■ *AUDACITY* AND *PRAAT* AS PEDAGOGICAL TOOLS: ANALYSING FLUENCY AND PRONUNCIATION ACCURACY

XAVIER MARTIN-RUBIÓ[1]
Universitat de Lleida
Lleida, Spain

Glavna pretpostavka ovog rada jeste da analiziranje usmene produkcije pomoću pristupa pod imenom Kompleksnost, tačnost, fluentnost (Complexity, Accuracy and Fluency) može biti od koristi studentima engleskog jezika. U radu se opisuje projekat koji su studenti Univerziteta u Ljeidi (Katalonija, Španija) sproveli u okviru kursa Engleska fonetika i fonologija tokom školske 2019/2020. godine. Ovaj projekat, koji je podrazumevao upotrebu programa *Praat* i *Audacity*, između ostalog je bio pogodan za studente budući da se fonetske teme obrađene u softverima preklapaju sa onim koje se obrađuju u okviru pomenutog kursa. Međutim, smatramo da bi se projekat mogao prilagoditi i potrebama drugih studenata i predmeta.

Ključne reči: fluentnost, tačnost izgovora, pedagoški alati, samorefleksija.

1. INTRODUCTION

Asking students of English to analyse their oral productions using measures from the (Second Language Acquisition) Complexity, Accuracy and Fluency approach may be beneficial for their English language learning process. In this article, I intend to describe an individual project that students of the English Phonetics and Phonology course of the English Studies degree at Universitat de Lleida (in Catalonia, Spain) are required to carry out every year in order to explore in what ways such a task can contribute to the students' improvement in this area of English language. In order to do this, students are requested to follow the detailed instructions of the project for the 2019/2020 academic year (provided through video) and to use the specific submission from a past student as a model of success in the activity.

---

1   Kontakt podaci (E-mail): xavier.martinrubio@udl.cat

The goal of this academic project is twofold: firstly, to help students improve their transcription skills and knowledge through extensive transcribing practice; and, secondly, to promote the students' reflection regarding their own productions, which might lead to pronunciation problems detection, although it will ultimately be the students' choice whether to address them or not. The project was devised so that students can learn practical tasks such as phonemic transcription, the calculation of phonetic values, and the interpretation of spectrograms. All these aspects contribute to raising the students' awareness of the complexity of pronunciation accuracy and its relationships with the contextual reality in which speech is produced. I believe adaptations of this project with a less technical approach could be designed for students in other degrees or contexts; these versions would keep the core idea of encouraging students to record their own oral production and subsequently reflect on the outcome.

This article is structured into three parts: in the first one, a theoretical overview of Complexity, Accuracy and Fluency analysis within the Second Language Acquisition field is provided, and its connection to phonology is justified. In the second part, a step-by-step description of what students are expected to do for their individual project of the English Phonetics and Phonology course in the 2020/2021 academic year is supplied. Ultimately, in the third part, the individual project from one of last year's students (anonymised as Laia) is employed to reflect upon the possibilities that this task may offer to English language students.

Prior to this, some background information regarding the setting of the project might be helpful. The course 'English Phonetics and Phonology' is a 6 ECTS year 1 course of the English Studies degree at Universitat de Lleida. Universitat de Lleida is the only university in the province, as it is a sparsely populated area in the hinterland of Catalonia. Lleida province represents 38% of the total surface of Catalonia but it contains only 5.6% of its population. A majority of the students pursuing the degree are local students, most of whom have Catalan and/or Spanish as their first language (L1), but a growing number of local students are newcomers or the children of newcomers. The mother tongue of these students could thus be Romanian, Ukrainian, Moroccan Arabic or Amazigh, for instance. For many, one of the varieties of Spanish spoken in Latin America is their native language.

The course is structured into two parts: the first part deals with phonemes, the syllable, and connected speech, whereas part two focuses on the three main components of intonation, i.e. tonality, tonicity, and tone. Students are expected to have a B2 level of English at the time of enrolment, and this is true for most of them.

There were 55 students enrolled in the course in the 2019/2020 academic year. The course started in mid-February, with regular onsite sessions, but the lock-down that began in mid-March as a result of the COVID-19 pandemic resulted in a shift towards online sessions until the end of the course. This online teaching had to be, to a certain extent, improvised: there were video conferences of different types and students submitted tests and assignments online, although the originally planned course structure was maintained as much as possible.

For the 2020/2021 course, with 39 students enrolled, a hybrid teaching methodology has been planned; of the two programmed weekly sessions, only one will take place onsite. Video conferences with small groups and online tasks will complement the

onsite sessions. Another change from the previous academic year is that the intonation part of the course will be dealt with first this year, leaving phonemes and connected speech for the second part of the course.

2. CAF, PRONUNCIATION ACCURACY AND PHONOLOGY

CAF stands for 'Complexity, Accuracy, and Fluency' and has had a long tradition within Second Language Acquisition studies as a way to analyse an individual's learning process. 'Complexity' is defined by Michel (2017) as "the size, elaborateness, richness, and diversity of the L2 performance". 'Accuracy', on the other hand, has been described as "the ability to produce error-free language" (Czwenar 2014: 82), although what counts as error is rather debatable and subject of controversy. Finally, 'fluency' has been defined as "the ability to talk with normal levels of continuity, rate and effort" (Starkweather 1987: 12). When speech is constantly interrupted by pauses, continuity is affected; when only a small number of syllables are uttered in a given amount of time, the rate is low; and when the speaker struggles to find the suitable word, an unusual amount of effort is required.

Determining the CAF levels of a production depends on at least two elements: the task to be developed and the level of language proficiency of the performer. Apart from these, the level of accuracy might be influenced by the socio-educational context of the speaker, whereas the level of fluency might be affected by the speaker's personality, such as the degrees of self-confidence and extroversion (MacIntyre *et al.* 1998; Arnold 1999), as well as by personal traits such as speech impairments, like stuttering (Fortunato-Tavares *et al.* 2017).

Different tasks present different demands on the people that engage in them. This can be easily gathered from a task comparison. One could, for instance, consider the three following tasks: reading a text, describing a comic strip, and participating in a debate. While reading a text, the reader must only process the language written on the page; there is no need to improvise, even though the reader must still make sense of the words in the page. The person reading might produce pronunciation errors, but they would not be responsible for the lexical and grammatical elements in the text; the author of the text would be. When asked to describe a comic strip, one needs to choose what to say; although the task itself narrows the vocabulary to be employed, one has a larger range of lexical choices than while reading. Furthermore, in a monological comic-strip description, the participant has to solve the task in a self-sufficient way. In debates with other people, conversely, one can pick and reuse lexical and grammatical elements from the other participants. Moreover, one needs to process the information conveyed by the other participants and decide what to say and how to say it, while simultaneously dealing with issues of taking, holding, and yielding the floor. A learner will, thus, speak with different levels of fluency and accuracy across these different tasks.

Regarding fluency, language learners at the early stages of their learning processes may certainly struggle to formulate long meaningful chunks, consequently producing a higher number of pauses in their speech. A portion of such pauses will be situated in the middle of phrases, thus giving the impression to the listeners that they are not fluent

(Wennerstrom 2000). As Chambers (1997: 540) points out, "becoming fluent therefore is ... about pausing less often and pausing at the appropriate junctures in an utterance" and, at the phonological level, Hieke (1985) argued that fluent speech equals connected speech. When the number of uninterrupted syllables is measured, what is actually being measured is how capable the speaker is of connecting syllables without pausing; this is a great indicator of the progress in the language. This is further exacerbated in English, if one considers that it is a stress-timed language; therefore, the learner needs to learn to connect syllables with an alternation of stressed and unstressed syllables that is different in syllable-timed languages like Spanish (Leal 1995).

Accuracy, on the other hand, is intimately related with the notions of "error" and "mistake". According to Brown (2004: 216), "a mistake refers to a performance error in that it is a failure to utilize a known system correctly; while an error is a noticeable deviation from the adult grammar of a native speaker, reflecting the interlanguage competence of the learner". In this definition, the adult grammar of a native speaker is taken as the norm, and noticeable deviations are deemed as errors, whereas performance errors are described as mistakes. In the same line, Canagarajah (2015) argues that errors, unlike mistakes, are systematic and indicate the personal "grammar" of the language user. Errors are an integral part in the learning of an additional language, and one good indicator of progress is precisely the learner's decrease in the number of errors. Errors can be classified as lexical (word choice), phonological (pronunciation), semantic (meaning), syntactic (grammar), and pragmatic (content) (Jiménez Arias 2004: 177).

Pronunciation accuracy inevitably leads us to phonetics, the study of sounds, and to phonology, the study of the phonemes of a language. If the oral productions under analysis came from native speakers of a specific variety of a language, one would expect these productions to be error-free; there might be some production mistakes at most. In our case, however, the oral productions are performed by multilingual speakers who have English in their linguistic repertoires. Nevertheless, English for them is not simply a Lingua Franca used to make themselves understood and for whom intelligibility is the main goal (Boyd/Rudvin 2018); they are English language learners (and future English language experts) who want to improve their current level in the foreign language, which varies from learner to learner quite remarkably. These different levels in English translate into their oral production in a variety of ways.

Some of the students, for instance, may be unaware of the aspiration of voiceless plosives in English. When one such speaker utters [pɪn] rather than [pʰɪn], for example, the result is that although the two productions sound different, they actually correspond to /pɪn/, phonemically speaking. Although this could easily be regarded as a deviation from standard English, is this susceptible of being referred to as an error? Or is this just systematic L1-influenced pronunciation with little impact on intelligibility? Another common source of non-standard pronunciation comes from the assumption that a word is pronounced as it is written down. Typically, a student knows the written form of a word like "sword" and assumes the /w/ must be pronounced. This is also a deviation from the standard. This time, however, the impact of the non-standard pronunciation on intelligibility tends to be greater. In the next section, I provide the details of a project which my English Phonetics and Phonology students are required to carry out in order to reflect on their oral production in relation to (dis)fluency and (in)accuracy.

3. THE INDIVIDUAL PROJECT

The individual project, which accounts for 20% of the student's final grade, must be carried out in four steps and it entails two different submissions: (1) the spreadsheet with the transcription and the (dis)fluency and accuracy measures; and (2) the analysis of the phonological phenomena.

STEP 1: COMIC STRIP DESCRIPTION

Students are directed to the following website: https://fmart37.wixsite.com/fluency-accuracy. In the section entitled 'Steps and Rules', links to two comic strips are provided (CS1 and CS2). Students are instructed to open the links, spend up to a minute looking at each comic strip, and then audio-record a description of what they see in each comic strip. The two descriptions must come one after the other, and they are warned against looking up the words in dictionaries or writing down a description to be read aloud.

The students can use any audio recording device, in most cases their phones, but the Audacity software is recommended by default. Even when the description is recorded with another device, students are asked to open the resulting audio file in Audacity[2] anyway in order to "clean it up". The cleaning process entails deleting any non-meaningful sounds or silence before the description properly starts and after the last syllable of said description has been produced. The resulting file is exported as: student's surname_student's name_CSD_clean.wav.

STEP 2: BREAKING DOWN THE AUDIO FILE

For the second step of the process, LibreOffice Calc[3] and Audacity are required. As explained on the above-mentioned webpage, there are three types of audio chunks that will need to be identified: silent pauses (sp), filled pauses (fp) and what I call between-pauses units (bp-units), which basically correspond to runs in the CAF literature. A template of the spreadsheet to be used is provided in the Spreadsheets section of the website. There are two tabs in this spreadsheet, called *original* and *ST_name*. The *original* tab contains six columns: the first indicates the chunk number; the second informs us of the type of chunk (sp, fp or bp); the third shows the length of the chunk (in milliseconds); the fourth indicates the number of syllables of the bp-unit (which will be zero for the other types of chunks); the fifth calculates the rate of the bp-unit (syllables divided by length of the bp-unit); and the last column, much wider than the previous ones, contains the phonemic transcription of the bp-unit. It is advisable to number the elements of each type of chunk (sp01, sp02, etc.) and to use different colours for the types of chunks so one can more easily identify them; red can be used for silent pauses, blue for filled pauses, and black for bp-units, for instance.

---

2   The latest version 2.4.2 of the free software available from: https://www.audacityteam.org/

3   LibreOffice 7 is available for free at: https://www.libreoffice.org/ and it contains LibreOffice Writer and LibreOffice Calc.

Once the clean audio file is already in Audacity, and after identifying each type of chunk, the information corresponding to this chunk is typed into row 1 of the spreadsheet. If the audio starts with a bp-unit in which the students says "I can see a woman" during 1.193 seconds, the information in the different columns will be: 001, bp01, 1.193, 6, 5.03, and /ˈaɪ.kæn̩ˌsiˑ.əˈwʊm.ən/. When the chunk is a bp-unit, the selected portion of the file will be cut and pasted into a new Audacity project and exported as bp01.wav. A pause must inevitably come after a bp-unit. This can be a silent pause of 1.042 seconds in length. If this is the case, the information in row 2 will be: 002, sp01, 1.042, 0 and an empty cell. This new portion of the audio is simply deleted. Filled pauses can also be deleted. The process will be repeated until there is only one final chunk left. Once this last chunk is dealt with, step 2 will have been completed.

STEP 3: CALCULATING THE MEASURES

In order to complete the *original* tab, the elements in columns three and four must be added. All the time lengths added in column three constitute a quantity measure: the Total Response Time (TRT henceforth). By adding the number of syllables of all the bp-units, indicated in column 4, we obtain the total amount of syllables produced. We then apply the Data Sort tool to column 2; in this way, we group all the different types of chunks together. The following move involves copying and pasting the rows corresponding to bp-units to the other tab: *ST_name*. This other tab has four extra columns: repetitions; false-starts and self-corrections; pronunciation accuracy; and lexical and grammatical accuracy. Right below the rows for the bp-units, there is a row with measures: column three now contains the time length of just one type of chunk. This measure corresponds to Speech Time (thus the name of the tab). The number of syllables will be the same, as we have only considered syllables within the bp-unit chunks.

51	bp23	0.579	3	5.18	/ˈfaɪ.nəl.i/				
53	bp24	4.282	14	3.27	/hɪ.dɪˈtaɪ.dɪsˈaɪds.tuˈgoʊˈbæk.ænˈstərtsˈrʌn.iŋ.ənˈtɪl/			1	
55	bp25	3.325	15	4.51	/hɪəˈrɑɪvs.tu.dɛˈpleɪs.wɛrˈhɪ.hʌdˈlɛft.dɛˈwɔˑ.mən.æn.dɛˈdɒg/				
58	bp26	2.610	9	3.45	/ˈdɛn.dɛˈmʌn.dɪsˈaɪds.tuˈkɪk.dɛˈdɒg/				
60	bp27	3.072	13	4.23	/ˈlɪv.ɪŋ.dɛˈwɔˑ.mənˈasˈtɒn.ɪʃd:baɪt.wæt.hɪ.hʌdˈdɔˑn/				2
64	bp28	4.766	21	4.41	/ˈɔˑl.dɔˑ.dɛˈwɔˑ.mən.wʊˑdˈprɒb.ə.bli.bɪˈfɪl.ɪŋˈæŋ.gər.ænd.sʊrˈpraɪs.ʌt.dɛ.seɪmˈtɑɪm/				
66	bp29	4.246	17	4.00	/dɛˈwæ.dɛˈmæn.wəs.sæt.isˈfaɪd.bɪˈkɔˑs.hɪ.hʌd.əˈtʃiˑvd.hɪs.rɪˈvɛndʒ/			1	
		65.149	261	120.12			0	3	5	2
		ST	syl				rept	fssc	pro	lex & gr
	9.00	MSR		65.15	ST		0	1.15	1.92	0.77
	4.01	ROST		87.71	TRT		rept	fssc	pro	lex & gr
	4.14	Arbpu					/100syllables			
	74.28	STR								

Figure 1. Measures in the spreadsheet

Figure 1 above shows how students must present the measures in the spreadsheet. MSR stands for Mean Syllables per Run, and provides information about how long, in syllables, the runs are, on average. To calculate MSR, we divide the number of syllables by the total number of runs/bp-units (261 divided by 29 in Figure 1 below). ROST stands for Rate of Speech Time, and it results from dividing the total number of syllables by the Speech Time (261 divided by 65.149 seconds in the example). This tells us that the student has produced 4.01 syllables per second on average. ARbpu stands for Articulation Rate per bp-unit, and it is a slightly different way of calculating the speech rate: this time, we first calculate the average rate of each bp-unit (something we have indicated in column 5), and then, we add these rates (120.12 in the example) and divide this by the number of bp-units. This number should not differ greatly from the ROST. In the example, ROST is 4.01 and ARbpu is 4.14. Finally, STR stands for Speech Time Ratio, and it informs us about the percentage of time which the student has devoted producing meaningful syllables. We multiply the Speech Time by 100 and divide it by the TRT. In this case, the student has spent 74.28% of the time producing syllables.

Apart from these quantity and fluency measures, the four columns on the right side refer to (dis)fluency and accuracy measures. In Figure 1, in the cells from the first row which have a grey background, the total number of cases is written down. Three rows down, the number of cases per 100 syllables is provided, so these measures can be compared to much shorter or longer productions. Only consecutive repetitions are counted, but as many as there are. Some examples might help: "in in in this case", two repetitions; "in this_ in that case", one self-correction; "in this_what I thought was that", one false start. Fluency and disfluency measures need to be analysed together: a description could have, on the one hand, a high MSR and high ROST, which would indicate good fluency, but, on the other, a high number of repetitions, false-starts, and self-corrections, which would point in the opposite direction.

Regarding accuracy, which refers to deviation from the standard, it is essential to take both the speaker and the task into consideration. Students are asked to focus on pronunciation, but they can also complete the grammar and lexical errors column. As it was previously explained, these students are multilingual speakers with English in their repertoire, but with varying degrees of command of the language. Laia, the student discussed in part 3 of this article, for instance, was born in a village near Lleida and her L1 is Catalan. Her father is from the neighbouring Autonomous Community of Aragon, but they have always spoken Catalan within the nuclear family. She uses Spanish with relatives on her father's side. The student has attended English language academies for several years and has spent time abroad: two weeks in London in 2017; three weeks in Ireland in 2018; and a month in Toronto in 2019. She has a B2 level of French and is currently studying Italian. She also expresses a wish to learn other languages, like German and Chinese, in the future. Moreover, she is also enrolled in an English Studies programme.

Level variation within the group goes from students with a low B2 level to a student born in London to a British father and Catalan mother, whose mother tongue is English. Needless to say, therefore, the number of errors will vary greatly among students. Identifying the errors is a very controversial task that will be discussed in section 3, taking Laia's performance as a guide, but at this point it needs to be emphasised that having the students identify their own errors and then reflect upon them is an effective

way to increase their awareness regarding their language level and the features of their production in English. Furthermore, students also benefit from such reflection in that they realize that sounding like native speakers is not necessarily the ultimate goal; becoming fluent and intelligible speakers of English is.

STEP 4: ANALYSIS OF PHENOMENA

For this part of the assignment, *LibreOffice Writer*, *Audacity* and *Praat* are required. The analytical report comprises three main blocks. Block 1 contains a brief description of the (dis)fluency and accuracy measures obtained in the previous assignment. Students are offered spreadsheets made by students in the previous years, so that they can compare their measures and speculate on their meanings. At this point, false starts and self-corrections need to be commented. In Laia's description, for instance, there are three such cases: bp16, bp24, and bp29. In bp16, Laia says "decided" followed by /ku/, but then she starts again with /tu/ and completes the idea with "give her a smile". There are different options for what might have occurred in this instance: Laia may have wanted to say "decided could_" and realised the verb "to decide" takes the form 'to+infinitive' and therefore self-corrected; or maybe she wanted to say "to give" but made a mistake.

The other two cases are clearly cases of self-correction. In bp24, Laia starts with /hɪ.dɪˈtaɪ/ and then self-corrects to /dɪˈsaɪds/. It is very likely that she had always meant to say "he decides", but had inadvertently produced the plosive sound /t/ rather than the fricative /s/ as required. After realising the mistake, the student would have stopped mid-word and self-corrected in the following syllable. In bp29, Laia starts with "the wo(man)" and realises her mistake, so she starts again with "the man".

Block 2 corresponds to the transcription of the comic strip description in story format. The student turns the phonemic transcription into standard writing, pauses are indicated between brackets, rounding the length to one decimal of a second, and filled pauses are transcribed with either "mm" or "e:". Here is an example, corresponding to the description of one of the comic strips by Laia:

in this comic strip (0.3) I can see a man (0.4) who is holding his hat with his left hand (0.5) and a woman (0.3) who is holding his dog (0.9) I imagine that the woman was going for a walk (0.4) with her dog when suddenly (0.4) they bumped into a stranger (1.2) e: in that moment the man decide to touch (0.3) the dog but the animal (0.5) instead of being quiet (0.3) he decided to bite his hand (1.3) mm the woman e: looked angrily at the man (0.9) e: however although he was feeling pain in his hand (0.6) decided cou_ to give her a smile (0.5) while holding his hat with his left hand (2.2) mm after that the man decides to go away (0.6) and leave the woman e: and her dog behind (1.7) however after some steps (0.8) mm he thought that it wasn't a good option (0.6) finally (0.3) he deti_ decides to go back and starts running until e: he arrives to the place where he lad left the woman and the dog (1.1) mm then the man decides to kick the dog (0.4) living the woman astonished by what hi had done (1.1) e: (1.3) although the woman would probably be feeling anger and surprise at the same time (0.5) the wa_ man was satisfied because he had achieved his revenge

The last block is further divided into subsections, one per phonemic or phonetic phenomenon analysed. Students are advised to deal with between three and five phenomena, and they must select these phenomena after listening to their productions several times, something they will have surely done while filling in the spreadsheet. Having the audio files of the different bp-units is helpful in this selection process. Folders can be created for each phenomenon, for instance, and the audio files that contain the chosen phenomenon can be stored in the corresponding folder. A file can obviously contain more than one phenomenon and thus be stored in different folders. How the different phenomena are analysed using *Praat* will be exemplified in the last section of this article.

4. LAIA'S INDIVIDUAL PROJECT

The three elements chosen from Laia's description are: (1) the production of the three phonemes in the word "dog"; (2) a case of epenthesis between /n/ and /s/; and (3) the presence of a fricative rather than an affricate sound in the word "imagine".

For Item 1, a folder that would store the six audio files which contain the word "dog" in Laia's description was created; these files are bp-units 5, 7, 10, 20, 25 and 26. In three out of these six examples, "dog" is placed at the end of the bp-unit (bp05, bp25 and bp26) whereas in the other three "dog" is followed by another syllable (bp07, bp10 and bp20).

In Figure 2 below, corresponding to bp05, "dog" is placed at the end of the bp-unit. The syllable contains two plosives, an initial voiced plosive /d/ immediately followed by a vowel, as voiced plosives are not aspirated. After the vowel, the medial phase during which air builds up follows, and the release burst of a velar plosive more consistent with a /k/ sound closes the bp-unit. We can see how the medial phase before the burst of the plosive takes even longer than the time the vowel is being produced. Ladefoged (1982: 49) observed that the major difference between the final stops in "nap, mat, knack" and "nab, mad, nag [...] is in the vowel length, not in the voicing of the final consonants".
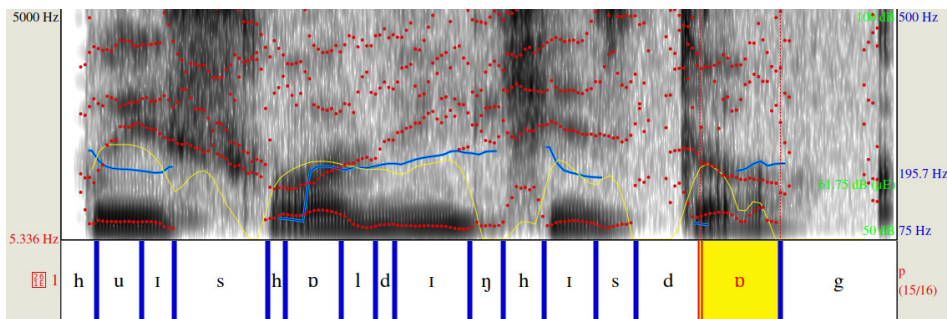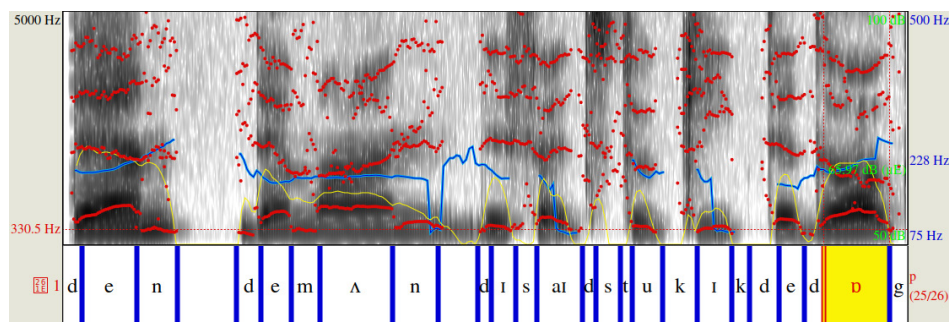

Figure 2. bp05 of Laia's first description: /g/ or /k/?

The best way to indicate that a /g/ phoneme is coming is to lengthen the previous vowel. However, in this case the student has produced a vowel consistent with a /k/ phoneme, gathered much air and released the plosive. Therefore, it is only the context that allows in fact to identify the word as "dog" rather than "dock".

The rendering of "dog" in bp25 is very similar to this one, but the one in bp26 is different in that the vowel takes longer (see Figure 3 below). It actually represents 67% of the time spent delivering the word "dog", whereas in bp05 above, it represented 31%. And this time, in perceptual terms, there is no possible confusion with "dock". It is also worth paying attention to final /k/ in "kick", which is not released at all, thus leading to potential intelligibility issues, as it could just as easily be "keep the dog" rather than "kick the dog".

Figure 3. bp26 of Laia's first description: appropriate vowel length

The second analysed item actually uses one of the audio files above: the one for bp07. In Figure 4 below we can identify a plosive between final /n/ of "when" and the initial /s/ of "suddenly". Tench (2011: 100) points out that "although /n/ and /s/ share an alveolar point of articulation, the tongue changes from a flat 'broad' contact to a grooved shape. As in the other transitions, the tongue movement may lag behind, leaving the flat 'broad' contact fractionally longer; this helps to produce a transitional /t/". Tench (2011) is the coursebook used, and this is the kind of connection students are expected to make in the project. They have a spectrogram in which there is a white area and a very black line, which is something we normally associate with the medial and burst phases of plosives. We hear an /n/ and an /s/, but if we just play the section around the colour change, we actually hear a /t/. And then the student –one hopes – either remembers or seeks what might be occurring here. Is this an error? Or is it this a typical phenomenon of the language? As it turns out, it is a very common phenomenon.
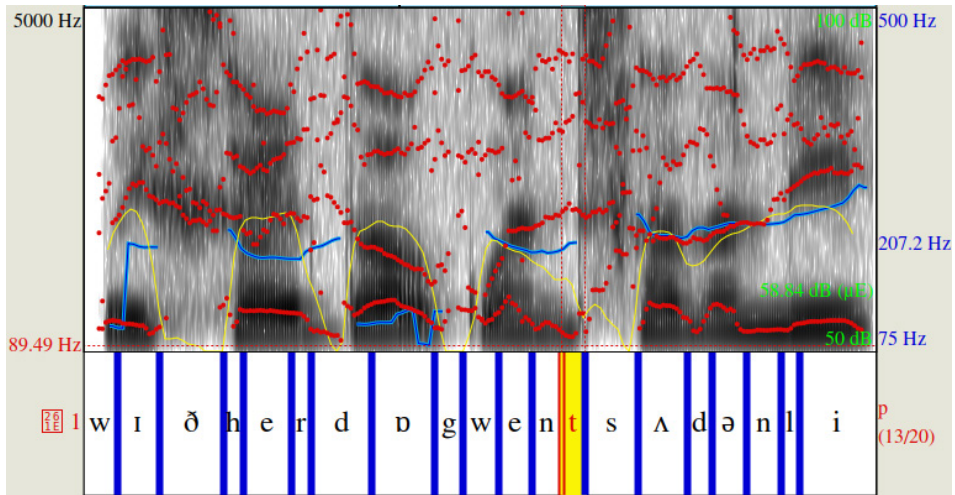
Figure 4. bp07 of Laia's first description: epenthesis

The last item of the analysis refers to bp06 /aj.ɹˈmæʒ.ɪn.dæt.dəˈwɔː.mən.wɔːsˈgo.ɪŋ.fɔːr.əˈwɔːk/, and, more in particular, to the phoneme /ʒ/. The standard pronunciation of "imagine" contains an affricate sound at the end of the second syllable. Tench (2011: 44) describes affricates as "a close-knit sequence of plosive and fricative produced by a single organ of speech". In the column immediately above the phoneme in the spectrogram in Figure 5 below, two distinct sections can be identified: the whiter section corresponding to the plosive part of the affricate sound, and the darker section that corresponds to the fricative part.
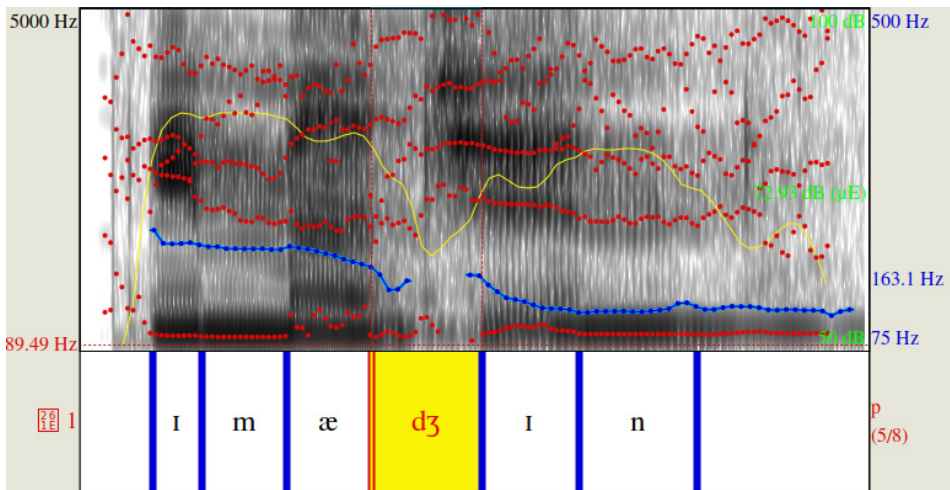

Figure 5. Standard pronunciation of /ɪˈmædʒ.ɪn/

This is not the case in Laia's rendering of the word, as shown in the spectrogram in Figure 6 below. In that case, only the fricative section can be observed. Laia has produced the fricative sound /ʒ/ rather than its affricate counterpart. It is worth mentioning that the same verb is used in Catalan ("*imaginar*") and that the same fricative sound /ʒ/ is used. In Spanish, it may be worth adding, "*imaginar*" is pronounced with the voiced velar fricative /x/.

This non-standard pronunciation of "imagine" was counted as an error in the spreadsheet, whereas the non-standard pronunciations of "dog" in bp05 and bp25 were not. The student is the one who must decide on these matters, but part of the feedback the student receives is related to the complexity of these decisions. Students obtain two types of feedback: one is more technical in nature and refers to how accurately they have divided the chunks, calculated the measures, or phonemically transcribed.
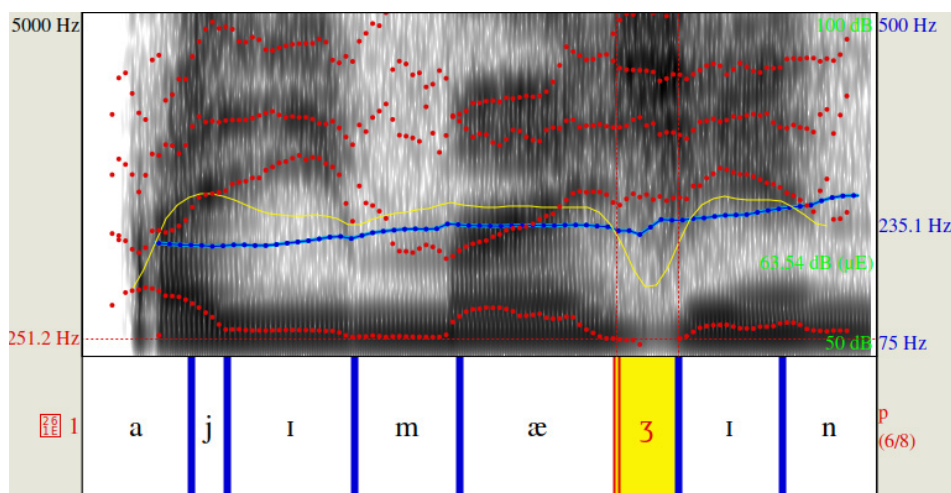

Figure 6. bp06: fricative?

The second kind of feedback aims at helping students reflect upon their productions in terms of (dis)fluency and accuracy. In the case of this last item for Laia's production, what is most relevant to highlight is that Laia realises that she has produced a fricative sound when standard varieties of English would use an affricate sound. Whether this should be considered a pronunciation error or not, or even whether this could pose intelligibility problems in certain contexts or not, is debatable.

5. DISCUSSION AND CONCLUDING REMARKS

The individual project described in these pages is a tool devised so that students can learn practical tasks such as transcribing phonemically, calculating measures, or interpreting spectrograms, while at the same time becoming aware of the intricacies of aspects like pronunciation accuracy and their connection to the contextual reality in which speech is produced. In English as a Lingua Franca scenarios, non-standard

pronunciation which does not lead to intelligibility problems will tend to be overlooked, as the main aim in such settings is to communicate the message effectively and intelligibly. However, in the context of a course that is part of an English Studies programme, it appears both necessary and suitable to target standard pronunciation, as well as to use specialized software like *Audacity* and *Praat* to raise students' awareness regarding how their productions as non-native English speakers stand in relation to the dominant varieties of English. However, it needs to be underlined that this comparison between native and non-native pronunciation should not lead to the idea that the native version should always predominate or be the ultimate goal for English language learners.

An example I often use in class concerns the already-mentioned aspiration of voiceless plosives in English. In the word "potato", we have three such cases in a row. The Catalan and Spanish words for "potato" is "*patata*", but none of the three plosives are aspirated in Catalan or Spanish. I tell students about a famous British author who has been living in Catalonia for many years, who speaks very fluent Catalan, and who is often in the national television channel as a commentator. When he talks, I tell them, he aspirates the plosives in Catalan, marking him as a native speaker of English in Catalonia. When I imitate this man, students laugh, but their laughter contains an element of surprise because they realise how different they feel about this man sounding British while using Catalan in relation to how they feel about themselves sounding Catalan while using English. To me, what matters the most is that they learn about aspiration of plosives, the phonemic contexts in which plosives are aspirated in English and the ones in which they are not, and the fact that this is different in Catalan and Spanish. But then, whether this is something they want to modify or not is up to them. I do not think the aforementioned British author cares too much about the aspiration of his plosives when speaking Catalan. I suppose he might even like it.

REFERENCES

Arnold, J. (ed). 1999. *Affect in language learning*. Cambridge: CUP.

Boyd, J. and M. Rudvin. 2018. Accommodation, Cooperation and Face Strategies in ELF: A Case Study of Institutional Power in Migrant Settings. In C. Martin-Rubió (ed.) *Contextualising English as a Lingua Franca: from Data to Insights*. Newcastle-upon-Tyne: Cambridge Scholars, 170–189.

Brown, H. D. 2004. *Principles of Language Learning and Teaching*. 4th Edition. New York: Longman.

Canagarajah, S. 2015. Clarifying the Relationship between Translingual Practice and L2 Writing: Addressing Learner Identities. *Applied Linguistics Review 6(4),* 415–440.

Chambers, F. 1997. What do we mean by fluency? *System 25(4),* 535–544.

Czwenar, I. 2014. Analysing Spoken Language for Complexity, Accuracy and Fluency: Some Methodological Considerations. In W. Szubko-Sitarek, Ł. Salski and P. Stalmaszczyk (eds.) *Language Learning, Discourse and Communication*. Cham: Springer, 81–92.

Fortunato-Tavares, T. *et al.* 2017. Children who stutter exchange linguistic accuracy for processing speed in sentence comprehension. *Applied Psycholinguistics 38(2),* 263–287.

Hieke, A. E. 1985. A componential approach to oral fluency evaluation. *Modern Language Journal 69(2),* 135–142.

Jiménez Arias, I. 2004. Treating Students' Errors in Oral Production. *Letras 36,* 175–188.

Ladefoged, P. 1982. *A Course in Phonetics*. San Diego: Harcourt Brace Jovanovich.

Leal, C. F. 1995. *A pause as a cause of change*. ERIC document ED 379935. Rockville, MD: Educational Resources Information Center.

MacIntyre, P. D. *et al.* 1998. Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal 82(4)*, 545–562.

Michel, M. 2017. Complexity, Accuracy and Fluency (CAF). In S. Loewen and M. Sato (eds.) *The Routledge Handbook of Instructed Second Language Acquisition*. London: Routledge, 50–68.

Starkweather, C. W. 1987. *Fluency and stuttering*. Englewood Cliffs, NJ: Prentice Hall.

Tench, P. 2011. *Transcribing the Sound of English: A Phonetics Workbook for Words and Discourse*. Cambridge: Cambridge University Press.

Wennerstrom, A. 2000. The role of intonation in second language fluency. In H. Riggenbach (ed.) *Perspectives on fluency*. Ann Arbor: University of Michigan Press, 102–127.

SUMMARY

AUDACITY AND *PRAAT* AS PEDAGOGICAL TOOLS: ANALYSING FLUENCY AND PRONUNCIATION ACCURACY

The main premise of this article is that asking students of English to analyse their oral productions using measures from the Complexity, Accuracy and Fluency approach might be beneficial for their language learning process. We describe a project that students of the English Phonetics and Phonology course of the English Studies degree at Universitat de Lleida (Catalonia, Spain) carried out during the 2019/2020 academic year. The project helps students improve their transcription skills while stimulating reflection on pronunciation aspects. For this purpose, Praat and Audacity software packages are employed, which are especially suitable to deal with phonemic phenomena that need to be covered in the course, although it is possible to think of adaptations of the project to other types of students so that they could still benefit from it.