aspects of (1) to what extent the measures of pronunciation constructs should be valid and reliable – those two ideas are important to avoid bias in an assessment; and (2) the role of the rater in English pronunciation assessment, which presents unique challenges in drawing valid inferences from performance, scoring assignments, and ultimately to decisions for which a pronunciation assessment was intended, to the pedagogical and social consequences beyond, and to provide the link between concerns in giving judgments; (3) the rater should have both a familiarity with the accent and a familiarity with the content; (4) a suprasegmental approach (stress, rhythm, and intonation) is an important complement to the teaching of segmental; (5) rhythm plays a big role and is defined as a “continuum” of functions and effects in pronunciation assessment; (6) trends in testing using automated assessment of pronunciation for the technology-minded is for the goal of gauging accuracy in imitating a native speaker, and is reflected in the use of Praat as a device.

ACKNOWLEDGMENTS

The authors would like to deliver their gratitude to the honorable editor and reviewers of the Indonesian TESOL Journal who reviewed the final version of this article

REFERENCES

- Baratta, A. (2019). *World Englishes in English Language Teaching*. Palgrave Macmillan.
- Behr, N. S. (2022). English Diphthong Characteristics Produced by Thai EFL Learners: Individual Practice Using PRAAT. *Computer Assisted Language Learning Electronic Journal (CALL-EJ)*, 23(1), 401–424. <https://dl.ndl.go.jp/info:ndljp/pid/12863064>

- Boersma, P., Weenink, D. J. M., & Van Heuven, V. (2007). Speak and unSpeak with PRAAT. <https://www.researchgate.net/publication/259810776>
- Brown, D., & Abeywickrama, P. (2015). *Language Assessment: Principles and Classroom Practices (Third Edition)*. Pearson.
- Brown, D., & Abeywickrama, P. (2019). *Language Assessment: Principles and Classroom Practices (Fourth Edition)*. Pearson.
- Brown, H. D., & Lee, H. (2015). *Teaching by Principles: an Interactive Approach to Language Pedagogy (Fourth Edition)*. Pearson Education.
- Burns, A. (1998). Teaching Speaking. *Annual Review of Applied Linguistics*, 18, 102-123. <https://doi.org/10.1017/S0267190500003500>
- Burns, Anne., & Claire, S. (2003). *Clearly Speaking: Pronunciation in Action for Teachers*. National Centre for English Language Teaching and Research Macquarie University.
- Chong, Y. S., Hamid, B. A., Amzah, N., & Salehuddin, K. (2022). Investigating L2 Speech Production through Verbal Reports: A Systematic Review of Recent Studies. *GEMA Online Journal of Language Studies*, 22(2), 24–42. <https://doi.org/10.17576/gema-2022-2202-02>
- Chun, D. (2002). *Discourse Intonation in L2: From Theory and Research to Practice*. John Benjamins Publishing Company.
- Cummins, F., & Port, R. (1998). Rhythmic Constraints on Stress Timing in English. *Journal of Phonetics*, 26, 145–171. <https://doi.org/10.1006/jpho.1998.0070>
- Darmi, R., Puteh Behak, F., Abdulllah, H., Darmi, R., & Wahi, W. (2018). Variations of L1 Use in the English Language Class. *GEMA Online Journal of Language Studies*, 18(2), 284–311. <https://doi.org/10.17576/gema-2018-1802-19>
- De Jong, N. H. (2018). Fluency in Second Language Testing: Insights From Different Disciplines. *Language Assessment Quarterly*, 15(3), 237–254. <https://doi.org/10.1080/15434303.2018.1477780>
- Dimova, S. (2018). Pronunciation Assessment in the Context of World Englishes. In O. Kang & A. Ginther (Eds.), *Assessment in Second Language Pronunciation (First Published)*. Taylor and Francis Group.
- Georgiou, G. P. (2019). EFL Teachers' Cognitions about Pronunciation in Cyprus. *Journal of Multilingual and Multicultural Development*, 40(6), 538–550. <https://doi.org/10.1080/01434632.2018.1539090>
- Ghanem, R., & Kang, O. (2018). Pronunciation Features in Rating Criteria. In O. Kang & A. Ginther (Eds.), *Assessment in Second Language Pronunciation (First Published)*. Taylor and Francis Group.

- Goh, C. C. M., & Burns, A. (2012). *Teaching Speaking: A Holistic Approach*. Cambridge University Press.
- Goh, P. S. C. (2019). Preschool Teachers' Perspectives on Using English Language to Teach. *GEMA Online Journal of Language Studies*, 19(4), 346–362. <https://doi.org/10.17576/gema-2019-1904-18>
- Harding, L. (2018). Validity in Pronunciation Assessment. In O. Kang & A. Ginther (Eds.), *Assessment in Second Language Pronunciation (First Published, pp. 30–48)*. Taylor and Francis Group.
- Huang, Y., & Hashim, A. (2020). A Quantitative Study of Chinese Learners' Identities as Reflected in Their Attitudes toward English Accents. *GEMA Online Journal of Language Studies*, 20(1), 151–168. <https://doi.org/10.17576/gema-2020-2001-10>
- Isaacs, T. (2018). Shifting Sands in Second Language Pronunciation Teaching and Assessment Research and Practice. *Language Assessment Quarterly*, 15(3), 273–293. <https://doi.org/10.1080/15434303.2018.1472264>
- Isaacs, T., & Harding, L. (2017). Pronunciation Assessment. *Language Teaching*, 50(3), 347–366. <https://doi.org/10.1017/S0261444817000118>
- Isaacs, T., & Trofimovich, P. (2017). Key Themes, Constructs and Interdisciplinary Perspectives in Second Language Pronunciation Assessment. In *Second Language Pronunciation Assessment: Interdisciplinary Perspectives (pp. 3–8)*. Multilingual Matters Ltd. <https://doi.org/10.21832/ISAACS6848>
- Kallio, H., Kautonen, M., & Kuronen, M. (2023). Prosody and Fluency of Finland Swedish as a Second Language: Investigating Global Parameters for Automated Speaking Assessment. *Speech Communication*, 148, 66–80. <https://doi.org/10.1016/j.specom.2023.02.003>
- Kang, O., & Ginther, A. (2018). *Assessment in Second Language Pronunciation*. Taylor and Francis Group.
- Larsen-Freeman, D., & Anderson, M. (2011). *Techniques & Principles in Language Teaching*. Oxford University Press.
- Lieshout, P. Van. (2017). *PRAAT Short Tutorial: an Introduction*. University of Toronto, Department of Speech Language Pathology, Faculty of Medicine, Oral Dynamics Lab.
- Mairano, P., & Santiago, F. (2020). What Vocabulary Size Tells Us about Pronunciation Skills: Issues in Assessing L2 Learners. *Journal of French Language Studies*, 30(2), 141–160. <https://doi.org/10.1017/S0959269520000010>

- Maryn, Y. (2017). Practical Acoustics in Clinical Voice Assessment: A Praat Primer. *Perspectives of the ASHA Special Interest Groups*, 2(1). <https://doi.org/10.1044/persp2.SIG3.14>
- McMillan, B., & Turnbull, M. (2009). Teachers' Use of the First Language in French Immersion: Revisiting a Core Principle. In *First Language Use in Second and Foreign Language Learning. Multilingual Matters*. <http://www.multilingual-matters.com>,
- Moere, V. A., & Suzuki, M. (2018). Using Speech Processing Technology in Assessing Pronunciation. In O. Kang & A. Ginther (Eds.), *Assessment in Second Language Pronunciation (First Published, pp. 137–152)*. Taylor and Francis Group.
- Mompean, J. A., & Gonzales, J. F. (2015). *Investigating English Pronunciation: Trends and Directions (First Published)*. Palgrave Macmillan.
- Monfared, A. (2018). Ownership of English in the Outer and Expanding Circles: Teachers' Attitudes toward Pronunciation in ESL/EFL Teaching Contexts. *Asian Englishes*, 21(2), 207–222. <https://doi.org/10.1080/13488678.2018.1554303>
- Nagle, C., Sachs, R., & Zárate-Sández, G. (2020). Spanish Teachers' Beliefs on the Usefulness of Pronunciation Knowledge, Skills, and Activities and their Confidence in Implementing them. *Language Teaching Research*, 1–27. <https://doi.org/10.1177/1362168820957037>
- Nagy, K., & Robertson, D. (2009). Target Language Use in English Classes in Hungarian Primary Schools. In *First Language Use in Second and Foreign Language Learning. Multilingual Matters*. <http://www.multilingual-matters.com>,
- Norton, P. (2003). Stress at the Suprasegmental Level. In A. Burns & S. Claire (Eds.), *Clearly Speaking: Pronunciation in Action for Teachers*. National Centre for English Language Teaching and Research Macquarie University.
- Nunan, D. (2023). The Changing Landscape of English Language Teaching and Learning. In E. Hinkel (Ed.), *Handbook of Practical Second Language Teaching and Learning (First Published, pp. 2–24)*. Routledge Taylor and Francis Group.
- Osatananda, V., & Thinchana, W. (2021). Using Praat for English Pronunciation Self-practice outside the Classroom: Strengths, Weaknesses, and its Application. *LEARN Journal: Language Education and Acquisition Research Network*, 14(2), 372–396. <https://so04.tci-thaijo.org/index.php/LEARN/index>
- Park, C. U., & Kim, H. J. (2015). Measurement of Inter-Rater Reliability in Systematic Review. *Hanyang Medical Reviews*, 35(1), 44–49. <https://doi.org/10.7599/hmr.2015.35.1.44>

- Pennington, M. C. (2015). Research, Theory and Practice in L2 Phonology: a Review and Directions for the Future. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English Pronunciation: Trends and Directions (First Published, pp. 149–174)*. Palgrave Macmillan.
- Pennington, M. C., & Rogerson-Revell, P. (2019). *English Pronunciation Teaching and Research: Contemporary Perspectives*. Palgrave Macmillan. <http://www.palgrave.com/gp/series/14504>
- Quinto, E. J. M., & Macayan, J. V. (2019). Exploring English Speaking Anxiety among Filipino Engineering Students: Its Influence on Task Performance and Its Sources. *GEMA Online Journal of Language Studies*, 19(3), 57–74. <https://doi.org/10.17576/gema-2019-1903-04>
- Richards, J. C. (2004). *Communicative Language Teaching Today*. Cambridge University Press.
- Richards, J. C., & Rodgers, T. S. (1999). *Approaches and Methods in Language Teaching: A Description and Analysis*. Cambridge University Press.
- Richards, J. C., & Rodgers, T. S. (2014). *Approaches and Methods in Language Teaching (Third Edition)*. Cambridge University Press.
- Rozaimée, F. (2018). The Intonation of English and Malay Questions in Brunei Darussalam. *Asian Englishes*, 21(1), 38–51. <https://doi.org/10.1080/13488678.2018.1448963>
- Ryan, S., & Ryuji, T. (2021). Assessing the Practicality of Using an Automatic Speech Recognition Tool to Teach English Pronunciation Online. *STEM Journal*, 22(2), 93–104. <https://doi.org/10.16875/stem.2021.22.2.93>
- Saito, K., Macmillan, K., Kachlicka, M., Kunihara, T., & Minematsu, N. (2022). Automated Assessment of Second Language Comprehensibility: Review, Training, Validation, and Generalization Studies. *Studies in Second Language Acquisition*, 1(1), 1–30. <https://doi.org/10.1017/S0272263122000080>
- Seargeant, P. (2016). World Englishes and English as a Lingua Franca: a Changing Context for ELT. In G. Hall (Ed.), *The Routledge Handbook of English Language Teaching (First Published, pp. 13–26)*. Routledge Taylor and Francis Group.
- Stanley, J., & Lipani, L. (2019). *Stanley Praat Basics: Introduction to the Software*. Faculty Publications. <https://scholarsarchive.byu.edu/facpub/6106>
- Starfield, S. (2016). English for Specific Purposes. In G. Hall (Ed.), *The Routledge Handbook of English Language Teaching (First Published, pp. 150–164)*. Routledge Taylor and Francis Group.

- Styler, W. (2023). *Using Praat for Linguistic Research*. Associate Teaching Professor of Linguistics at UC San Diego Director of UCSD's Computational Social Science Program. <https://github.com/stylerw/usingpraat>.
- Thomas, S., & Scobbie, J. M. (2015). Mixed Accents: Scottish Children with English Parents. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English Pronunciation: Trends and Directions (First Published, pp. 83–107)*. Palgrave Macmillan.
- Thornbury, S. (2005). *How To Teach Speaking*. Longman Pearson.
- Thornbury, S., & Slade, D. (2006). *Conversation: from Description to Pedagogy*. Cambridge University Press.
- Turnbull, M., & Dailey-O'Cain, J. (2009). Introduction. In *First Language Use in Second and Foreign Language Learning (pp. 1–14)*. Multilingual Matters.
- Wang, Y. (2019). An Empirical Study on the Effectiveness of Speech Visualization Software Praat in Assisting English Phonetic Rhythm Teaching: a Phonetic Study Based on Minnan-accented Learner. *Advances in Social Science, Education and Humanities Research*, 329(4), 732–735.
- Wells, J. C. (2009). *English Intonation: An Introduction*. Cambridge University Press.
- Woldetsadik, G. T., Bachore, M. M., Woldeab, T. A., & Gezahegn, T. H. (2022). Mobile-Based Aural Oral Skill Lessons: The Effects on EFL Teacher Trainees' Oral Performance. *Cogent Education*, 9(1). <https://doi.org/10.1080/2331186X.2022.2112943>
- Yan, X., & Ginther, A. (2018). Listeners and Rater: Similarities and Differences in Evaluation of Accented Speech. In O. Kang & A. Ginther (Eds.), *Assessment in Second Language Pronunciation (First Published)*. Taylor and Francis group.
- Yang, W., & Zhao, X. (2021). Research on the Function of Visual Phonetic Software Praat in Vocational English Phonetics Teaching. *Journal of Physics: Conference Series*, 1856(1), 1–7. <https://doi.org/10.1088/1742-6596/1856/1/012057>
- Young, J. C., & Shishido, M. (2022). Evaluation of Offline Automated Speech Recognition for English as Second Language Learning Application. EdMedia Innovate Learning Online.
- Zhang, D., Lin, C. H., Zhang, Y., & Choi, Y. (2017). Pinyin or No Pinyin: Does Access to Word Pronunciation Matter in the Assessment of Chinese learners' Vocabulary Knowledge? *Language Learning Journal*, 47(3), 344–353. <https://doi.org/10.1080/09571736.2017.1289237>