



Università
Ca' Foscari
Venezia

UNIVERSITÀ CA' FOSCARI VENEZIA

**Master's Degree Programme
in Scienze del Linguaggio
Linguistica Educativa**

Tesi di Laurea

**SPEECH VISUALIZATION TECHNOLOGY
AND TEENAGE EFL LEARNERS**

**The Effects on English Prosody of Praat Use
in a Venetian High School**

Relatore:

Ch.ma Prof.ssa Marcella Menegale

Correlatore:

Ch.ma Prof.ssa Geraldine Ludbrook

Laureanda:

Alice Percivalli
808156

Anno Accademico 2022 / 2023

ABSTRACT

Due to its rather undefined nature, prosody may be considered a particularly challenging aspect of second language (L2) pronunciation. In this regard, speech analysis software *Praat* has proved to be a very helpful tool for L2 learners, who seem to benefit from visualization of waveforms and pitch contours in the enhancement of English suprasegmental features and especially of intonation.

So far, most research has focused on the effects of Praat audio-visual feedback at the university level, essentially neglecting upper secondary educational contexts. To fill this gap, this study presents one possible pedagogical implementation of the software within two second year classes of a Venetian high school (Italy).

Building upon an exploratory sequential mixed method design, evidence from preliminary classroom observation and interviews with teachers determine the structure of the visual feedback paradigm (VFP) used in the successive pronunciation activity with Praat, at the end of which a satisfaction questionnaire is administered.

Results are promising for three reasons: 1) the preliminary phase reveals students' increased interest towards English pronunciation issues; 2) the 12 subjects recording the target questions before and after Praat training show to have benefitted from audio-visual feedback of English models as their intonation contours result being substantially more native like; 3) questionnaire data exhibit an overall favourable response from the subject participants to the use of Praat in the scope of pronunciation instruction.

Keywords: L2 pronunciation, audio-visual feedback, Praat, teenage learners

ACKNOWLEDGEMENTS

I would like to express my most heartfelt thanks to Professor Ludbrook for supporting me since day one of this journey with her precious advice.

I would like to express my sincere gratitude to Professor Menegale for trusting in my dissertation project and for her constant availability.

I am particularly grateful to all the students who took part in this study and their teachers who made it possible.

I am also grateful to Ellie, Esther and Sebastian who lent their voice as native speakers as well as to Cristiano for assisting me with the analysis of the data.

A heartfelt ‘thank you’ goes to all my ex-coworkers for their moral support and constant encouragement.

Finally, I thank my parents, without whose help all this would not have been possible and my friends for continuing to believe in me.

Alice Percivalli

TABLE OF CONTENTS

INTRODUCTION	p.1
LITERATURE REVIEW	p.5
1.1 The vocal instrument	p.5
1.2 Pronunciation and second-language teaching	p.9
1.2.1 The four — plus one — waves of L2 pronunciation teaching history	p.13
1.2.2 Core dichotomies of L2 pronunciation instruction	p.17
1.3 Pronunciation of Italian learners of English	p.25
1.3.1 Italian and English phonological systems: Segmental features	p.26
1.3.2 Italian and English phonological systems: Suprasegmental features	p.31
1.4 Technology and pronunciation instruction	p.36
1.4.1 Visualization technology and acoustic analysis of speech	p.42
1.4.2 <i>Praat</i> : a free voice analyzer to train pronunciation	p.45
RESEARCH PROJECT	p.51
2.1 Research questions and hypotheses	p.51
2.2 Research setting and participants	p.52
2.3 Research design	p.53
2.4 Data collection procedures	p.56
2.4.1 EFL teaching context outlining via observation and interviews	p.57
2.4.2 Pronunciation activity with <i>Praat</i>	p.59
2.4.3 Online questionnaire	p.68
DATA ANALYSIS	p.69
3.1 Interview data	p.69
3.2 Pronunciation activity data	p.73
3.3 Questionnaire data	p.82
DISCUSSION OF RESULTS	p.89
4.1 Research question 1	p.89
4.2 Research question 2	p.91

4.3 Research question 3	p.93
4.4 Summary and conclusions	p.94
REFERENCES	p.99
Appendix A. Interview questions	p.107
Appendix B. Praat activity schedule	p.109
Appendix C. Images drawn from the PowerPoint slideshow	p.111
Appendix D. Item list to be recorded (words and questions)	p.113
Appendix E. Written guidelines: how to download Praat	p.115
Appendix F. Written guidelines: how to record, visualize and read your voice with Praat, and how to save your audio files	p.117
Appendix G. Guess the intonation	p.123
Appendix H. Intonation + Context	p.125
Appendix I. Written guidelines: how to conduct a comparison between the subject's initial recordings and the native speaker models	p.127
Appendix L. Intonation in discourse: Dialogue 1	p.129
Appendix M. Intonation in discourse: Dialogue 2	p.131
Appendix N. List of questionnaire items	p.133
Appendix O. Intonation contours by student 1	p.135
Appendix P. Intonation contours by student 2	p.137
Appendix Q. Intonation contours by student 3	p.139
Appendix R. Intonation contours by student 4	p.141
Appendix S. Intonation contours by student 5	p.143
Appendix T. Intonation contours by student 6	p.145
Appendix U. Intonation contours by student 7	p.147
Appendix V. Intonation contours by student 8	p.149
Appendix W. Intonation contours by student 9	p.151
Appendix X. Intonation contours by student 10	p.153
Appendix Y. Intonation contours by student 11	p.155
Appendix Z. Intonation contours by student 12	p.157

INTRODUCTION

Pronunciation instruction is generally cautiously welcomed by both learners and teachers of a second (L2) or foreign (FL) language. Yet, it is undeniable that good pronunciation plays a central role in determining the success or failure of oral communication (Setter & Jenkins, 2005). If on one hand, the unlikelyhood of attaining a native-like L2 phonological competence after a certain age, thus keeping a foreign accent, may be discouraging for some L2 speakers (Derwing & Munro, 2005; Munro, 2008; Busà, 2008a), on the other, instructors often feel they lack the necessary formal preparation and the right instruments to address such a multifaceted and complex aspect of L2 learning (MacDonald, 2002).

Given that mutual intelligibility is more important than native-like accent in verbal exchanges (Munro & Derwing, 2015), learners must be able to perceive the difference between their first language (L1) and the L2 phonological system, before producing speech not overly affected by L1 transfers. In this regard, several studies have shown that explicit pronunciation instruction may be pivotal in improving comprehension of problematic segmental as well as suprasegmental differences (De Bot & Mailfert, 1982; Derwing et al., 1998; Lord, 2005). Moreover, the need of constant feedback to enhance this kind of awareness has led L2 instructors to increasingly rely on the ever growing number of CAPT (Computer Assisted Pronunciation Training) technologies available (Fouz-González, 2015; Revell-Rogerson, 2021).

Among these technologies, speech analysis software have proved to be particularly effective in helping learners increase pronunciation accuracy, especially as far as prosodic features and intonation are concerned (Hincks, 2015). Although originally designed for experimental phonetics, these software provide learners with various types of audio-visual feedback, such as waveform, spectrogram or pitch contour displays (Busà, 2008b). Research results have demonstrated that a dual feedback modality leads to better gains than aural feedback only (De Bot, 1983; Olson, 2014b; Gorjian & Arvand, 2022).

Acoustic analysis software *Praat* (Boersma & Heuven, 2001), a free and versatile computer program developed at the University of Amsterdam, has long been used in the classroom setting at tertiary level, mainly to test its effectiveness on the perception and production of English intonation (Busà, 2008b; Le & Brook, 2011; Olson, 2014a;

Larassati et al., 2022). Examples of implementation of the so called visual feedback paradigm (VFP) by means of Praat are multiple and the outcomes obtained are overall positive. Nevertheless, only a limited amount of research has been conducted at the upper secondary level of education (Gorjian & Arvand, 2022) and no evidence has been found of investigations carried out within the Italian upper secondary school context.

To fill this lack of information, this study explores the feasibility of adapting the VFP, through a pronunciation activity with Praat, to the needs of two second year classes of a Venetian high school and examines whether Italian learners of such a young age can benefit from the visualization of their own productions of English questions as compared to native speaker models in the enhancement of prosody and more specifically of intonation. To this end, first, classroom observation and interviews with teachers are conducted, in order to gather valuable data on which to shape the VFP; second, a multi-phase practical activity is carried out, in which participants are asked to record a series of English questions characterized by specific intonation patterns and to re-record them, after comparison with audio-visual feedback of English models, trying to reproduce the target intonation; third, a satisfaction questionnaire is administered to assess participants' response both on the activity as a whole and on Praat usability.

The study results indicate that: 1) classroom observation and interviews with teachers provide a more favourable framework than the one outlined in much of the literature with regard to students' and instructors' attitude towards L2 pronunciation; 2) subject participants' final recordings of English questions exhibit substantial improvement in pronunciation of intonation patterns, as a result of visual-feedback comparison with native productions; 3) questionnaire findings reveal a broad appreciation for the pronunciation activity and, except for some perplexities on Praat technicalities, the software potential as a valid tool for English pronunciation instruction is acknowledged by the young respondents.

This dissertation will start with a description of some important acoustic features of the human voice in an attempt to highlight an often overlooked facet of pronunciation and prepare the reader to better grasp the topic of speech visualization technology. L2 pronunciation is then analyzed in all its aspects: large space is dedicated to Italian and English major phonological differences as well as to the principal technologies fostering L2 pronunciation learning. In the research project chapter, motivation of the study and research questions/hypotheses will be presented followed by the outlining of the setting,

the participants and the experimental design. A section is then dedicated to a detailed illustration of the three sources of data of this investigation, namely the observations/interviews, the Praat activity and the questionnaire. The gathered data will be separately analyzed and eventually discussed in relation to the three research questions.

LITERATURE REVIEW

This introductory chapter discusses some important aspects concerning foreign language pronunciation for pedagogical purposes. It begins by describing what lies behind speech from a vocal perspective. It continues by outlining the delicate role of pronunciation in second and foreign language teaching: first, with an explanation of what is meant by *foreign accent*; second, with an overview of the principal historical stages of pronunciation teaching; third, through the analysis of some key issues animating the theoretical debate. A section is then dedicated to a comparison between the Italian and English phonological systems, while the last part of the chapter presents an account of the main technologies applied to pronunciation training.

1.1 The Vocal Instrument

The frenzy of daily life is often characterized by a strong focus on external reality, both material and virtual, rather than the individual. Although in this reality communication has assumed an increasingly important role, the specific nature of the instrument at the basis of human communication, namely *voice*, continues to be largely ignored.

Unlike the aura of mystery surrounding some vital internal organs, such as the heart and brain, the human voice does not generally arouse the same appeal. During the highly sophisticated process, also known as *the speech chain* (Denes & Pinson, 2015), whereby two people succeed in exchanging verbal messages, speech sounds seem to have no secrets for the protagonists of the interchange. Not only will they effortlessly produce and perceive speech sounds, providing that they share the same language, but they will presumably recognize the outermost portion of the human vocal system, i.e., the mouth, as the primary sound source (Denes & Pinson, 2015).

Nevertheless, it is worth noting that, despite this mutual understanding, even minor distractions in controlling the vocal apparatus may lead to speech misinterpretations, such as the perception of erroneous sounds due to different regional backgrounds, or mistaking a statement for a question, or even conveying an obnoxious attitude in place of a friendlier one. Before such inconveniences, the human voice appears to acquire a more complex character.

With regard to its manifold nature, Laver (1994) argued that:

The voice is the very emblem of the speaker, indelibly woven into the fabric of speech. In this sense, each of our utterances of spoken language carries not only its own message, but through accent, tone of voice and habitual voice quality it is at the same time an audible declaration of our membership of particular social and regional groups, of our individual physical and psychological identity, and of our momentary mood. (p.2)

A propensity to consider speech as the product of a limited number of organs situated on a person's face is likewise present in numerous idiomatic expressions, such as *bite your tongue*, *my lips are sealed*, or *lie through your teeth*, to mention but a few (Denes & Pinson, 2015). However, in certain religious and pastoral communities located in some of the world's most remote places, there is a particular singing tradition handed down for generations, whose name puts the accent on another major sound producing area, in other words *throat-singing* (Levin & Edgerton, 1999). Among them, the seminomadic herdsman of Tuva, a region at the border of Russia and Mongolia, call their chant *khoomii*, meaning *throat* in Mongolian language. According to this peculiar singing practice, throat-singers are able to master their voice as if they had an actual musical instrument at their disposal without the need of any formal instruction. By controlling the rate and manner of vocal folds vibration and adapting the position of jaws and lips, they take full advantage of the resonance in the vocal tract, producing as a result the two distinct and simultaneous tones characteristic of throat-singing: a constant low one and a varying high one mainly used for echoing different sounds of nature (Levin & Edgerton, 1999).

This genuine awareness of Tuvan singers about their vocal system may appear a remarkable skill, however it should be noted that the human voice behaves very similarly to a man-made instrument even during spoken communication. Whenever a speech sound is generated, the same three basic components essential to music come into play: a sound source, namely the vibration of vocal folds in the larynx due to the airstream coming from the lungs, a resonator extending over the vocal tract and a sound radiator in the form of the mouth opening (Titze, 2008). The use of terms such as *vibration* and *resonance* suggests that, in the scope of linguistic sounds production, the physiological and acoustic dimensions are closely interrelated (Denes & Pinson, 2015).

With regard to *vibration*, this notion may be referred to two important moments of the sound generating process. The first moment takes place at the very beginning, when, as

previously mentioned, these two small veil-like elastic bands situated in the larynx, named *vocal folds*, begin to rapidly oscillate under the pressure of air flowing from the lungs (Titze, 2008). This open and close movement cuts the air column into multiple tiny puffs generating a basic sound whose *fundamental frequency* depends, in turn, on the folds' vibration rate (Denes & Pinson, 2015). Conversely, the second moment is concerned with the process outcome. After passing through the vocal tract, the resulting sound is projected from the lips to the listener's ears in the form of a composite *sound wave* whose pressure acts on the surrounding air molecules setting them in motion. More precisely, speaking in terms of wave motion physics, it can be said that the moving masses of the system (i.e., air molecules) assume an oscillating rhythm also known as *vibration*, which causes the sound wave with its distinctive acoustic features to be transmitted to the receiver. Among these features, *amplitude* (i.e., the maximum displacement of air particles from their rest position) and *frequency* (i.e., the number of complete oscillations of air particles within one second) are of major importance, in that they are perceived by human hearing as the sound's *loudness* and *pitch*, respectively (Denes & Pinson, 2015).

Resonance, on the other hand, deals with the sound shaping mechanism taking place between the two passages outlined above. While on a physical level it involves the laryngeal segment over the vocal folds and the *vocal tract* with its *articulators* (e.g., tongue, lips, soft palate), on an acoustic level it alludes to the sound-boosting effect to which these organs contribute to (Titze, 2008). As previously mentioned, the air stream is initially assigned a basic sound profile by the vocal folds' frequency vibration; together with the fundamental frequency, a blend of other tones called *harmonics* are generated, whose frequencies are multiples of the fundamental one (Levin & Edgerton, 1999). At this point, the vocal tract acts on the simple sound similarly to an air-filled tube, that is as a *resonator*. However, unlike a classic air-filled tube (e.g., an organ pipe) with its own natural frequency of vibration, the vocal resonator shows a far more sophisticated mechanism due to both its unique structure, characterized by a myriad of nooks and crannies, and its tendency to constantly alter its shape during speech. At each consecutive configuration, the vocal tract will only emphasize the raw sound's harmonics that share the same frequencies of its temporary set of natural (or *resonant*) frequencies. Ultimately, these resonant components, also known as *formants*, will be

recognized by the listener as the most prominent features of the final speech sound's identity (Denes & Pinson, 2015).

In light of these considerations, the human voice may appear as a multifaceted phenomenon that can be of interest to professional singers or phoneticians rather than common language users. In fact, in the context of everyday conversation, speech assumes the form of “a dynamic, continuously varying process” that is generally taken for granted by interlocutors (Denes & Pinson, 2015, p.143). Yet situations may arise in which speakers are not able to rely on the same kind of prompt response from their vocal system; communicating in a *foreign language* being a case in point.

Indeed, it is worth noting that all grammatical domains of a foreign language (i.e., phonology, morphology, syntax, semantics and pragmatics) may constitute a hindrance to smooth communication, since on a cognitive level, new linguistic categories from each of these areas need to be detected, organized and applied in order to encode and decode messages in the new language (Warren, 2013). However, with respect to *phonology*, most intuitively connected with the mechanics of voice, a considerable effort has to be made by the speaker in the perception and production of new phonological patterns. Focusing on production, the imitative effort interests both the individual sound units or *phonemes* as well as those features superimposed on larger sequences of sounds, also known as *prosody*, which convey further meaning and musicality to speech (Kulshreshtha et al., 2012, as cited in Nordquist, 2020).

Considering that the number of foreign languages in the world is approximately seven thousand, not to mention countless dialects, there is probably an equal number of corresponding phonological systems distinguished by a specific sound inventory as well as a distinctive rhythmic and melodic profile (Graffi & Scalise, 2013). This remarkable heterogeneity highlights the intrinsic adaptability of our vocal instrument and, moreover, legitimizes its composite character beyond the singing dimension.

Thus, as we realize the uniqueness of the communication means at our disposal, verbal interaction in a foreign language appears far from being a simple mimicry exercise. Rather, it seems likely that a greater awareness of the articulatory and acoustic properties of the means in question might prove beneficial to *pronunciation practice*.

The following sections explore how the interest towards second and foreign language pronunciation teaching and research has changed over time, assuming recently, due

especially to speech technology development (Hincks, 2015), a more holistic approach for which the articulatory and acoustic aspects of the vocal system are considered complementary as they can benefit one another.

1.2 Pronunciation and Second-Language Teaching

According to the Oxford English Dictionary, *pronunciation* refers to “the action of pronouncing a word or words ... especially with reference to a recognized standard”. This definition would suggest a certain rigour in speech within a specific locutionary context.

Nevertheless, it should be noted that each communicative act of any natural language reveals what is called *linguistic variability*. Inherent to speech, this phenomenon may occur in a myriad of forms; in the language itself, for instance, as regional dialects are often adopted in place of more standard varieties; in the speaker’s voice, reflecting general characteristics, such as age, sex, geographical origin as well as idiosyncratic features relative to physiological differences; and also, in the choice of a particular linguistic register, formal or informal, that can hide information about the speaker’s attitude or social status (Busà, 1995).

The term *pronunciation* would therefore seem to have a broader meaning which includes at least three factors: the *production* of speech, the *perception* of speech and *variability* (Setter & Jenkins, 2005). It is precisely the type and degree of *linguistic variation* to guide the hearer in determining whether the stream of segmentals (single sound units) and suprasegmentals (prosodic features) produced by the speaker could be regarded as belonging to their native idiom or as deviating from it to the extent that it would be decoded as non-native (Busà, 1999). Although, as it will emerge below, even a strong non-native pronunciation does not necessarily compromise the understanding of the message on the part of the listener (Busà, 2008a), so as to induce Munro (2008) to suggest that it should be accepted as “part of normal variation in human speech” (p.213), the importance of pronunciation in determining the success or failure of oral communication cannot be underestimated (Setter & Jenkins, 2005).

To date, there is a general consensus among researchers that learning a second language (L2)¹ after early childhood inevitably entails the development of a *foreign accent* (e.g., Derwing & Munro, 2005; Munro, 2008; Busà, 2008a). Interestingly, the ease with which *accented speech* is produced by non-native speakers seems to go hand in hand with the immediacy showed by native listeners (as well as non-native) in perceiving the oddity in the signal (Derwing & Munro, 2005), as it has been assessed by several studies on accent detection (e.g., Flege, 1984 and Munro et al., 2003, as cited in Munro, 2008).

Despite dealing with a widely recognized and attested phenomenon such as that of foreign accent, L2 pronunciation still remains a more sensitive aspect of L2 grammar compared to other components, for example syntax or morphology (Setter & Jenkins, 2005). One of the reasons is to be found in the *identity* trait intrinsic to language sounds and in the extent to which this trait might be projected outwardly. (Setter & Jenkins, 2005). Pronunciation key features and possible ways identity can influence L2 spoken interactions are explained below.

From an *acquisitional perspective*, foreign accent is generally thought to be the result of phonological interference from the first language (L1) to the L2 (e.g., Busà, 1995; Munro, 2008; Setter & Jenkins, 2005). To be more detailed, as highlighted by Major (1987), the process also known as *loan phonology* would primarily occur in initial phases of the acquisition process, when the learner, unable to elaborate new phonological categories for some unfamiliar L2 sounds, would tend to substitute them using similar L1 phonemes. Afterwards, the incidence of L1 *transfer* is expected to gradually decrease, leaving space to developmental processes based on linguistic universals, just as with the progression of other levels of language grammar (Major, 1987, as cited in Busà, 1995).

However, as Setter and Jenkins (2005) point out: “pronunciation tends to operate at a subconscious level ... and so is often not easily amenable to manipulation” (p.2). In this respect, it may be noted how identity could play a major role in the speaker’s resistance to change. In fact, some L2 learners will be more reluctant, though subconsciously, to

¹ In this dissertation the term L2 will be used with a general meaning, to refer to both *second language* (L2) and *foreign language* (FL) contexts. Such choice was made on the basis of recent considerations that confirm similarities in L2 and FL linguistic developmental processes (Balboni, 2014, as cited in Busà, 2021) and lack of clear boundaries, nowadays, between these two realms as for the ways in which language learning is approached (Diadori et al., 2009, as cited in Busà, 2021).

abandon their L1 pronunciation because of its bond with their geographical and cultural identity, while others will be more willing to adapt to new phonological patterns driven by reasons such as social and economic integration in the L2 community or professional advancement and desire for prestige in an international environment (Busà, 2021).

On the other hand, from a *perceptual perspective*, the way in which native and non-native listeners with different L1 may judge the degree of perceived accent is still a matter of debate (Busà, 2008a; Setter & Jenkins, 2005). Broadly speaking, poor pronunciation skills are likely to result in at least partially unintelligible speech (Busà, 1995), nonetheless much of the research up to now has revealed that *accentedness* is essentially independent of *intelligibility* and *comprehensibility*, the two other notions most currently linked to L2 understanding (e.g., Munro & Derwing, 1995a, as cited in Busà, 1999; Derwing & Munro, 1997, as cited in Derwing & Munro, 2005; Derwing, 2018). Thus intelligibility, namely “the extent to which a speaker’s message is understood by a listener” (Munro & Derwing, 1995a, as cited in Munro & Derwing, 2015, p.379) might be recognized regardless of the degree of difficulty in understanding the message, i.e., comprehensibility, and the degree of difference from some pronunciation pattern, i.e., accentedness (see Table 21.1 and 21.2 in Munro & Derwing, 2015 for a detailed description of possible combinations of the three dimensions).

While on a mere linguistic level evidence suggests that intelligibility problems would be more likely to derive from the *type* of speakers’ errors rather than from their *number* (Busà, 2008a), and in particular from the alteration of certain phonological categories (often vocalic) and prosodic structures (Busà, 1995), interestingly, the question of identity seems once again to interfere with the listener’s comprehension of an L2 speaker’s message. In this regard, Busà (2021) provides a thorough account of what might impact on intelligibility judgement. Here follow some main points: 1) *social expectations*, which arise whenever variability in the speech signal is linked to a particular social or ethnic group; according to this, while a negative attitude on the part of the listener (often based on prejudice) towards the people thought to be native speakers of the language producing that peculiar L2 variation could hinder intelligibility, a more positive attitude could facilitate it; 2) *esthetic expectations*, principally resulting from the speaker’s facial features, which may create a further link between the L2 signal and a specific social/ethnic and linguistic identity, yet again inhibiting or helping comprehension; 3) *familiarity*, probably the most important factor,

as it may somehow neutralize possible negative effects on understanding caused by expectations; it appears, in fact, that increased familiarity with a particular accent would gradually drive the listener to ignore variability in L2 productions, since differences from the target language would become much more predictable (Gibson, 1969, as cited in Busà, 1999).

Overall, this brief outline of pronunciation's various facets seems to underline its highly variable, subjective as well as delicate nature, apparently capable of interfering with intelligibility in L2 communication at any time. Thus, it is reasonable to assume that these qualities could represent some of the reasons why pronunciation is generally considered one of the most difficult parts of both L2 learning and teaching (Setter & Jenkins, 2005) and therefore regarded with distrust by many learners and teachers.

It might, moreover, be argued that similar prudence is also reflected in the historical path of L2 pronunciation research and instruction, in that the prominence given to pronunciation issues by researchers and practitioners has repeatedly and non-linearly changed in the past 150 years, that is, since phonetics and phonology of foreign languages became a growing matter of debate in the Western world (Murphy & Baker, 2015). Positively or negatively influenced by the different theoretical approaches to L1 acquisition (Busà, 1995) and their corresponding L2 teaching methods such as Audiolingualism and Communicative Language Teaching (Derwing, 2018), pronunciation has long been viewed as the *neglected* element of L2 empirical research and classroom practice (Derwing, 2010), to the point of being labelled by numerous experts the *Cinderella* of second language acquisition (see for example Kelly, 1969, as cited in Levis, 2019; Celce-Murcia et al., 2010, as cited in Derwing, 2018).

Despite that, an increased interest in L2 pronunciation-related themes is being witnessed today. Within an instructional framework that is almost entirely communicatively oriented (Busà, 2021), L2 pronunciation teaching practices and materials are now largely informed by a growing body of empirical research findings and supported by a more efficient use of technology (Murphy & Baker, 2015). This latest change of course has led some researchers to distance themselves from the Cinderella metaphor, announcing instead the beginning of a new phase of pronunciation instruction (see Murphy & Baker, 2015; Derwing, 2018; Levis, 2019; Busà, 2021).

To gain a better understanding of present-day L2 pronunciation status, the following insights will provide a brief overview of some key historical stages in L2 pronunciation research and instruction and an analysis of two main leitmotifs of the field, namely the *nativeness/intelligibility* and *segmentals/suprasegmentals* dichotomies. Additionally, it is appropriate to clarify that the primary focus will hereafter be on English L2 pronunciation, as this is the L2 in which most studies have been undertaken (Derwing, 2018) as well as the foreign language (FL) examined in the experimental part of this dissertation. However, because all second languages appear to share basic phonological acquisition patterns (Busà, 1995), subsequent considerations should for the most part be regarded as universal.

1.2.1 The Four — Plus One — Waves of L2 Pronunciation Teaching History²

L2 pronunciation may be deemed to have followed a separate path to that of other domains of L2 grammar, primarily for two reasons. First, up until the 1800s, the widespread application of the so-called Grammar Translation Method in the teaching of foreign languages, with its focus on written texts and their translation (Murphy & Baker, 2015), had in all likelihood promoted a detailed study of areas such as lexicology and syntax, at the expenses of pronunciation (Levis, 2019). Second, the intangible quality of voice had made it impossible to gather direct evidence of pronunciation teaching activities until audio recorders were implemented, that is around the mid-twentieth century (Murphy & Baker, 2015).

The narrowing of this time gap began in the 1850s, when the work of a small group of innovators from Europe and the United States led to a change in priorities in L2 and FL teaching. Rather than focusing on the written word, these specialists prioritized the spoken language, claiming the importance of spontaneous speech practice and advocating an approach to pronunciation teaching based on an *intuitive imitation* of the auditory input (Murphy & Baker, 2015). Although these innovations struggled to reach school environments, remaining mostly confined within academic circles, they served as the basis for the following Reform Movement (see below) and can therefore be said to form what Murphy and Baker (2015) have named the *first wave* of pronunciation teaching history.

² This historical synthesis follows, in broad terms, the four-part time division proposed by Murphy and Baker (2015) in their *History of ESL Pronunciation Teaching*.

Starting in the 1880s, further emphasis was placed on L2 pronunciation. Its role in language teaching became fully acknowledged through the initiative of several European phoneticians, whose primary pedagogical intent was to transfer the findings of phonetics into the language classroom (Murphy & Baker, 2015). Joined in the *International Phonetic Association*, one of their greatest achievements was the development of the *International Phonetic Alphabet* or IPA (Deterding, 2015), which could be described as a set of symbols “capable of representing the full inventory of sounds of all known languages” (Setter & Jenkins, 2005, p.2). The use of IPA as a tool for pronunciation instruction — which continues to this day — is part of a broader teaching approach termed *analytical*, that was first introduced by the same late-nineteenth-century *Reform Movement* exponents. As opposed to the intuitive way, teaching pronunciation analytically entails an explicit action on the part of the teacher aimed at raising learners’ phonological awareness. Through the application of tools such as a phonetic alphabet, articulatory descriptions of unfamiliar phonemes, contrastive pronunciation exercises, the well-trained teacher should provide learners with the phonetic/phonological background useful to improve their natural skills for imitation of models, implicitly admitting the existence of an intuitive component too. (Celce-Murcia et al., 2010, as cited in Murphy & Baker, 2015). This innovative pedagogical framework, consisting of the coexistence of two instructional orientations (analytical and intuitive) to underpin pronunciation classroom practice, represents a clear change compared to the past and ensures to the Reform Movement period the title of pronunciation teaching’s *second wave* (Murphy & Baker, 2015).

Over the following decades, one of the most prominent examples of the lasting impact of this mixed approach was *Audiolingualism*, that became very popular between the late 1950s and the early 1970s (Derwing, 2018). This method emphasized the value of pronunciation in L2 teaching and utilized a series of systematic instructional techniques designed to minimize the risk of errors on the part of the learner from the very beginning (Derwing, 2010). Audio-lingual principles may be regarded as reflecting the general belief in those days according to which, using the words of Levis (2019), “the perfect learning of a new language was not only possible but was the logical outcome of correct teaching and learning strategies” (p.7). Supported by the teacher’s explicit-deductive phonological explanations (Mahalingappa & Polat, 2019, as cited in Busà, 2021) and the earliest forms of language learning technology, such as language

laboratories and portable cassette tape players, L2 students would practice discrimination, articulation and memorization of sounds through a wide amount of controlled repetition exercises (Murphy & Baker, 2015).

Moving across the ebb and flow of pronunciation in L2 teaching history, the following twenty years witnessed an increasing lack of interest in L2 phonological competence. While the theoretical foundations — i.e., structural linguistics and behaviourist psychology — of Audiolingualism and many other early-twentieth-century L2 teaching methods had started cracking, new diverse theoretical perspectives were emerging (Levis, 2019). Initially, at the turn of the 1960s and 1970s, the *innatist-cognitive* approach, based on Chomsky's theory, displaced the earlier school of thought, according to which language acquisition (as well as L2 acquisition) was seen as a mechanical process of habit formation achievable through mimicry and constant repetition (Busà, 1995). Instead, emphasis was given to learners' innate cognitive abilities, which would enable them to gradually proceed with the learning of grammar by processing easier central aspects first and leaving more complex and peripheral aspects afterwards (Busà, 1995). Thus, due to this attention to grammar, L2 teaching methods relying on innatist-cognitive principles aimed at formal correctness of sentences rather than speaking skills (Busà, 2021). Concurrently, the *critical period* hypothesis, whereby individuals would be unable to acquire a native-like accent after puberty, regardless of their L2 proficiency level, raised further questions about the usefulness of pronunciation instruction (Busà, 2021). Finally, at the turn of the 1970s and 1980s, *Krashen's theory of L2 acquisition* again diverted attention from the teaching of L2 phonology. According to his acquisition-learning hypothesis, there would be no formal instruction capable of matching the effect on language development of unconscious exposure to linguistic input (Busà, 1995; Derwing, 2018).

It is within this unfavourable climate for L2 pronunciation that, by the start of the 1980s, the *communicative approach* appeared on the scene. Drawing upon functionalist theories of language acquisition, which (as opposed to cognitive perspectives) considered the social and communicative function of language of paramount importance in the development of linguistic structures, relegating grammar to a subordinate role (Busà, 1995), it can be observed how, at the early stages of what is referred to as Communicative Language Teaching (CLT), available evidence of pronunciation instruction was almost nonexistent (Derwing, 2018). The reason is to be found in the

fact that traditional forms of pronunciation teaching (based on audio-lingual models) were then regarded as incompatible with CLT's main objectives, namely *communication* and *fluency*. Levis (2019) provides a thorough account of this "mismatch", claiming that teaching communicatively entails a clear separation from the search for *accuracy* and *nativeness* typical of more traditional methods; priority to discrete phonemes, use of systematic repetition and error correction (especially pronunciation errors) were deemed detrimental to speakers' spontaneous fluency. To put it concisely, it could be argued that teaching pronunciation communicatively seemed at the time unconceivable (Levis & Sonsaat, 2017, as cited in Busà, 2021).

Nevertheless, by the mid-1980s, the innovative work of several academic researchers investigating effective ways to teach English L2 pronunciation (EL2) within a communicative approach, had major impact on materials developers and as a consequence classroom teachers, giving rise to the *third wave* of pronunciation teaching history (Murphy & Baker, 2015). With the support of several new publications dedicated to the theme, such as classroom textbooks, activity recipe collections and teacher preparation texts, EL2 educators could now devise activities in which pronunciation would boost the pragmatic use of contextualized language (Murphy & Baker, 2015). Accordingly, the notion of native-like accent was replaced by that of *intelligible pronunciation*, which became a core element of communicative competence (Busà, 2008a). Thus, it may be stated that "the key role of pronunciation in improving the learner's oral skills, and in contributing to ensuring the success of oral communication" was eventually acknowledged by the advocates of CLT (Busà, 2008a, p.166).

Although the decade between the mid-1980s and the mid-1990s constituted a first example of how research and greater networking opportunities (e.g., serial publications, conventions, professional associations) could benefit pronunciation instruction (Murphy & Baker, 2015), it should be highlighted that experts' contributions at that time were not supported by empirical investigations but relied for the most part on *intuition* and personal experience as pronunciation instructors (see, for example, Derwing, 2010, reporting on possible flaws in her attempts to teach EL2 pronunciation to a group of Vietnamese speakers).

It is precisely the development of an *empirical research* base underpinning pronunciation teaching to characterize the field's most recent *fourth wave* (Murphy &

Baker, 2015). Starting from the mid-1990s, empirical researchers' work has increasingly informed pronunciation instruction trying to provide educators with more exhaustive answers to some areas still considered problematic, such as: what priority should be given to different aspects of L2 phonology, what target language model should be adopted (especially in the case of English) or what is the level of teachers' preparation in the subject matter and possible ways to improve it (Derwing, 2018). Moreover, with regard to classroom practice, it may be noted how today pronunciation instruction is essentially distinguished by a constant combination of previous approaches; in other words, intuitive, analytical and communicative techniques are used together to enhance learners' phonological competence (Murphy & Baker, 2015).

In conclusion, assuming that the advent of a *fifth wave* of pronunciation teaching history may to date only be hypothesized, as suggested by Murphy and Baker (2015), a further step towards a higher visibility of L2 pronunciation seems to have been made. Although some researchers argue that L2 pronunciation teaching has not yet incorporated knowledge made available through research effectively enough (see, for example, Derwing, 2018), it might be supposed that this aspect is, at least partly, being compensated by increasing access to online resources. The endless amount of *authentic materials* offered by the Internet together with more user-friendly software and apps have possibly made a difference, in that both teachers and students seem, nowadays, to approach to L2 pronunciation more confidently and successfully than in the past (Busà, 2021). Noteworthy is that a continuous analysis of new technological tools (aiming at exploring pros and cons) and their adjustment to instructors' context of teaching or learners' specific needs, might represent one of the main traits of this incoming wave. Ultimately, the constant progress achieved over recent years and the ongoing efforts towards an effective synthesis between L2 pronunciation teaching's multiple components and latest instructional innovations, would seem to support the view that the Cinderella metaphor has lost enough strength to be replaced by, as Derwing (2018) suggests, a more appropriate metaphor: that is, pronunciation may now be considered the *Belle of the Ball* of second language instruction.

1.2.2 Core Dichotomies of L2 Pronunciation Instruction

As the latter part of this historical review highlighted, once preliminary communicative ways of L2 teaching revealed that extensive exposure to input did not necessarily mean

fluency in communication (this way calling into question Krashen's monitor model, as cited in Busà, 1995), a new awareness towards the role of *formal instruction* became widespread. Dickerson (1987), in particular, argued that explicit formal instruction could positively affect L2 acquisition both in the short and long term, in that consciously manipulated rules would act on the interlanguages' rule systems, successively emerging on a subconscious level, leading gradually to a greater accuracy in speaking (Dickerson, 1987, as cited in Busà, 1995). Although, this formal approach concerned all areas of language development, the need of addressing pronunciation explicitly was particularly strong due to its role played in oral interactions. As a result, some long unresolved issues relative to pronunciation instruction began again to be tackled, this time within the new communicative framework. In order to analyze them, reference can be made to two meaningful pairs of opposite terms, which in turn encompass a whole array of no less important questions.

a) The first opposition is that between *nativeness* and *intelligibility*.

Which of these two principles is to be prioritized in the teaching of L2 pronunciation has long been a matter of debate (e.g., Sweet, 1900, as cited in Munro & Derwing, 2015; Abercrombie, 1949a and 1949b, as cited in Murphy & Baker, 2015). If until the advent of CLT, there was consensus among scholars and practitioners that sounding like a native speaker could be considered the ultimate goal for L2 learners, the communicatively oriented teaching context required a shift in perspective (see subsection 1.2.1). Having as primary objective fluency in communication, CLT could not aim at eradication of foreign accent but rather at facilitating the development of speech patterns that would not hinder oral interactions, despite a certain degree of accentedness (Busà, 2008a; Munro & Derwing, 2015).

It could be said that, since Lenneberg's *Critical Period Hypothesis* had identified a biological limit to native-like L2 attainment (arguably related to cerebral language lateralization), indicating pre-puberty as the period in which L2 learners still have full access to their L1 acquisition strategies (Lenneberg, 1967, as cited in Busà, 1995 and Setter & Jenkins, 2005), the assumption that adult speakers could rarely, if ever, reach a native-like L2 phonological — as well as morphosyntactic — competence had been increasingly accepted and supported by a growing body of literature (e.g., Tarone, 1978, Coppieters, 1987 and Long, 1990, as cited in Busà, 1995). In 2009, for example,

Abrahamsson and Hyltenstam published a study on the “nativelikeness” of 200 Spanish/Swedish bilinguals that well illustrated the effects of L2 acquisition onset age. Spanish was the participants’ L1 and their onset age of Swedish acquisition spanned from 1 to 47. All subjects were selected for their high L2 proficiency and because they self-evaluated their Swedish pronunciation as native-like. Despite that, only early learners’ tests scores (i.e., those produced by participants who started to acquire the L2 before age 12) could be regarded as equivalent or close to native speaker’s scores (Abrahamsson & Hyltenstam, 2009, as cited in Derwing, 2010).

Although this and other investigations (e.g., Fathman, 1975 and Oyama, 1982, as cited in Busà, 1995) seem to underpin Lenneberg’s hypothesis, the precise time frame during which learners would be deemed capable of reaching native proficiency has long been a controversial matter (for different age limits see Lenneberg, 1967 and Flege, 1988, as cited in Busà, 1995). In addition to this, several other factors have been called to account for different intensity of accented speech between child and adult speakers (Flege, 1987, as cited in Setter & Jenkins, 2005). These include: amount and length of *exposure to L2 input* (Walburg et al., 1978, as cited in Busà, 1995), degree of *formal instruction* (Flege & Fletcher, 1992, as cited in Busà, 1995) and a series of other variables related to the speaker’s personality, such as *language aptitude* (Krashen & Terrel, 1983, as cited in Busà, 1995), *motivation* and *anxiety* (Krashen, 1982, as cited in Busà, 1995). These variables would be added to the age factor impeding, in the vast majority of cases, a full acquisition of L2 pronunciation. It may be asserted that, adult speakers, given both the cognitive and personal effort, would naturally rely on an intelligibility principle, as they would tend to improve their L2 phonological competence enough to successfully communicate (Busà, 1995). By contrast, as Derwing and Munro (2005) prudently observe: “most learners who strive for nativeness are likely to become disheartened” (p.384).

Furthermore, another aspect worth noting is that the pedagogical need to teach pronunciation within an intelligibility framework has gradually increased over the years due to *globalization*. Its impact on people’s mobility has, indeed, made moving and travelling abroad a very common phenomenon. As a consequence, in the case of some widely spoken languages, such as English, Spanish and French, the fact of being used as *lingua franca* by a large number of people with different linguistic backgrounds has

increasingly blurred the line between what is referred to as native accent and what is perceived as foreign (Munro, 2008; Busà, 2021).

Specifically with regard to *English*, its global role in communication has put a spotlight on the concept of *intelligibility among non-native speakers* (Busà, 2008b). This, in turn, raises two questions. Firstly, whether traditional *speech models* adopted to teach pronunciation are still adequate or a wider variety of models is needed; in other words, British and American English would no longer be able to fully meet today's learners' needs, as they will most likely interact with people having the most varied English accents. In this scenario, pronunciation instruction would be called to rely on exposure to multiple native and non-native models so as to help learners to handle the great amount of input they could encounter (Busà, 2008b; Derwing, 2018; Busà, 2021). Secondly, whether in a context characterized by a constant emphasis on communication skills enhancement as well as limited time for pronunciation instruction, educators might opt for a targeted selection of certain aspects of L2 phonology to be taught (Derwing, 2018). In this case, higher priority should be given to those features "which are deemed particularly detrimental to intelligibility if not corrected" (Busà, 2008b, p. 114). This priority assessment may be conducted on the basis of several criteria, as the following examples suggest: by focusing on *high functional load* phonemic contrasts (Derwing, 2018), meaning those sound units oppositions differentiating a considerable number of minimal pairs and as such causing greater intelligibility problems than phonemes with lower functional load (Catford, 1987, as cited in Derwing, 2018); by evaluating the right balance between the teaching of segmentals and suprasegmentals (see below), in that, given their interwoven nature, raising learners' awareness on both phonological levels would seem to positively affect L2 perception as well as production (Setter & Jenkins, 2005; Busà, 2008b); or even starting from a classroom needs analysis, especially in the presence of different linguistic backgrounds, as students' L1 may contribute to a different order of acquisition of L2 phonological categories (Busà, 1995; 2021).

All things considered, it may be argued that, at least with respect to the so-called international languages, the attainment of a faultless pronunciation becomes an "unnecessary" goal for both learners and educators (Munro, 2008). As shown by specialists working in the scope of English as International Language, the rather outdated concept of *accent reduction* ought to be replaced by a more inclusive principle

of *accent addition*, which means that, once basic mutual intelligibility has been achieved, “L2 pronunciation features” should be added to “learners’ repertoires in accordance with their needs and preferences”, this way helping individuals to possibly preserve their L1 identity (Setter & Jenkins, 2005, p.6).

b) The second major opposition is that between *segmentals* and *suprasegmentals*.

In trying to identify which elements of L2 accent are more likely to affect intelligibility and comprehensibility, communicative contexts of teaching contributed to widen some researchers’ scope of investigation to an area of L2 pronunciation that had long been overlooked, the components of which may be summarized under the name of *prosody* (Busà, 1995; Busà, 2008a; Zielinski, 2015).

Prosodic features or suprasegmental phenomena may be considered a challenging aspect of language in general, in that their extremely variable nature makes them hard to be described consistently even among native speakers; it would be highly unlikely to gain the same feedback regarding temporal and melodic properties such as *stress*, *rhythm* and *intonation* (Busà, 2008a). Nevertheless, given the pivotal role played by prosody both in L1 development (being the first phonological element to be perceived by infants as well as the first linguistic means of self-expression) and adult speakers interactions (contributing significantly to meaning), it comes with no surprise that its relevance has been acknowledged even as regards L2 perception and production (Busà, 2008a).

As shown by a considerable amount of literature on EL2, prosody has been proven to influence listeners’ perception and intelligibility judgements of L2 pronunciation (Setter & Jenkins, 2005; Zielinski, 2015; Busà, 2021). Particularly, native English speakers have been found to be susceptible to errors relative to certain suprasegmental traits as they would strongly impact on the correct parsing of speech input (Setter & Jenkins, 2005). Among these, worth mentioning are: *word stress*, that, as highlighted by Cutler (1984), seems to be fundamental for correct word retrieval during speech perception, as it constitutes an essential part of the phonological profile of any lexical item stored in a native English speaker’s mental lexicon (Cutler, 1984, as cited in Setter & Jenkins, 2005); *stressed syllables*, which more broadly are shown by Grosjean and Gee (1987) to be responsible for triggering lexical research (Grosjean & Gee, 1987, as cited in Setter & Jenkins, 2005); and *nuclear stress* (closely tied to the first two items as well as to intonation, see below), that by marking phrases’ most prominent syllables, provides

long stretches of speech with the proper melodic contour (Levis & Wichmann, 2015). Evidence of the impact of nuclear stress on intelligibility may be found, for instance, in Hahn (2004), whose study compared listeners' perception of the same accented speech (Korean accent of English), whose nuclear stress was manipulated in various ways. Results show that the group presented with the correct nuclear stress talk had less difficulty understanding, compared to the other stress conditions (Hahn, 2004, as cited in Derwing & Munro, 2005).

After all, stress patterns are at the basis of English *speech rhythm*, which by no coincidence is often described as *stress-timed*, meaning that intervals between stresses tend to be equidistant time-wise, regardless the number of unstressed syllables in between (Busà, 2008b). Having acknowledged the importance of language specific rhythmic rules in speech segmentation (Cutler, 1993, as cited in Setter & Jenkins, 2005), English listeners will be therefore expected to segment what they hear based on word stress, while EL2 speakers whose L1 displays a diverse rhythm (i.e., syllable-timed or mora-timed) will most probably transfer its timing patterns onto their English productions, creating “an anomalous rhythm” that might hinder the success of verbal communicative exchanges (Adams, 1979, as cited in Setter & Jenkins, 2005, p.8).

Furthermore, emphasis has been placed on *intonation*, namely the variation of *pitch* in the voice, since it has been shown that it may serve important linguistic functions (Levis & Wichmann, 2015). While traditional models were primarily concerned with *attitudinal* meanings of English intonation, as exemplified by the four pitch levels system elaborated by Pike (1945), who claimed that each combination of pitch levels could express a specific attitude (Pike, 1945, as cited in Levis & Wichmann, 2015), more recent approaches have focused on its *grammatical* and *discourse* functions. Accordingly, it has been argued that, through intonation, speakers would be able to: indicate *phrase boundaries*, as each phrase generally coincides with a so-called tone unit; assign *prominence* to some discourse elements, by placing an extra (intonational) stress³ on certain syllables, for e.g., to mark new information in contrast to given information; control *conversation* mechanisms, such as turn-taking or introduction and conclusion of topics, by choosing a specific pitch contour, for e.g., a final rising or

³ *Nuclear stress* and *intonational stress* are often regarded as overlapping suprasegmental phenomena. The meaning can change on the basis of the acoustic parameters being analyzed. To be considered that, the main acoustic correlates of stress and intonation are: *fundamental frequency* (F0), *duration* and *intensity*. Generally, while *stress* is associated with all three physical sound features, *intonation* is mostly related to F0 patterns (Denes & Pinson, 2015).

falling contour to indicate openness or closure respectively (Setter & Jenkins, 2005; Busà, 2008b; Levis & Wichmann, 2015). Thus, it may be assumed that EL2 speakers who tend to transfer different intonation patterns from their L1 or even use other strategies typical of their native language to express the same functions, will be likely to convey incorrect attitudinal, grammatical or pragmatic meanings, causing again possible breakdowns in intelligibility (Busà, 2008b).

Traditionally, the teaching of segmental features have been regarded as the primary goal of L2 pronunciation pedagogy. To this day, many instructional techniques still aim at drawing learners' attention to the articulation of individual sounds, as in the case of the support provided by articulatory diagrams and charts to explain the positioning of articulators during consonants and vowels production or the use of minimal pair drills and auditory discrimination exercises to enhance learners' perception of L2 phonemes (Busà, 2008a; 2008b). However, having a good understanding of the L2 phonemic inventory has been proven to be insufficient for the successful articulation of sounds in spoken language. Longer stretches of speech may, in fact, be conceived not only as glued together by prosody but also as shaped by a combination of suprasegmental phenomena, which are very likely to modify the basic articulatory and acoustic properties of distinct sound units (Busà, 2008a). A case in point is the influence that word stress, rhythm and intonation exert on English vowels pronunciation, which leads to a constant alternation of full and reduced vowels. In the attempt to reproduce the same patterning, learners whose native sound system features alternative prosodic properties (as in the case of syllable-timed languages) could show a tendency to over-articulate vowel phonemes, producing the full form in place of the reduced one (Busà, 1995; 2008b).

In light of the above, continuing to view segmentals and suprasegmentals as separate entities of L2 phonology might prove to be counter-productive for pronunciation instruction. As pointed out by Busà (2008b), because these two aspects “overlap and contribute to each other in many important ways, in pronunciation classes they should be taught together” (p.118). In this perspective, Derwing et al. (1998) compared the effects of three types of pronunciation teaching on the productions of three groups of EL2 learners. The following approaches were used: focus on segmentals; attention to prosody in the scope of interactive discourse; and no specific pronunciation instruction (assigned to the control group). Overall, the three variables undergoing native speakers'

rating, i.e., comprehensibility, accentedness and fluency, showed improvement in the two pronunciation-focused groups compared to the no instruction group. However, while in the sentence reading task both experimental groups exhibited improvement in comprehensibility, in the extemporaneous narrative description task only the group whose training was centered on prosody showed better ratings concerning comprehensibility and fluency, meaning that in conversational contexts segmental production may benefit from insights into suprasegmental features.

These findings further highlight the advantages of integrating the two levels of phonology in L2 classroom practices (Murcia et al., 2010, as cited in Zielinski, 2015). Nevertheless, a need is felt to address, in the first place, the long-standing issue of instructors being often reluctant to teach pronunciation, let alone prosody (Busà, 2008b; Derwing, 2010). What emerges from several surveys conducted among EL2 teachers (a list of studies can be found in Derwing, 2018) is a sense of *inadequacy* towards this area of instruction due to numerous factors, such as the small presence of pronunciation in curricula, the unavailability of suitable material and, more notably, the *lack of formal preparation* (see MacDonald, 2002 for more detail). These aspects seem to affect even more severely the teaching of suprasegmentals, in that, in order to provide learners with comprehensive explanations and precise corrective feedback, instructors should deepen their knowledge of both the L1 and L2 prosodic systems focusing on a wide range of issues which contribute to the highly complex character of the subject, as for example: articulatory and acoustic properties, communicative functions and contextual factors (Busà, 2008a). In this respect, teachers again report a lack of training and didactic materials (mostly segment-based) together with missing guidelines for a targeted selection of L2 prosodic features to be taught (Busà, 2008b). Therefore, it may be hypothesized that further efforts are needed in this regard on the part of curricula designers, material developers and those responsible for teacher-training programs, who, besides enhancing instructors' pronunciation-related pedagogical skills, should provide them with the means "to critically evaluate materials and curriculum on the basis of empirical research" (Derwing & Munro, 2005, p.391).

To conclude, what seems to emerge from recent literature on L2 pronunciation is a propensity for a *broader framework* of instruction, reflected both in a greater tolerance for a wider range of accents and a gradual weakening of the debate concerning segmentals and suprasegmentals (Zielinski, 2015). Despite some teachers' concerns and

methodological flaws in educational practice, it is reasonable to assume that individuals can indeed benefit (especially passed puberty) from a type of explicit education that aims at improving comprehension of the differences between the L1 and L2 on both phonological levels (segmental and suprasegmental). This, together with personalized feedback (whenever possible) and contextualized communicative practice, would yield positive outcomes not only in speaking abilities, but also in other domains of the grammatical competence as well as in listening and reading (Busà, 2021).

The next section deals with the specific case of Italian-speaking learners of EL2. A comparison between the two phonological systems will highlight which L2 segments and prosodic features are more likely to hamper the efforts of Italian students to communicate in English.

Afterwards, in section 1.4, the focus will shift to the many solutions technology can provide in the attempt to overcome those issues related to L2 pronunciation teaching reviewed thus far. Among these solutions, the potential offered by the software *Praat* will be explored in detail.

1.3 Pronunciation of Italian Learners of English

Nowadays, the greater accessibility to audiovisual material allows individuals to find plenty of examples of Italian accents in English. To the more classic gangster-like movie characters, a series of today's politicians, businessmen, singers and cooks, to cite a few, can be added, whose videos available all over the Internet showing their speech performance in English, may represent a valuable source of information on the ways in which Italian pronunciation features (i.e., L1) can manifest themselves in English productions (i.e., L2), possibly impacting on speech understandability. It is not uncommon for these figures to opt quite consciously (given an acceptable L2 proficiency) for maintaining their accent intact, aiming at receiving recognition from a wider audience (Busà, 2021). In such cases, their talk combined with their specific professional role would evoke positive images of *Italianness*; by way of example, a fashion designer's accent recalling the imagery of the Italian clothing style.

Negative phonological *transfer* of this kind — i.e., the L2 speaker's tendency, in presence of asymmetrical structures, to activate the closest L1 sound in place of the correct target phoneme (Busà, 1995) — would lead to errors that are unlikely to hinder message comprehension. Viewed from a pedagogical perspective then, these features

might not be given priority in the selection of “ideal candidate[s] for instruction” to Italian learners of English, at least at an early stage (Derwing, 2010, p.27).

Instances of such errors include the following: pronunciation of English / r /⁴ as alveolar trill [r] instead of as postalveolar approximant [ɹ] (Busà, 2021); pronunciation of unaspirated voiceless stop consonants [p, t, k] in place of the English aspirated ones [p^h, t^h, k^h] (Wheelock, 2016); both these substitutions deal with English *allophones*, meaning that the use of one phonetic variant over the other does not imply the creation of a new word (Deterding, 2015). Moreover, the insertion of an epenthetic vowel similar to the *schwa* sound [ə] on completion of English words ending in a consonant, in particular stop consonants (Busà, 2008b; Wheelock, 2016); this phenomenon being a consequence of Italian *phonotactic constraints*, such as the preference for open syllables (ending in a vowel) over closed syllables (ending in a consonant) in word-final position (Busà, 1999; Brown, 2015), as well as a reflection of Italian rhythmic patterns (see Wheelock, 2016, p.47 for an example of vowel epenthesis’ effect on rhythm).

Thus, the urge to correct these error types will mostly depend on learners’ will and personal needs (e.g., social, professional, etc.), since they can contribute to the — sometimes detrimental — stereotyped image of Italian speakers of English (Busà, 2008b). Instead, as for ordinary contexts of instruction, having communication as primary goal should imply that, in designing EL2 pronunciation activities tailored on Italian students, instructors would focus on those sound features that might constitute “a source of difficulty” for the learner and at the same time “a source of problems for the listener” (Busà, 1999, p.70). As it will be observed below, research may offer valuable insights on the underlying causes and effects of these important phonological deviations, helping teachers in the discernment of learning material that aims at increasing their students’ intelligibility.

1.3.1 Italian and English Phonological Systems: Segmental Features

Although at first glance, Italian and English language seem to share a few essential aspects, such as SVO word order, a very similar alphabet and numerous cognates, the illusion of some Italian native speakers of being able to effortlessly learn English could

⁴ According to IPA conventions for brackets: *slashes* // are used to enclose consonant and vowel *phonemes* belonging to the inventory of sounds of a specific language; *square brackets* [] indicate language-independent sounds or *phones* (Deterding, 2015).

easily vanish as soon as they are confronted with the large number of differences in pronunciation (Wheelock, 2016).

In this regard, it is noteworthy that, along with important dissimilarities at both the segmental and suprasegmental level (analyzed in more detail below), a further hurdle might be represented by *orthography*, more specifically phonemic orthography. In fact, while Italian presents a high degree of correspondence between graphemes and phonemes, which makes the system more predictable, in English such correlation is much weaker (Brett, 2004; Wheelock, 2016); consider, for instance, the [ʃ] sound, writable in English in 14 different ways (Celce-Murcia et al., 2010, as cited in Wheelock, 2016). Therefore, Italian learners' tendency to pronounce words as they are spelled could be expected to give rise to negative transfer (Wheelock, 2016).

With reference to *segmental differences*, the discussion may be divided into two parts: one dedicated to consonant segments and the other dedicated to vowel segments. This choice is based on the fact that we are dealing with two sound categories very different from each other (Busà, 1995) and that, at least in this specific learning context (i.e., Italian as L1 and English as L2) vowels seem to pose more complex issues when compared to consonants in both learners' production and native speakers' perception of foreign accent, thus requiring greater attention.

Starting with *consonants*, by comparing the two languages' phonemic inventories, what emerges is the following (Deterding, 2015; Wheelock, 2016):

- Italian and English have almost an equal number of consonant phonemes, 23 and 24 respectively;
- a large number of phonemes have the same *manner* and *place of articulation*, for e.g., / p, f, m, n /; some differ only in place of articulation, such as Italian / t / being a dental stop, while English / t / an alveolar stop;
- some English phonemes, i.e., / θ, ð, h / are not part of the Italian phonemic inventory.

As for the last point, it may be assumed that the total absence of these segments from the L1 inventory could represent a further challenge in the process of *interlanguage development*, as it would be harder for Italian learners to find, at first, a similar L1 structure to replace the missing phone. These assumptions, drawing from principles of *contrastive* and *error analysis* (Weinrich, 1953 and Corder, 1967 respectively, as cited

in Busà, 1995), collide with other theories according to which L2 speakers would tend to produce new (unfamiliar) sounds with greater ease than similar sounds, since the non-occurrence of interlanguage phonemic identification would more readily lead to the creation of a new sound category (Flege, 1987, as cited in Busà, 1995). However, evidence from error analysis would seem to confirm some systematicity in the misformation of unfamiliar consonant segments. Wheelock (2016), for instance, has shown how, from an error analysis conducted on the phonetic transcriptions of 27 Italian EL2 learners (all reading aloud the same English passage), the pronunciation of English dental fricatives / ð / (voiced) and / θ / (voiceless) as alveolar stops [d] (voiced) and [t] (voiceless) was among the six most recurring consonant misformations detected, taking the third and the fourth place respectively as to the average group misformation rate. These results, although concerning a small sample size, by confirming a certain degree of identification of the two English phonemes with the closest Italian counterparts, suggest a possible area of difficulty on which instructors could focus on.

Turning now to *vowels*, there is a general consensus among researchers and practitioners that this group of segments is to be regarded as more problematic compared to consonant segments (see for e.g., Busà, 1995 and Piske et al., 2002 for English vowel production by Italian speakers; Saito, 2007 for English vowel production by Japanese speakers; Sebastián-Gallès & Soto-Faraco, 1999 for perception of Catalan vowels by Spanish speakers, as cited in Piske et al., 2002). Brett (2004) goes as far as to state that “the acquisition of new vowel sounds is one of the most challenging tasks for the adult learner of a second language” and that “it is also an aspect which requires a great deal of time for the learner and individual attention from the teachers” (p.103). Such claim would seem to hold true also in the learning/teaching context under consideration.

What emerges from a comparison between the Italian and English vowel systems may be summarized as follow (Busà, 1995; 1999; 2008b; Deterding, 2015; Wheelock, 2016):

- while Italian has only 7 vowel phonemes in its inventory, English presents a larger repertoire that counts around 11-13 vowel phonemes (depending on the language variety considered);
- traditionally, their *articulatory properties* are shown on a two-dimensional figure, sketchily representing the mouth shape, which consists of a *triangle* for

the Italian system and a *quadrilateral* for the English system. This kind of chart describes the articulatory organs action according to three variables: open/close (or low/high) and front/back, referring to tongue position, rounded/unrounded, referring to mouth and lips shape;

- comparing the two diagrams, attention may be drawn to the following aspects: while Italian has only one low unrounded vowel, i.e., / a /, English has three, i.e., / æ /, / ɑ: / and / ʌ /; while Italian is missing high lax⁵ vowels, English has two, i.e., / ɪ / and / ʊ /;
- the phonemic symbols inside the two diagrams indicate fully pronounced vowels, which represent the norm in Italian, since vowel quality tend to remain constant regardless of the degree of syllable stress, but not in English, where instead vowel quality is much more dependent on stress assignment rules and other suprasegmental events, causing *full vowels* to occur only in stressed syllables. In fact, due to vowel reduction processes, English additionally counts 5 *reduced vowels* occurring in unstressed syllables, the most common of which being the schwa / ə /.

Thus, it may be argued that the two systems numerical asymmetry alone would be enough to account for Italian speakers' difficulties in perceiving and producing English vowels. In such context, cases of *hypo-differentiation* of vowel contrasts are anything but rare (Busà, 2008b). With respect to the listed segments, for example, it is likely that the English vowels / æ / and / ʌ / will be pronounced as the Italian /a/ (see Wheelock, 2016 for more explicative examples). It should not be surprising, then, that differentiating words in English distinguished solely by a vowel (e.g., *bad/bed, sit/seat*) appears to be one major obstacle to Italian learners' intelligible pronunciation (Busà, 2008b). As for discerning between full and reduced vowels, the occurrence of phonemic *over articulation* is highly plausible; Italian speakers, whose L1 presents only limited vowel reduction at the phonetic level and no vowel reduction at the phonological level, might be unable to perceive vowel quality differences so as to pronounce all target segments in their full form (Busà, 2008b).

However, it is worth noting that problems related to the uttering of vowels may arise, not only from differences in terms of phonemic variety between the L1 and the L2

⁵ *Lax vowels* are produced with less muscular tension and are usually shorter than *tense vowels*, which instead are pronounced with greater muscular tension (Ladefoged & Maddieson, 1990, as cited in Busà, 1995).

systems, but also from the inherent *relative nature* of this specific set of segmentals (Busà, 1995). On this matter, given that each linguistic sound can be described on the basis of three spheres of existence, that is *articulatory*, *acoustic* and *perceptive*, Ladefoged (1967) highlights that, while consonants tend to be identified by listeners according to their articulatory properties, i.e., movements and positions of articulatory organs, vowels are instead recognized according to their acoustic properties, i.e., the way sound propagates through air (Ladefoged, 1967, as cited in Busà, 1995). Here follow some of the reasons behind this claim.

Consonant and vowel sounds seem to differ above all in the way they are produced articulatorily, in that the former are characterized by obstructions or semi-obstructions of various articulators, whereas the latter are obtained solely by tongue movements, with no blockage in the vocal tract (Busà, 1995; 2008b). Thus, perceptively speaking, consonants are likely to be identified more neatly than vowels (Repp, 1984, as cited in Busà, 1995), which in contrast are hardly distinguishable, especially in isolation (Fry et al., 1962, as cited in Busà, 1995).

Having said that, tongue approximate coordinates can only partially explain vowel quality. Indeed, more emphasis should be given to the role played by the whole vocal tract, including the laryngeal portion over the vocal folds. In fact, the larynx constant shaping — together with specific configurations of the upper articulators — functions as an acoustic filter, which, by emphasizing only certain frequencies of the signal, allows the quasi-periodic raw sound (originating from the vocal folds quasi-periodic vibration) to assume a peculiar vowel sound profile (see section 1.1 for acoustic details). More precisely, having reached some peaks of intensity, these emphasized frequencies, named *formants*, represent the acoustic correlates by which the listener is able to perceive different vowel qualities (Busà, 1995; Denes & Pinson, 2015).

Due to today's greater availability of user-friendly speech analyzers, the *acoustic analysis* of vowels (and of speech in general) has gone beyond the scope of experimental phonetics, becoming a common practice also in applied linguistics and occasionally in the language classroom (see section 1.4). The use of vowel diagrams based on acoustic data, that is on the intersection between the first two formants frequency values (F1 and F2), though apparently resembling the articulatory representation, may provide additional insight on the *vowel space*. By plotting the F1 and F2 values of different acoustic realizations of each phoneme, it is possible to obtain

a more objective description of the space occupied by each vowel (given a cohort of subjects with similar voice quality), which in turn can lead to further reflection on vowels indefinite character (Busà, 1995). For example, by juxtaposing the Italian and English acoustic diagrams what may be observed is that the English system covers a larger area of the diagram, hence showing an overall greater phonetic complexity (see figure at p.56 in Busà, 1995); in addition, even those vowel phonemes that, on the articulatory level, are deemed as equivalent in the two languages, acoustically speaking appear to share only in part the same phonetic quality as they occupy quite different areas of the diagram (Ferrero et al., 1979, as cited in Busà, 1995).

In light of this brief review of Italian and English segmentals, it may be hypothesized that the perceived degree of Italian accent in English is more likely to stem from a faulty vowel production rather than a faulty consonant production (see the study conducted by Busà, 1995; 1999 for further details on such correlation). Therefore, aiming at improving vowel pronunciation accuracy of Italian students, instructors should give priority to the perceptive/acoustic dimension of particularly problematic English vowel sounds, by proposing for instance a wide variety of aural discrimination exercises (Busà, 2008b). Equally importantly, they should provide learners with “individual feedback on their productions, as their self-correcting ability is limited by the influence of their native language” (Busà, 2008b, p.118). As it will be pointed out in the second part of this phonological comparison, a similar approach is also recommended in the teaching of suprasegmentals (being non-identifiable articulatorily), where, in particular, acoustic analysis will play a key role.

1.3.2 Italian and English Phonological Systems: Suprasegmental Features

The transfer of Italian prosodic patterns into English speech by native Italian speakers tends to frequently fall into the category of forgivable pronunciation mistakes, acquiring in some cases, as mentioned above, a positive connotation. Such lenient attitude seems to be shared also by many instructors, who generally give greater priority to the teaching of segmentals rather than suprasegmentals (Busà, 2008a; 2008b). Be it the result of time constraints or lack of teachers’ preparation, the tendency to neglect this topic, thus overlooking the impact of suprasegmental phenomena on (mis)pronunciation of single segments, could represent an obstacle to learners’ attainment of successful communicative outcomes.

As with challenging phonemic asymmetries, learners should be explicitly instructed about problematic differences between their L1 prosodic system and that of the target language. In this respect, Spaai and Hermes (1993) assert that teaching prosody implicitly, through simple listening and repetition, can hardly lead to its real learning (Spaai & Hermes, 1993, as cited in Busà, 2008a).

Following this principle, a comparison between some significant Italian and English suprasegmental aspects will be presented. Note that all of them are somehow interwoven with vowel perception and production processes.

Syllables

These language units, that usually receive little consideration in the description of language, may be regarded in all respects as suprasegmentals, as they consist mostly of combinations of sounds that, despite having blurry edges, are “psychologically real to language users” presumably due to the fact that they “correspond to peaks in sonority”(Brown, 2015). This would be demonstrated by the fact that individuals are capable of identifying quite easily the number of syllables in words, regardless of their native language and writing system (Brown, 2015). Moreover, word and sentence stress may be deemed as falling on syllables rather than individual phonemes (Brown, 2015). Hence the importance of analyzing language specific syllable structures.

Italian syllable structure is primarily of the CV type, having one vowel in peak or nuclear position (like all syllable structures), one consonant in onset position and an empty coda position. As such, it can also be defined as an *open syllable*. As for other types of structures, Italian allows only limited combinations of consonants in onset or coda position (Busà, 2008b; Brown, 2015).

English syllable structure, on the other hand, is primarily of the CVC type, hence a *closed syllable* due to the final consonant. In addition to this structure, English uses many other syllable forms, allowing complex consonant clusters both in onset position (up to three consonants) and coda position (up to four consonants) (Setter & Jenkins, 2005; Busà, 2008a; 2008b). As underlined by Brown (2015), English presents a more complex syllabic system than that of most languages, often causing problems to foreign speakers, who would naturally try to simplify English pronunciation.

Stress and segmental duration

Given that consonant length is neither a distinctive feature of Italian nor of English and that word and sentence stress most commonly fall on the vocalic nucleus of syllables, altering among others its duration, this part of the discussion will address the correlation between stress and vowel duration.

Vowel length in Italian is phonetic, meaning that it does not distinguish different vowel phonemes (Busà, 1995). Moreover, vowel duration is generally rather constant, mainly due to rhythmic reasons (see below), despite vowel position in words or sentences (Farnetani & Kori, 1990, as cited in Busà, 1999). Indeed, duration becomes relevant in *word stress* contexts, where the stressed syllable presents *de facto* increased duration. Here, the type of syllabic structure will determine the relative length of the vocalic nucleus, which will be longer in open syllables and shorter in closed syllables (Fava & Magno Caldognetto, 1974, as cited in Busà, 1999). Although stress might be perceived mainly as a rise in sound intensity, the data provided by acoustic analysis indicate duration as the main acoustic correlate of word stress in Italian (Farnetani & Kori, 1986; 1990, as cited in Busà, 1995).

Vowel length in English, conversely, is phonemic, since it contributes to create vowel oppositions, as in *pull/pool* (Busà, 1995). In addition, English vowels seem to show an overall greater variability in duration than Italian vowels; this would be due not only to rhythmic reasons, but also to the context in which these segments occur, such as a particular consonant environment (e.g., longer vowels before voiced consonants vs shorter vowels before voiceless consonants) or the position in words and phrases (e.g., vowel lengthening before pauses) (Klatt, 1976, as cited in Busà, 1999). The distinction between stressed and unstressed syllables, in the case of English, is determined by the co-occurrence of several other elements besides duration, among which the presence or absence of vowel reduction phenomena (see below) and a combination of higher level prosodic phenomena on stressed syllables, such as rhythmic and intonational stress (Dauer, 1983 and Bertinetto, 1988, as cited in Busà, 1999; Cutler, 2015). Thus, duration cannot be considered the main acoustic correlate of stress in English (Ladefoged, 1975, as cited in Busà, 1995), as, due to the factors just mentioned, along with being longer, stressed syllables might be louder (acoustic intensity) and higher in pitch (acoustic frequency) (see Figure 6.1 in Cutler, 2015 for an acoustic comparison by means of waveforms and spectrograms of the same syllable in stressed and unstressed condition).

Speech rhythm and isochrony

Rhythmically speaking, languages have long been considered to present what is referred to as *isochrony*, namely a form of equality in timing based on the recurrence at regular intervals of either syllables or stresses (Low, 2015). While Italian has been defined as a *syllable-timed language*, with syllables being the rhythmic units recurring at regular intervals, English has been described as a *stress-timed language*, meaning that stresses are the units recurring at regular intervals (Pike, 1945 and Abercrombie, 1967, as cited in Busà, 1995). Having said that, such rhythmic categorization does not seem to be confirmed by acoustic measurements, remaining mostly a *perceptive contrast* (Busà, 1995; Low, 2015). That is why “isochrony should be described as a tendency rather than as an absolute” (Low, 2015, p.125) or else it could be seen as the result of the merging of already mentioned aspects, such as syllable structure, stress patterning and presence or absence of vowel reduction (Dauer, 1983; 1987, as cited in Low, 2015).

In this perspective, it may be noted that Italian speech is mainly characterized by the presence of a fixed type of lexical stresses interspaced by long sequences of unstressed syllables with similar duration, a typical feature of syllable-based languages (Busà, 1995). In contrast, English speech shows a less-fixed type of lexical stresses together with different stress levels, which leads to a sort of continuous alternation of stressed and unstressed syllables (Cutler, 2015). The effect of this variation on segments may be regarded as a key element in the perception of English stress-timed rhythm. In fact, whenever a syllable receives the *primary stress*, meaning segmentally a full vowel in peak position, one or more surrounding syllables undergo *vowel reduction* (Ladefoged, 1975 and Mackay, 1987, as cited in Lepage & Busà, 2014), in other words the centralization of the vocalic nucleus, which may take the form of the indistinct schwa sound / ə / or even disappear entirely in some instances. This phenomenon, by compressing the unstressed syllables, would serve to maintain constant time intervals between primary stresses (Busà, 1995). Experimental evidence on the impact that incorrect word stress placement and faulty vowel reduction could have on EL2 intelligibility may be found in Lepage and Busà (2014), whose study examines native English listeners perception of Italian and Canadian French accented words containing these two “natural occurring prosodic errors”. Results show lower percentages of correct identification for words categories containing both errors, thus confirming that the co-occurrence of incorrect stress and vowel reduction may impair L2 intelligibility.

Because rhythm, in the words of Taylor, is “perhaps the most widely encountered difficulty among foreign learners of English”, it would be worth considering these two prosodic aspects as good candidates for in-class pronunciation activities (Taylor, 1981, as cited in Setter & Jenkins, 2005, p.8).

Intonation

Also known as *speech melody* or *variation of tone* (or *pitch*) across phrases, intonation is probably the suprasegmental trait that most closely recalls the analogy between our vocal system and a musical instrument (Denes & Pinson, 2015). In support of this claim, it may be noted that, although proved to be unsuitable, some attempts to transcribe and teach intonation involved the use of notes on musical staves to represent different tone levels (see Steele, 1775 as cited in Levis & Wichmann, 2015 and Chan, 1985, as cited in Levis, 1999).

Despite its ample use during speech, finding an adequate method for depicting intonation in written material seems complicated (see Levis & Wichmann, 2015 for an account of different notation systems). Moreover, there is often no unanimous opinion on the way intonation is perceived. Here, acoustic analysis, by providing more objective information, may be of great help in complementing listening (Levis & Wichmann, 2015). In fact, today’s speech analysis software allow *real-time tracing* of *fundamental frequency (F0)*, namely the main acoustic correlate of pitch. More specifically, an F0 display is capable of tracking vocal folds vibration rate for each *sonorant segment* (while no or little trace is left by fricatives and plosives, since their production does not involve folds vibration) giving as a result an *intonation* or *pitch contour* of the speech signal (Levis & Wichmann, 2015). Considering, for instance, an English complex sentence (see for e.g., Figure 8.1 in Levis & Wichmann, 2015), the reading of its intonation contour may provide valuable insights on the way pitch variation communicates both *phrasing* and *discourse meaning*. While phrasing is generally indicated by breaks in the F0 trace (matching pauses), meaning may be conveyed by *pitch prominence* on certain syllables to mark focus (i.e., usually an upwards movement of the pitch line) or by *final pitch movement* to indicate openness or closure of phrases (e.g., a final rising or falling contour) (Levis, 1999; Levis & Wichmann, 2015).

Based on the above, it may be argued that acoustic analysis of intonation could be a source of quick and reliable *feedback* for EL2 learners whose native language involves a

more restricted use of intonation compared to English (see section 1.4). This being the case of native Italian speakers who, due to such suprasegmental contrast, would be likely to hypo-differentiate English intonational structure (Busà, 2008b; 2021). Here is why in more detail.

It should be pointed out that there is a number of morphological and syntactic features affecting the extent and the way in which intonation may be used in a language. In the context at hand, Italian presents a rich system of inflectional morphology and a flexible word order, while English has fewer inflections and a rather fixed word order (Busà, 2008b). As a consequence, Italian will tend to move syntactic elements in the sentence and use lexical items for signaling salient information; English, conversely, will either use an intonational or primary phrase stress (i.e., an extra rise in pitch), normally on the last major word of a tone unit, or move the same extra stress to other words of the unit in order to shift focus on new or emphatic information, thus de-accenting old or shared information (Busà, 2008b; Levis & Wichmann, 2015). Lastly, besides a possible transfer of the wrong focus marking device, a faulty use of intonation by EL2 Italian learners may concern final intonation patterns, most typically, *question intonation*. As an example, evidence show that the peculiar falling intonation of English open questions (i.e., starting with a question word) appears particularly challenging for Italian speakers, who instead would tend to use a rising pitch to end this question type, mirroring the more commonly used pattern in their native language (Levis, 1999; Busà, 2008b).

1.4 Technology and Pronunciation Instruction

The use of technology for training L2 pronunciation can be traced back to the period between the late 1800s and early 1900s with the first recorded models played by the *phonograph* and the *gramophone* (Hincks, 2015; O'Brien et al., 2018). These tools were displaced after World War II by *tape-recording* machines, which allowed learners to listen to models as well as to record their own speech; these machines were an essential component of *language laboratories*, which remained in vogue until the end of the 1970s (Roby, 2004, as cited in Hincks, 2015). With the advent of early digital technologies, *computer laboratories* for language learning came into being; however, in the 1990s, *personal computers* (with audio input and output capacity) gained in popularity, leading to the development of rudimentary pronunciation software, whose

functions were almost comparable to those of a recording device. Only starting from the 2000s, with the increasingly widespread use of the *Internet* and *digital tools*, the availability of L2 pronunciation models grew exponentially and speech technology evolved to the point of providing feedback, although limited, on students' perception of pronunciation. Considering that the provision of feedback on production is a more cumbersome task for digital systems (Hincks, 2015).

While the field of research termed CALL (i.e., Computer-Assisted Language Learning) has long been investigating how technological advances could assist language teaching and learning, more recently a peculiar branch of CALL, denominated CAPT (i.e., *Computer-Assisted Pronunciation Training*), has been developed with the purpose of exploring how technology could improve specifically pronunciation training (Fouz-González, 2015; Pokrivčáková, 2015). However, it should be highlighted that CAPT is an ever-expanding field of research characterized by the constant emergence of new sophisticated software and approaches. Thus, far from attempting to provide the most up-to-date picture of the state of the art, this section will focus on the adaptability of various web materials to pronunciation training as well as outline the main types of speech technology most commonly used as an aid to pronunciation learning in and outside the classroom (Fouz-González, 2015).

Since technology started being used more extensively in language education, it has become increasingly clear how its contribution could represent a possible solution to the many issues entailed in pronunciation instruction. Here is a three-point description as to how.

First of all, the need for *constant feedback* on the part of the learners seems to be fulfilled. If traditionally, students could rely solely on teachers' observations in order to overcome the perceptual and productive constraints caused by their L1 (Fouz-González, 2015), many CAPT systems are able to provide immediate and possibly customized pronunciation feedback (Busà, 2008a; Revell-Rogerson, 2021). This may give learners the chance to practice not only when the teacher is available but also autonomously, at their own pace and according to their priorities (Hincks, 2015; Busà, 2021). With particular reference to FL contexts, the enhancement of oral skills can thus take place within a private environment outside the classroom, benefitting, above all, those less confident learners who often experience anxiety when practicing in front of others (Fouz-González, 2015). These systems include, for instance, free online courses, as the

one offered on the BBC website⁶ or many applications for PCs, mobiles and tablets dedicated to pronunciation training, such as *Duolingo*, a free app, or *My English Tutor*⁷ (MyET), a paid app (Pokrivčáková, 2015; Busà, 2021; Revell-Rogerson, 2021).

Connected to this, it may be asserted that the endless source of linguistic input provided by the Internet have become central for the improvement of L2 communicative competence; movies, TV series, tutorials, online games, conference talks (e.g., *TED Talks*), all available on several platforms, can be conceived as examples of *authentic* — yet complex — *material*, which exposes L2 speakers to an extremely wide linguistic variability, both in terms of accents, linguistic registers and also communicative functions (Busà, 2021). These resources might be used in class to foster greater interaction among students or, again, they might be used autonomously whenever the learner wishes, being particularly suitable for what is called “ubiquitous learning”, that is the possibility of studying anytime and anywhere thanks to the aid of mobile digital devices (Busà, 2021; Li et al., 2005 and Yang, 2006, as cited in Revell-Rogerson, 2021).

Ultimately, the advantages offered by computer-assisted instruction may contribute to reduce many *teachers’ reluctance* towards pronunciation teaching (Pokrivčáková, 2015): from those who claim a lack of confidence due to insufficient training, given the online availability of huge amounts of well-structured and ready-to-use pronunciation exercises, as well as self-training materials; to those who, despite being willing to teach pronunciation, might be held back by time or curriculum constraints, in which case suggestions could be made by instructors, after careful evaluation, on which technology could best suit their students’ needs when used by themselves (O’Brien et al., 2018). This echoes Derwing’s “make good use of technology” principle, according to which, among the options available on the web, there could also be “poor quality offerings, some of which may actually do more harm than good”, hence the importance of teachers’ recommendations on “what the students should work on at home” (Derwing, 2018, p.16).

Delving into the matter, the technology used in CAPT systems, in large part originally developed for the treatment of communication disorders (Pokrivčáková, 2015; O’Brien

⁶ <https://www.bbc.co.uk/learningenglish/features/pronunciation>

⁷ <https://www.duolingo.com/learn>; <https://tw.myet.com/MyETWeb/PersonalizedPage.aspx>

et al., 2018), may be considered as divided into three different basic typologies: speech synthesis, speech recognition and speech analysis (Hincks, 2015).

Speech synthesis consists in the production of computer-generated speech (Busà, 2008a). Also known as *text-to-speech*, these systems are most commonly used to provide synthesized input by artificially transforming written texts into spoken language (O'Brien et al., 2018). The main advantage of this technology lies in the fact that individuals, no longer needing to rely on pre-recorded items, would be able to generate the pronunciation of any utterance at anytime, just by typing it on a computer or a similar device, hence obtaining an unlimited number of pronunciation models (Fouz-González, 2015; Hincks, 2015). This is the case of many online *translator software*, such as *Google Translate*, which, despite being among the most known, still features a robotic-like voice, or the one offered by *www.howjsay.com*⁸, which by integrating artificial intelligence (AI) can produce a much more natural-sounding voice (Fouz-González, 2015; Pokrivčáková, 2015).

If combined with visual models of the face and oral cavity, speech synthesis may supply learners with augmented input and/or feedback, i.e., auditory and visual simultaneously. These sort of virtual tutors, also called *Talking Heads*, can illustrate articulatory movements through either two- or three-dimensional animations of the mouth producing speech sounds (Busà, 2008a; Fouz-González, 2015; Revell-Rogerson, 2021).

Furthermore, speech synthesis may contribute to the enhancement of learners' perception of problematic sound contrasts. In fact, given its high level of controlled manipulation compared to natural speech, it allows researchers to create series of *synthetic stimuli* differing acoustically only subtly from one another, with the purpose of shifting learners' attention depending on the focus of the trial (see Wang & Munro, 2004, as cited in Hincks, 2015, for an example of effective use of synthetic stimuli in the perception of English vowel quality contrasts by Mandarin and Cantonese learners).

Until recently, one major concern about speech synthesis was represented by its potential *lack of naturalness* and, as a consequence, a hunch by some researchers (initially confirmed in studies such as Strange & Dittman, 1984, as cited in Fouz-González, 2015) that training pronunciation with synthetic stimuli could result in less positive gains than using natural stimuli. However, the latest progress in this field (e.g.

⁸ <https://howjsay.com/text-to-speech>

the ever greater integration of AI) has been such that several synthesis systems are now showing an extremely-realistic output capacity (Fouz-González, 2015; Revell-Rogerson, 2021).

On the other hand, *speech recognition* refers to the technology that enables computers to recognize the sounds produced by human speech and to respond to it in multiple ways (Fouz-González, 2015). Currently, the best-known commercial applications for automatic speech recognition (henceforth ASR) are: *dictation systems*, which can process and transcribe any speaker's voice, such as the speech-to-text feature of many messaging apps (e.g., *WhatsApp*); and *spoken dialogue systems*, which can follow different kinds of users' voice commands, from the simpler, such as setting alarms, to the more complex, at times resembling a sort of mini-conversation between the user and the device (Fouz-González, 2015; Hincks, 2015), as in the case of several virtual assistant technologies, like Apple's *Siri* or Amazon's *Alexa* (Busà, 2021).

This communicative potential of ASR may be of great appeal for CAPT, as ideally, if ASR software were able to process everything the user says, they would provide learners with objective and personalized feedback on any aspect of their pronunciation (without them needing specific knowledge of phonetics), this way mirroring the teacher's action, but without any time limit or personal bias (Fouz-González, 2015; Hincks, 2015).

Nevertheless, despite being at the basis of a large number of CAPT systems, ASR technology still seems to present several *limitations* and, while its use has proved to be quite effective in the assessment of perception and production of segmental features, the same cannot be said for suprasegmental features (Revell-Rogerson, 2021). Predictably, one of the main hurdle in ASR could be represented by its *limited storage capacity*, in other words, the maximum amount of native pronunciation models (varying depending on the software) against which the L2 signal is compared (Busà, 2008a). In fact, being the storing of all nuances of human speech still unconceivable, at least at present, ASR-based CAPT will only partially be able to detect pronunciation errors, given the myriad of non-native accents, while the type of feedback provided will still be more generalized (for e.g., global pronunciation scoring) rather than individualized and constructive (Fouz-González, 2015; Hincks, 2015). In addition, as Revell-Rogerson (2021) points out, users may even receive "erroneous feedback", since the system might "either not

detecting learner errors or falsely labelling acceptable pronunciation as errors, both of which can be frustrating and demotivating for learners” (p.194).

Similarly to speech synthesis, ASR technology has undergone significant changes in recent years. Due to the growing use of AI, which enables software to “adjust to learner speech with increased exposure, as a human does” and more sophisticated databases, improvements may be noted in many CAPT systems and applications, especially with regard to *feedback accuracy* (Revell-Rogerson, 2021, p.194). Generally speaking, speech recognition systems have now been tuned so as to more precisely identify every user’s pronunciation needs, thus providing more customized and timely feedback (Revell-Rogerson, 2021). One example of these advances can be found in the commercial application *ELSA Speak*⁹, an English pronunciation app particularly suitable for smartphones. ELSA combines speech technology and artificial intelligence, giving users the chance of a thorough and targeted practice (based on an initial evaluation test) in the perception and production of a wide variety of segmental features (e.g., vowel to vowel linking in connected speech) but also suprasegmental features (e.g., word and sentence stress). It also presents an AI English coach with whom the learner may interact in guided conversations on specific pronunciation issues; the dialogue continues each time the pronounced utterance receives positive feedback by the virtual coach.

Lastly, *speech analysis* (covered in more detail in the later section) concerns the evaluation of the acoustic properties of a speech signal through different visualization techniques, such as waveforms, spectrograms and pitch contours (Hincks, 2015). Having been originally designed for experimental phonetics, the application of *acoustic analysis software* in CAPT (with the main purpose of providing audio-visual feedback) has occasionally drawn criticism due to possible interpretation difficulties of the on-screen representations that some learners may encounter (see for e.g., Setter & Jenkins, 2005). However, a considerable amount of studies have revealed the effectiveness of visual displays in improving L2 pronunciation, both in terms of phonemic and prosodic accuracy (e.g., Weltens & De Bot, 1984; Lord, 2005; Hamlaoui & Bengrait, 2016).

Overall, CAPT resources would seem to offer endless advantages to L2 pronunciation teaching and learning. Their increasing availability and affordability have led to an ever greater integration of these tools and materials in the language classrooms as well as in self-learning activities. Nevertheless, caution is advisable, in that, despite often showing

⁹ <https://elsaspeak.com/en/>

an appealing and user-friendly interface, many CAPT apps and systems may lack solid pedagogic underpinnings, being their development determined by commercial rather than pedagogical reasons (Busà, 2008a; O'Brien et al., 2018; Revell-Rogerson, 2021). One main concern about CAPT tools is that they still tend to adopt the *one-size-fits all* approach, which might be counterproductive for students, who would be exposed to too generalized content and feedback, thus possibly wasting considerable time and losing motivation in the long run (Derwing, 2010; Revell-Rogerson, 2021). Ultimately, it could be argued that, although new advances in pronunciation technology may offer outstanding possibilities, a large part of it should primarily be considered for its *awareness raising* function (Setter & Jenkins, 2005) rather than its specific practical pedagogic value; hence, for the time being, the title of “best teachers, evaluators and correctors of learners’ performances in L2” would seem to remain in people’s hands (Busà, 2008a, p.173).

1.4.1 Visualization Technology and Acoustic Analysis of Speech

Arguing for the importance of individually tailored and objective feedback, the use of *visual displays* for pronunciation training has long been proven to provide a form of customized and unbiased evaluation of speech (Hincks, 2015). Typically, their contribution is defined as *multimodal*, since the feedback obtained from model inputs and learners’ own outputs is both auditory and visual (Revell-Rogerson, 2021). Through the use of speech analysis software, learners can in fact “record and visualize their speech output on their computer monitors”, as well as listen to and visualize native speakers models (Busà, 2008b, p.119). This dual feature may favour *perceptual* — and therefore productive — *enhancement*, in that students, as well as instructors, have more tangible information at their disposal while comparing and contrasting pre- and post-training outputs and/or L2 and native-speakers productions (Fouz-González, 2015; O'Brien et al., 2018).

That being said, rather than supplying actual explanatory and corrective information, the various patterns appearing on-screen carry *raw acoustic data* of the speech signal, which need to be correctly interpreted. This, at times, might be a challenging task for either learners or instructors, which is why the teaching of some basic notions of L1 and L2 phonology, as well as of acoustic phonetics before use is generally advocated (Revell-Rogerson, 2021; Busà, 2008a).

The first attempts to implement speech analysis systems in order to investigate the impact of *audio-visual feedback* on pronunciation training were directed to the study of prosody and in particular of *intonation*. It is worth noting that in the past the complex structure and costs of hardware and software systems for speech analysis required experiments to be conducted solely by language experts and within university computer laboratories (Busà, 2008a). Groundbreaking research of this sort was carried out in the Netherlands by De Bot (1983), who, by means of one of the earliest devices capable of plotting fundamental frequency (F0) data on a display in the form of a contour (see Figure 1), showed the efficacy of audio-visual feedback on Dutch students pronunciation of English intonation over auditory feedback only (see also De Bot & Mailfert, 1982; Weltens & De Bot, 1984). In addition, this experiment provided valuable insights into the potential benefits of audio-visual feedback on *learning behavior*, in that subjects exposed to the dual mode tended to repeat the target sentences more often than subjects receiving just auditory feedback, which in turn would suggest that, being *visually informed* about the distance or closeness of their own productions with respect to the model may “stimulate improvement through self-monitoring and self-correction” (Busà, 2008b, p.119).

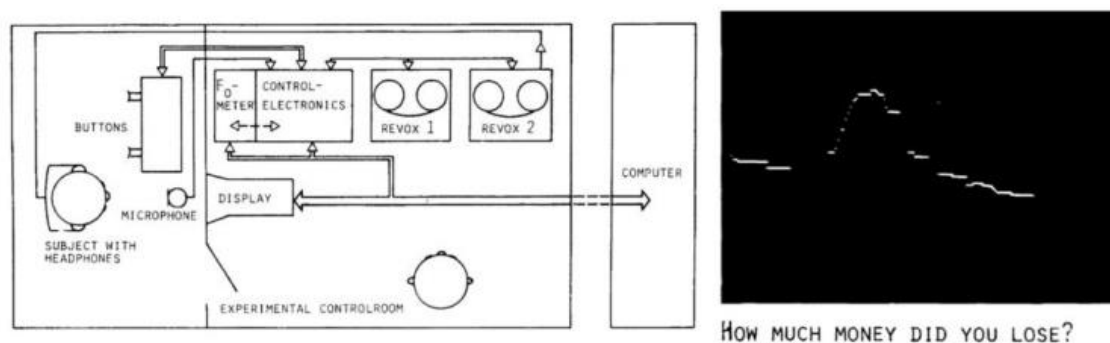


Fig.1. Experimental set-up and example of visualization used, i.e., pitch contour display.

(Adapted from De Bot, 1983)

Nowadays, technological development enables anyone to have access to a wide variety of speech analysis software, which are easily downloadable on PC (or other devices) at a relatively low cost, as in the case of *BetterAccent Tutor* (Busà, 2008a), or even for free, such as with *Praat* (see later section) and *WaveSurfer* (Hincks, 2015). These programs, despite showing some complexities in their advanced features, may be relatively straightforward when used within a *limited scope of application*. Moreover, the presence of a great deal of online material dedicated to them, such as tutorial videos

and shared scripts, can further facilitate the understanding and use of waveform, spectrogram and pitch displays (O'Brien et al., 2018).

Referring to this, *waveforms* may be regarded as the most versatile acoustic representation technique, capable of conveying important information on both the segmental and suprasegmental level of speech, as for instance, duration of problematic segments or intensity and durational differences between stressed and unstressed syllables (Fouz-González, 2015; Hincks, 2015). Consequently, waveform displays are often coupled with either spectrographic visualizations, in order to obtain feedback on vowels and consonants, or with pitch contour representations for intonation analysis.

On the other hand, the use of *spectrograms* (see Figure 2), depicting the amount of high and low frequencies carried by the signal (Boersma & Heuven, 2001), is usually considered with caution, in particular within the scope of classroom implementation. In fact, their correct interpretation is likely to require greater knowledge of articulatory phonetics and acoustic analysis (Busà, 2008a), as each sound needs to be associated with various frequency patterns (see for e.g., Saito, 2007). Nevertheless, the feasibility of introducing spectrographic representations in the language classroom (as well as a self-learning tool) is demonstrated by Olson (2014a), who, within a guided inductive framework (that does not entail any explicit phonetic or acoustic reference), presented English-speaking students with the chance of practicing problematic Spanish consonant sounds by means of *Praat*, which on the whole proved to be well received and properly used.

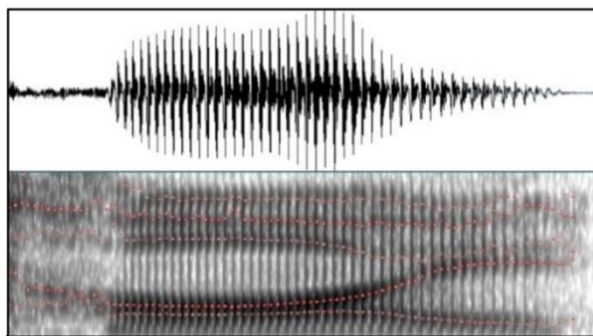


Fig.2. Example of double display of the English diphthong /ɔɪ/. In the upper box: the *sound wave*; in the lower box: the corresponding *spectrogram*. To be noted that the first two darker bands of the spectrogram, i.e., the first two formant frequencies (F1 and F2) show a change in the distribution of sound energy as a result of different articulation of the two vocalic elements forming the diphthong.

(Adapted from Behr, 2022)

Lastly, *pitch contours* (see Figure 3) are probably the most intuitive type of visualization, as they appear in the form of simple lines depicting our voice melody. The comparison between these fundamental frequency (F0) traces would then seem a fairly easy task even for students who have little or no training in speech analysis (Busà, 2008a). As a result, a large number of studies (within both laboratory and classroom settings) have explored the usefulness of pitch displays in the teaching and acquisition of L2 prosody (see for e.g., Ramirez-Verdugo, 2006, as cited in Fouz-González, 2015; Le & Brook, 2011; Chun et al., 2015, as cited in O'Brien et al., 2018;). However, it should be noted that their interpretation may as well not be so straightforward. Here is why (Hincks, 2015; Levis & Wichmann, 2015): 1) the continuous melody perceived while speaking is plotted on the screen fragmentarily, since unvoiced segments carry little or no pitch; this implies that the line could show *unexpected gaps*; 2) the trace can also present so-called *octave leaps*, in other words sudden spikes in pitch or brief sequences of much higher or lower contour that cannot be heard by the listener as they are caused by software miscalculation of the F0; 3) different natural *voice frequency ranges*, for example between the learner and the native speaker model, may result in significant mismatch when compared, despite correct imitation. Taken together, these inconveniences could be confusing and demotivating for the learner, who should therefore be informed in advance.

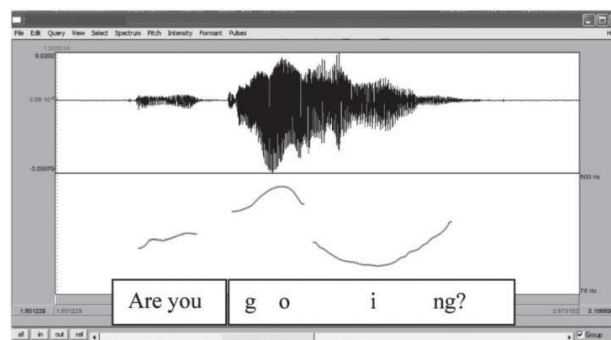


Fig.3. Example of double display of the question *Are you going?* uttered by an English native speaker. In this case the lower box shows the time-aligned *pitch contour* marking sentence stress (or peak of prominence) on the vowel /o/ and a final fall-rising pattern. (Busà, 2008a)

1.4.2 Praat: a Free Voice Analyzer to Train Pronunciation

Among the freely available speech analysis software, *Praat* (Boersma & Weenink, 2023) is perhaps one of the best known. Not only is it extensively used among

musicians and phoneticians, but it has gained increasing credibility as an effective tool in L2 pronunciation instruction (Boersma & Heuven, 2001).

Praat has been developed since 1992 by Paul Boersma and David Weenink at the Institute of Phonetics Sciences of the University of Amsterdam, however, due to its rather user-unfriendly interface, “designed more with the scientist/researcher in mind” (Wilson, 2008, p.112), its use as an aid to pronunciation teaching has occasionally raised controversy. For example, Derwing (2010), defines this form of application as “unrealistic”, suggesting that “only the most conscientious teacher is going to tread in that territory” (p.30); in the same vein, Setter and Jenkins (2005) highlight how challenging the reading of formant plots produced by Praat might be, arguing that it requires “a sophisticated level of understanding, which may be lacking in many teachers and learners” (p.10). Nevertheless, this computer program offers many advantages, first and foremost, to be freely downloadable from <https://www.fon.hum.uva.nl/praat/>, where a new updated version is often available.

Putting the emphasis on features that will be relevant in the experimental part of this dissertation, this software may be described as consisting of three main parts, or windows (Boersma & Heuven, 2001):

- the *Praat Objects window*, in which the recorded sounds are temporarily saved (see Figure 4);
- the *Sound window*, where the signal can be thoroughly analyzed through various acoustic visualizations (i.e., waveforms, spectrograms, pitch, intensity or formant contours), and functions (e.g., selection and zooming, listening of a time stretch), (see Figure 5);
- the separate *Picture window* into which the different visualizations may be drawn and from which any picture can be printed (see Figure 6).

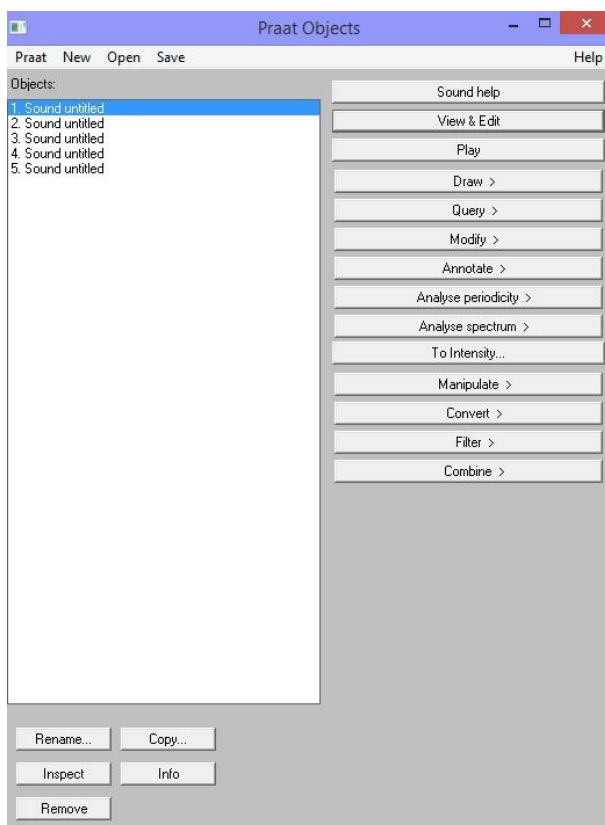


Fig.4. The Praat Objects window.

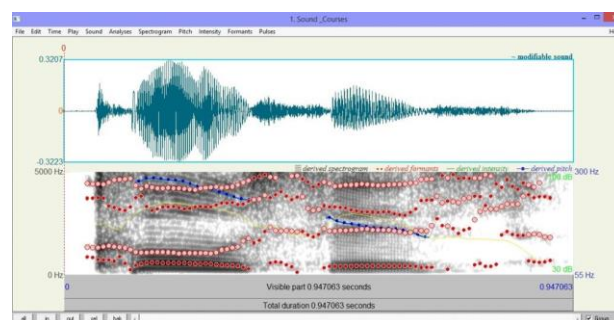


Fig.5. The Sound window.

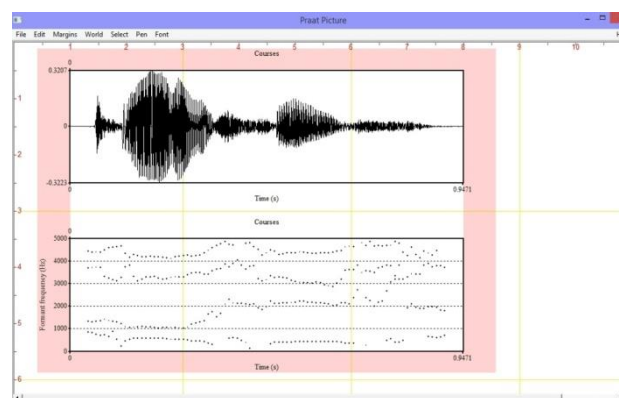


Fig.6. The Picture window

Overall, Praat, given its high level of functionality, may be regarded as a versatile software that likewise readily lends itself to a pedagogical use in the scope of pronunciation training (see Wilson, 2008 for a brief outline of Praat teaching functions). As further evidence of this, a large number of studies investigating the potential benefits of audio-visual feedback provided by Praat has been conducted *in the classroom setting*. Among the most recent, some have focused on individual phonemes or diphthongs (e.g., Olson, 2014a; Behr, 2022), while the majority have concerned *English intonation patterns* of various sentence types, such as yes/no questions, wh- and tag-questions or salutations (see for e.g., Busà, 2008b; Le & Brook, 2011; Hamlaoui & Bengrait, 2016; Larassati et al., 2022). Another aspect of the just cited studies worth mentioning is that the subject participants were advanced learners, most commonly higher-level university students, which could explain why the identification of errors by learners through visual feedback was in most cases successful and transferred to speech production.

However, it has to be said that general final improvement of oral skills seems as well to have been found in other scenarios, such as in the *lower-level, general skills language*

classroom (see Gorjian & Arvand, 2022 for the implementation of Praat at high school level) or even *outside the classroom*, in which case Praat was proposed as a self-training tool to a group of volunteer students in order to compensate the lack of opportunity to practice English pronunciation in class (Osatananda & Thinchana, 2021).

To conclude, from the many post-training questionnaires given to learners to assess the software accessibility, what emerges is that Praat is generally well received by students of lower and higher levels of instruction, who, despite reporting some technical difficulties, express the desire to continue practicing pronunciation by means of the software (see for e.g., Olson, 2014a; Osatananda & Thinchana, 2021; Larassati et al., 2022). Driven by such promising outcomes, the experimental part of this thesis will try to provide further evidence of Praat usability in the classroom context.

In summarizing this first chapter, the complexity of L2 pronunciation phenomena could be argued to be the reason why pronunciation was for a long time neglected and disconnected from L2 instruction (Underhill, 2010, as cited in Levis, 2019) and at the same time readmitted into the language teaching family (Levis, 2019), assuming eventually a prominent role. The advent of Communicative Language Teaching has in fact gradually emphasized the importance of pronunciation for successful oral interactions, encouraging researchers to investigate into its numerous issues within a communicative perspective. Altogether, these findings seem to confirm that native-likeness cannot be a realistic goal for most adult L2 speakers and that their foreign accent may be regarded as naturally part of linguistic variability; teaching pronunciation should rather aim at the intelligibility and comprehensibility of the message. In light of this, a more holistic and targeted approach has been suggested, which considers segmentals and suprasegmentals not as separated entities, but as equally relevant elements of L2 pronunciation that should be addressed together. Moreover, instructors should help learners recognize their errors by stressing those differences between the L1 and the target language which are more likely to impair intelligibility. Finally, the need of ongoing and personalized feedback has been further fulfilled by technology, which has proved to be a valid aid to the teaching and learning of pronunciation. Among the now countless applications available, speech analysis software seem to provide, through visualization displays, the most objective, language-independent form of feedback (Fouz-González, 2015), while giving learners a concrete chance of self-monitoring and self-assessment (Olson, 2014a). Nevertheless, in order to do this, some references to

sounds articulation and voice acoustics is recommended. Rather than being a drawback, this can make students more aware of the mechanisms beneath their productions, which would result in the fact that most students appear to “enjoy the hands-on experience of working with their own language and discovering facts about it” (Busà, 2008b, p.123).

RESEARCH PROJECT

This chapter is divided into four main sections. The first section explains the reasons behind the study and presents the research questions guiding the author throughout the investigation. The second section outlines the setting where the project took place and the various steps leading to the selection of the participants. The third section discusses the paradigms and approaches at the basis of the research design. Lastly, the fourth section provides a detailed description of the instruments and procedures adopted for collecting the data.

2.1 Research Questions and Hypotheses

Among the studies mentioned in the previous chapter on the use of Praat for L2 pronunciation enhancement purposes, only a small number was conducted in the lower-level language classroom (of tertiary education); quite understandably, the lower the level of instruction, the lower the implementation of the software. As a result, the involvement of upper secondary students in this kind of investigation has been found solely in one single study, i.e., Gorjian and Arvand (2022), which was carried out in a girls' high school in Dehloran, Iran. Moreover, as far as the researcher knows, the same scenario would apply to the Italian context (that of the current study), where Praat and more generally speech analysis technology, while being implemented at the university level (Busà, 2008b), still seem to be absent from upper secondary education.

Addressing this lack of literature, this small-scale study was undertaken to explore whether students attending an upper secondary school in the city of Venice, Italy, could benefit from audio-visual feedback in developing their English pronunciation skills. Based on the results of previous work (e.g., Weltens & De Bot, 1984; Le & Brook, 2011; Olson, 2014a; Hamlaoui & Bengrait, 2016; Gorjian & Arvand, 2022), it was expected that the use of Praat software for this purpose would prove beneficial to EFL learners within such educational context.

In order to investigate this general assumption, the following research questions and hypotheses were formulated:

- 1) Is the space dedicated to pronunciation instruction during class enough to meet the English language requirements outlined in the school syllabus? Additionally, is pronunciation instruction welcomed favourably by students?

It is hypothesized that pronunciation teaching time is regarded as inadequate by instructors in relation to the syllabus requirements; at the same time, students are expected to show a certain reluctance towards the subject matter, possibly due to a sense of inadequacy.

- 2) To which extent, if any, can the use of visual-feedback provided by Praat enhance students' awareness on English suprasegmentals as well as improve their prosodic abilities while practicing different types of English questions?

It is hypothesized that proposing the use of Praat within a design tailored to the needs of the classrooms involved, can prove to be effective, serving both as an awareness raising tool and as a valid learning aid for improving the prosody of English questions.

- 3) What is the students' response to the proposed pronunciation activity? Is their perception of Praat conducive to a possible integration of the software in the scope of pronunciation instruction?

It is hypothesized that students respond positively to the proposed exercises and that the interest aroused by the new technological tool could motivate them to adopt it for further pronunciation activities.

2.2 Research Setting and Participants

The study was conducted with the participation of students attending an art and design *high school* located in the heart of Venice, Italy. This upper secondary institution offers a single academic program during the first two years of schooling, diversifying, from the third to the fifth year, into six different curricula.

As regards foreign languages, the only available option throughout the school path is the learning of *English*, which is taught *three hours* per week. In addition, students may have the chance to attend extracurricular courses, held by native teachers, which are aimed at providing preparation for two Cambridge exams: the *B1 Preliminary* (former PET) and the *B2 First* (former FCE).

Premise to the selection of participants

The choice of the research setting was suggested to the author of this thesis by her co-supervisor, who is familiar with the availability of the teachers of this school to

cooperate in various extracurricular activities. Prior to the beginning of the study, the researcher met with the English teaching coordinator to explain the nature of the didactic activity proposal.

Initially, fifth year students had been considered to be the best candidates for this kind of experiment, due to its potential complexity. However, the coordinator heartily recommended opting for second year students, being generally more open and receptive to novelty compared to last year students, who, instead, are more likely to be concerned with the upcoming graduation exam. This suggestion, though, came with a caveat, that of not expecting a great response in relation to the part of the activity designed to be carried out at home.

That being said, the study participants can be comprehensively outlined as follows:

- *2 English teachers* (one of whom being the above mentioned coordinator) who served as subjects of a preliminary interview aimed at providing a picture, even though approximate, of the state of pronunciation instruction in relation to this particular educational framework;
- *2 second year classes*, i.e., 2A class (20 pupils) and 2B class (21 pupils), whose teachers of English were the two interviewees. All students, in fact, were the subjects of classroom observation and attended the in-class activity presentation, the in-class guided analysis, as well as the final follow-up lesson.
- *12 students*, aged between 15 and 16 (of whom 6 from 2A and 6 from 2B), namely the actual sample of this empirical investigation, being the only subjects who completed both the initial recordings and the final recordings and, consequently, the ones who were asked to fill in the closing questionnaire.

2.3 Research Design

Given the dearth of information about possible applications of speech analysis technology with the population in question, the present study is designed as a *pilot study* aiming primarily at exploring Praat pedagogical adaptability to this specific research context.

In order to achieve this goal, the approach adopted as a model for this work was originally that of Olson (2014a), who, by means of Praat, was able to test (obtaining

promising results) the so-called visual feedback paradigm, or VFP, as a pedagogical method for the teaching of segmentals in a lower-level, general skills language classroom, i.e., “the equivalent of the first two years of L2 instruction” at the university level (Olson, 2014a, p.49).

The VFP could be concisely described as consisting of:

- (a) a non-native speaker recording the stimuli; (b) a visual display of the speech feature, most commonly the intonation contour; (c) a visual display of a native speaker production for comparison, often accompanied by a corresponding auditory presentation; and (d) a re-recording on the part of the non-native participant attempting to match the native-speaker productions. (Olson, 2014b, p.175)

However, Olson (2014a) developed his approach relying particularly on the software self-learning potential, thus creating a student-centered activity, “in the face of such teacher-centered pronunciation instruction”, which was based on “a guided inductive method ... with a particular nod to the notion of self-monitoring and self-assessment” (p.54). As a consequence, the VFP was implemented as follows:

- a *pre-activity* completed for homework prior to class, in which students, following specific written instructions, were supposed to download Praat, record a list of target words containing the problematic sound, print the visual displays (to note that the word *spectrogram* was avoided, being considered too technical, and substituted with “visual representation”) and try to detect the segmental patterns;
- an *in-class guided analysis* in which learners, divided into small groups, were asked to make a visual comparison, aided exclusively by a series of guiding questions, between the characteristics of their own productions and those of the same words produced by a native speaker, in order to hypothesize about the corresponding auditory differences (hypotheses that were then tested by viewing and listening to repetitions of the two pronunciations);
- a *follow-up activity*, to be completed as an at-home assignment, where students, again following written guidelines, had to re-record the list of items using Praat, this time having at their disposal both the auditory and visual feedback of native speaker productions; the main aim being modelling their recordings (visually

and auditorily) after those of native speakers. To conclude the task, copies of their new and old spectrograms were to be turned in.

Such an approach, although being bold, in that it contradicts previous common objections regarding the software complex design and interface, soon proved to be, to some extent, unsuitable for the uncharted territory where the present study would be carried out, the main reasons being the *time constraints* and the *classroom environment dimension*. In fact, as Dörnyei (2007) highlights, the classroom represents a highly complex environment, which may pose numerous problems to the researcher. In his list of difficulties and challenges that might be hiding in classroom research, three items are worth noting, as they played a crucial role in the fine-tuning of this research design, namely:

- *working with students*, in other words, being the involvement into the proposed activity on a voluntary basis, “it is highly unlikely that every student will do his/her best for a project in which they have little interest and which has no direct bearing on their school grades”; (p.189)
- *unexpected events and interruptions*, since, despite the planned schedule, covering about twenty days, the project was prolonged to almost a month due to various inconveniences, for example a school strike;
- *technical difficulties*, encountered especially during the recording of some students’ audio carried out (on the author’s PC) in particularly noisy environments such as the school corridor.

Thus, in line with its emergent nature (Dörnyei, 2007) a preliminary, *qualitative phase* of investigation was undertaken, in which classroom observation and interviews with teachers allowed the gathering of useful information (soft data) on the context and the subjects. These data were, in turn, used to further adapt Olson’s model to the needs of the two classes (see later sections), with the aim of encouraging greater participation and collecting more precise evidence (hard data) in the subsequent *quantitative phases* of the study, i.e., the activity with Praat and the post-training questionnaire.

Essentially, this series of steps may be argued to meet the characteristics of a particular research design called *exploratory sequential mixed method*, which refers to a study that begins “with an exploratory, qualitative phase” and moves “sequentially to a quantitative phase” (Creswell & Plano Clark, 2018, as cited in Mihás & Odum Institute,

2019, p.2). Following the symbol system proposed by some scholars (for e.g., Johnson & Christensen, 2004, as cited in Dörnyei, 2007) to represent the wide range of mixed methods designs, this research could be described as qual → QUAN → quan, namely a three-phase research where the central quantitative component (i.e., pronunciation activity with Praat) prevails over the other two (i.e., observation and interviews first and questionnaire last). As a final point, this sequence of method constituents is also known as the *instrument development variant* of an exploratory sequential mixed method design, as the initial (minor), qualitative component serves as a source of important data for the development of one or more refined instruments to conduct the subsequent quantitative part of the study (Creswell & Plano Clark, 2018, as cited in Mihás & Odum Institute, 2019, p.5).

2.4 Data Collection Procedures

Having mainly to handle audio data from minor students, the researcher's primary concern was to request parental consent, which was readily obtained by means of an informative notice sent to the families by the coordinator, to which no refusal followed. As for the interviews, the two teachers accepted without hesitation.

The instruments used to collect the data throughout the investigation were created by the author on the basis of research tools used in previous works, more specifically:

- some of the main themes covered by the interview questions were drawn from a preliminary survey on pronunciation teaching administered by Olson (2014a) to thirty language instructors;
- as mentioned in the previous section, the didactic experience conducted by Olson (2014a) served as a template for the creation of the Praat activity major steps, while the selection of the stimuli draws inspiration from the question types proposed in other studies on the use of Praat for the teaching of L2 prosody, in particular, Busà (2008b), Hamlaoui and Bengrait (2016), and Larassati et al. (2022); as for the rest of the related material found in the appendices, it was created, along the way, by and large from scratch, according to the evolution of the activity;
- among the several examples of questionnaires on the usability of Praat software, again the one proposed by Olson (2014a) on conclusion of his project, appeared

to be the most adaptable to the current study, offering useful hints for the set-up of a coherent item list.

2.4.1 EFL Teaching Context Outlining via Observation and Interviews

Observation

Preliminary classroom observation has had a marginal role in this investigation, having been conducted in both classes only during one single English lesson. Accordingly, the use of observation sheets was considered unnecessary, opting instead for field notes, which constitute the basis of the next description. The first part of the description will focus on the observation of 2A class, whose teacher will henceforth be referred to as Teacher A, while the second part will focus on the observation of 2B class, whose teacher will henceforth be indicated as Teacher B.

2A class

This class consists of 20 pupils, all native speakers of Italian. The classroom is equipped with an interactive whiteboard (IWB) and the students' desks are lined up in three long parallel rows facing the teacher's desk.

The 60 minute lesson, taking place in the morning, features the study of a section of the textbook (i.e., *Identity B1*, Oxford) and a final role play on the part of two students.

The learners are, on the whole, very quiet and polite, both towards the instructor and towards each other. Teacher A remains seated for the first part of the lesson, going through the textbook unit step by step. In order to do this, she resorts to the online version of the book (visible by everyone on the IWB), having thus access to its interactive features and multimedia content, especially the numerous audio files. In the same vein, at the end of the unit an animated content is showed to summarize the grammar rules addressed; no other method is used during the lesson for similar purposes.

Each student reads in turn part of the textbook exercises, showing overall good pronunciation skills. In case of misreading of English sounds, Teacher A does not provide immediate corrective feedback, but rather she occasionally returns to some problematic sounds repeating them correctly.

Generally, the instructor tries to engage all students calling them by their first name; while a few of them seem to respond significantly more actively, it is interesting to notice that one particular pupil, sitting at the back of the classroom, is repeatedly called by his second name, seemingly to gain his attention and involvement.

The final role-play is apparently a very common and appreciated activity within the class and it usually does not imply teacher evaluation. It consists of choosing a video from the Internet, projecting it on mute on the IWB and reciting the characters' parts verbatim. In this case, two girls recite a long scene from their favourite animated show while projecting the relative YouTube video. Worth noting, as evidence of their sense of initiative, is that these two learners are not among those who, during the lesson, have stood out for their confidence in terms of English pronunciation.

2B class

This class consists of 21 pupils, all native speakers of Italian, except for one native Russian-speaking student and another Italian learner who, visiting regularly her relatives in Australia, presents an almost native-like English pronunciation. The classroom is equipped with a video projector, whose image quality is inevitably lower than that one offered by an IWB. The students' desks are placed differently, that is into four rows (made up of two or three desks) at each side of the classroom, thus leaving an empty space in front of the teacher's desk.

The 60 minute lesson, taking place in the afternoon, features several different phases, such as an initial summary written test, a discussion on the Commonwealth drawn from the textbook (i.e., *Identity B1*, Oxford) followed by a reading and listening activity on the subject (showed on the projector screen), and a final reading test for those learners who need to improve their grades.

The students, although being polite, show to be rather loud. However, at any instructor's signal they suddenly become silent. Teacher B tends to stand and move around the room rather than sitting down, this way establishing a quite effective contact with many of the learners.

Considerable time is dedicated to pronunciation, in that each phase of the lesson is accompanied by additional references to problematic English sounds. Teacher B often resorts to the use of a normal whiteboard on which he writes the sounds in minimal pairs, such as *bat/but/bet*, followed by *batman/butman/betman*. During the final reading

test, given the aim of the task, Teacher B provides each student, sitting in front of him, with personal feedback on their pronunciation. 2B class may also be said to show overall good phonological competence.

Interviews

Subsequent to classroom observation, Teacher A and Teacher B agreed to be interviewed.

Having a clear enough overview of the issues regarding pronunciation instruction that needed to be explored (in relation to the objective of the research), the *semi-structured* interview type was chosen (Dörnyei, 2007). As such, the two one-to-one interviews consisted of the same set of 20 pre-prepared guiding questions (see Appendix A for the full question list), which were asked to both interviewees in the same order. Predictably, the respondents elaborated on certain themes, partly answering to some of the questions beforehand, thus, the recording of both interviews was crucial for correct data collection (Dörnyei, 2007).

The questions may be grouped according to the *themes covered* as follows:

- Q.1-2: personal questions on the respondent's job;
- Q. 3-4-5: organization of English teaching throughout the school path and role held by its components in the curriculum;
- Q. 6-7: the classroom and its attitude towards the learning of English;
- Q. 8-9-10: students' present attitude towards English pronunciation;
- Q. 11-12: curriculum guidance on English pronunciation and implementation during lesson;
- Q. 13-14-15: English pronunciation teaching techniques, attention to prosody and feedback modes;
- Q. 16-17-18: instruments and technologies at teachers' disposal;
- Q.19: self-assessment of the respondent's pronunciation teaching skills;
- Q.20: final closing question that gives the respondent the chance to add anything that might have been overlooked.

2.4.2 Pronunciation Activity with *Praat*

In line with the framework outlined by the first stages of this investigation (see section 3.1 for details), several important considerations were made in the development of the Praat activity, among these: being participation on a voluntary basis, effort was made to find suitable ways to draw learners' attention to the topic of pronunciation and encourage them to take part in the project; attention was paid in the creation of additional guiding materials (e.g., video tutorials), in order to give students a certain degree of autonomy in carrying out the task; and care was taken to avoid unnecessarily technical terminology (both in relation to phonetics and acoustic analysis of speech), as well as to limit the discussion on Praat features to the few commands strictly necessary for the purpose of the activity.

Moreover, attempting to provide valuable and constant support to complete the task successfully, one of the first steps was to create two virtual classrooms with *Google Classroom* (one for 2A class and one for 2B class), which would allow both the researcher to upload all the necessary material together with explanations, and the subjects to upload the audio files of their recordings. In addition, this common virtual space would serve as a place in which any question or doubt could be raised. Thus, one week before the starting date, the *activity schedule* was uploaded (see Appendix B), in order to inform the learners on its content and timeframe, and allow them to start considering whether or not to participate.

Overall, this central part of the research may be outlined as consisting of *four main phases* framed by an initial presentation and a final questionnaire (see subsection 2.4.3 for further details on the questionnaire). Presented below is a brief list of such phases, which shows, among other things, the alternation of their place of completion. To be noted that the same procedure was applied to both classes.

- *Presentation* (in class): PowerPoint introduction to the subject;
- *Phase 1* (at home): initial recordings of the target questions with Praat;
- *Phase 2* (in class): guided analysis with Praat on rhythm and intonation of students' recordings and comparison with native English speakers' productions;
- *Phase 3* (at home): practice and re-recording attempting to model their productions after native speaker models;
- *Phase 4* (in class): assessment of skills acquired during the activity through the reading of dialogues;

- *Questionnaire* (at home): online questionnaire shared on Google Forms.

Presentation

The three major topics addressed by the PowerPoint presentation were the following:

- 1) Why do we often consider pronunciation to be a difficult aspect of foreign language learning?
- 2) What do we actually mean by *pronunciation*? What are its main components?
- 3) What is Praat software? How can it be useful for pronunciation learning?

In order to introduce the class to the first theme, a video showing two *Xhosa language speakers* was presented (<https://www.youtube.com/watch?v=DPaSfoNT4yQ>), followed by a description of *the human vocal apparatus*. The choice of such unfamiliar language was made both to capture the audience attention, being notoriously catchy for its *clicks*, and to highlight the adaptability of any individual's vocal system in the production of an extremely wide range of speech sounds. In this context, reference was made to a possible analogy between our vocal apparatus and a musical instrument, in that the first can be trained, (not only for singing but also for speaking) as much as the second can be tuned. Emphasis was thus placed on the concept of *linguistic interference* between the L1 and the L2 and on how this phenomenon may affect language learning, especially at the beginning (see Appendix C).

As for the second topic, a diagram of *pronunciation main components* was shown followed by a brief outline of the differences between vowels and consonants and a more in-depth explanation of rhythm and intonation. Here, *Italian rhythm* was associated to the regular periodic sound of a machine gun, in order to distinguish it from *English rhythm*, which, instead, was compared to a roller coaster for its stress patterning. A practical exercise was then proposed (drawn from https://www.youtube.com/watch?v=_P7_69FeqnU&t=12s), where students had to snap their fingers in correspondence with English stress placement. Finally, the part dedicated to *intonation*, despite being a complex subject per se, was able to catch learners' interest more than any other topic, presumably due to a somewhat impressive close-up animated image of the vocal folds movement, as well as to a practical exercise, in which, by pressing the fingers on the throat, the person was supposed to feel the folds vibrating at different rates based on the tone used. In addition, while pronouncing the word *Wednesday*, students were asked to identify whether and where in the word the

folds would stop vibrating, thus understanding the difference between *voiced* and *unvoiced* speech sounds (see Appendix C).

Praat was introduced in broad terms, since its functions would be more thoroughly explored in the following lesson. The researcher then, opened the program on her computer, providing the class with a first practical example of how to record a sample word using the software. Any reference to which aspects of pronunciation would be the objective of the in-class analysis was avoided, in order to have participants speak as naturally as possible. Learners were reminded that Praat can only be used on PC, as neither a tablet nor a mobile version of the software are available at present.

Phase 1: Initial Recordings

The items to be recorded are a list of *questions*, belonging to four different question types, each preceded by a context of reference and a single word taken from the same question. All the items are, in turn, linked together by a more general context (see Appendix D).

Example of word and question with context:

4a) *Portuguese*

Context: you are standing in the student residence hallway when you hear two girls speaking a language that draws your attention. You walk up to them and are pretty sure to have recognized one of your favorite languages. For confirmation, you take courage and ask them:

4b) *You're speaking Portuguese, aren't you?*

The recording of single words was only required in this initial phase, being considered a useful exercise for participants, who were supposed to familiarize with the software. On the other hand, the inclusion of a context for each question would acquire a more important meaning in the re-recording phase, that is, once the subjects had learnt about its relevance in relation to different intonation patterns.

It should be noted that, although initially the subject participants were expected to improve their English pronunciation with respect to three variables, i.e., duration, rhythm (stress placement) and intonation of questions, the data analyzed in the next chapter will regard exclusively pitch contours, being pitch the most straightforwardly accessible suprasegmental aspect through the use of Praat. In addition, it was considered

appropriate to limit the range of English intonation patterns to three major types, namely the *rising*, *falling* and *flat* tone, since referring also to the *falling-rising* and the *rising-falling* tones could have been overly complex.

The 8 questions include (Hewings, 2007; Hamlaoui & Bengrait, 2016; Larassati et al., 2022):

- 2 *wh- questions*: ending with a *falling* intonation contour, when used to find out information;
- 2 *yes/no questions*: ending with a *rising* intonation contour, when used to make sure about something;
- 2 *tag questions – up tags*: ending with a *rising* intonation contour, when the speaker is not sure about something and really wants to ask a question;
- 2 *tag questions – down tags*: ending with a *falling* intonation contour when the speaker is sure about their statement and expects confirmation.

Following the written guidelines and tutorial instructions, students were directed to: download Praat (see Appendix E); record each word and question separately, naming each temporary file with the same label used in the item list and their surname (e.g., 4a_Surname, 4b_Surname); save the temporary files in the *Praat Objects* window (see subsection 1.4.2) as wave files (WAV) for later playback and uploading on Google Classroom (see Appendix F). Moreover, as a hands-on introduction to some of the basic features of Praat for acoustic analysis, learners were invited to visualize one or more audio files, focusing on the waveform display, and follow some easy procedures, such as: moving the pointer to listen to different parts of the waveform, editing the waveform by cutting unnecessary edges (i.e., silent fragments), and reflecting upon possible segmental and suprasegmental associations between the waveform contour and the actual sounds with the help of the zooming function (see Appendix F).

At the end of phase 1, the number of students having completed the recordings was 12 for 2A class and 10 for 2B class. Among these, only 5 subjects from 2A and 4 subjects from 2B carried out the task at home, hence autonomously (uploading the files on Google Classroom), while the rest asked whether it was possible to do the exercise during school hours (not having a PC was one of the reasons). These subjects, then, sat next to the researcher, either in the school corridor or in an empty classroom, and used her computer to record the target items through Praat.

Phase 2: In-Class Guided Analysis

The objectives of the in-class analysis were threefold: 1) to have students focus on visualizations of their own productions, in an attempt to read some basic acoustic patterns; 2) to provide learners with audio-visual feedback on intonation by means of a comparison between their own productions and those of native English speakers; 3) to find a rule of thumb for English pitch patterns in relation to specific speech contexts.

As for the speech models, three native speakers (two females and one male) with different accents (British and Australian) made themselves available for recording. The productions selected as final models were chosen on the basis of their sound clarity and intonation precision.

Having collected in each classroom a number of recordings roughly equal to half of the students, *working in pairs* was regarded as the best solution to allow those learners who had not completed the task to participate actively by following the classmate's voice visualizations. After having edited all the subjects' audio files, by cutting the "silent" edges and adding the pitch contour (using the *show pitch* function), the images were printed directly by the researcher and handed out in class together with different color highlighters to be used during the analysis. It is worth noting that some of the visual representations were not as clear as expected, since samples can easily be contaminated by external noises coming from the environment or the hardware used.

In order to attain the first objective, a visual representation of sentence 1b (recorded by the author) was projected in front of the class (see Figure 7).

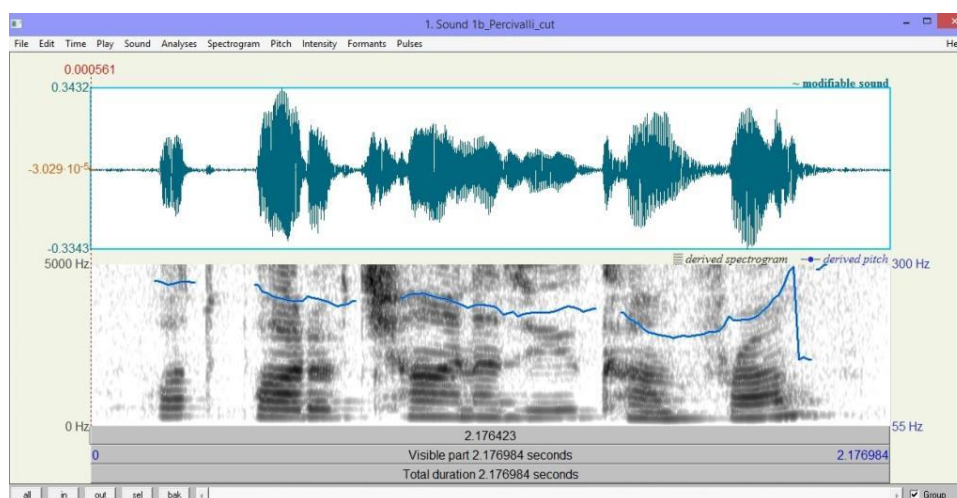


Fig.7. Visual representation (recorded by the author) of sentence 1b:

What part of China do you come from?

Through a series of questions paralleling those used in the at-home assignment, students' attention was drawn to the *waveform contour* and in particular to those wave portions pronounced with greater *intensity* (or “higher volume”). As an hands-on exercise, learners were invited to draw a circle around the more intensely pronounced parts of their waveforms, coming to the conclusion that, in speech, sound intensity is one major indicator of *word* and *sentence stress*.

Another acoustic feature of speech examined together was the difference between *voiced* and *unvoiced* sounds. By zooming into the word *China* and further into its first two phonemes, i.e., /tʃa/, pupils were shown how the wave fragment representing /a/ was characterized by a constant regular pattern, compared to the shape of /tʃ/, being instead highly irregular (see Figure 8). The reason of such visual discrepancy was understood once again by means of “the vocal folds exercise”: by touching their throat while uttering the two sounds, it became clear that the regular pattern reflected the vocal chords vibration, thus indicating a voiced sound, while the irregular one was the result of an unvoiced sound.

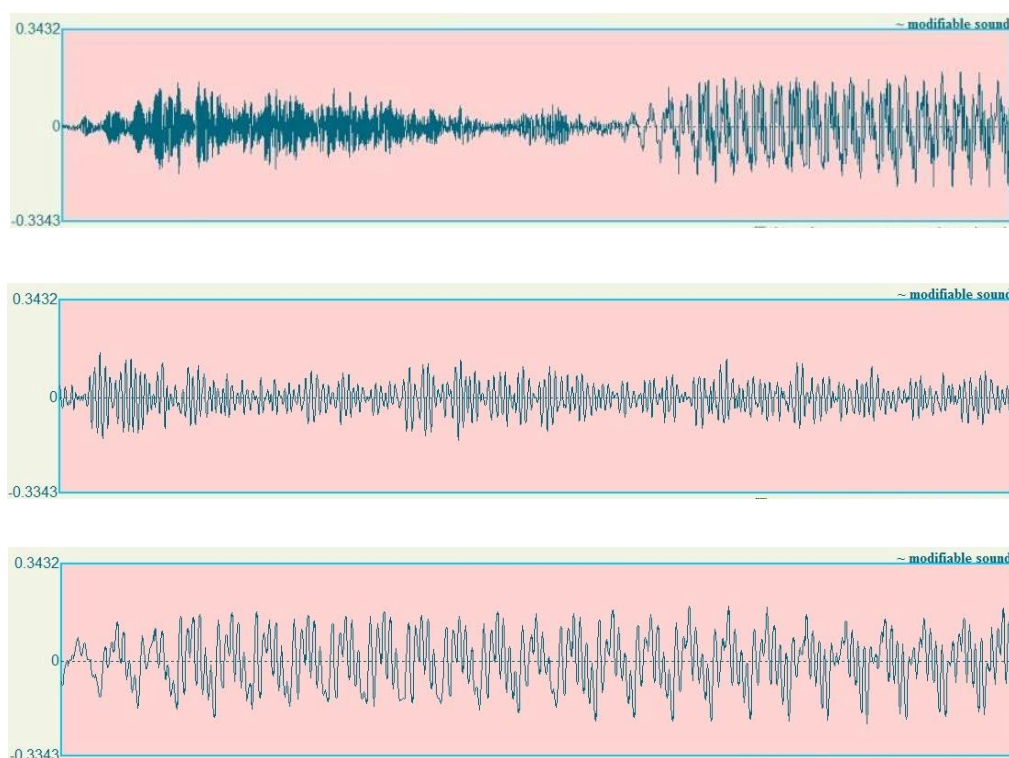


Fig.8. Zoomed-in images of /tʃa/, /tʃ/ and /a/ from the word *China*

Moving to the lower box of the display, attention was drawn to the *show pitch* function, used to plot the *intonation contour* (see the blue line in Figure 7). When asking why the contour was not a continuous line, several of the students were able to link this trait with the unvoiced segments. In relation to the second objective of the analysis, the following exercise focused on the *final pitch movement* of the eight questions. Learners were invited to take the sheet called *Guess the Intonation* (see Appendix G) and were introduced to three main final intonation patterns, namely *rising*, *falling* and *flat* (indicated respectively by \nearrow \searrow \rightarrow). After highlighting the final part of the pitch line on each handout, the task consisted of writing down the patterns used in the first column of the sheet, choosing among one of the three arrows. Having done that, the native speaker models were played, asking learners to guess and annotate (with the proper symbol) the final intonation pattern solely through audio input (hence the ear icon). At this point, visualizations of the English speakers productions were “played” (hence the eye icon), in order to confirm or disprove the aural hypotheses by comparing the arrows in the second and third column. The last column was meant for students to check the boxes in case their questions’ final pitch would match the model patterns.

Predictably, the intonation of wh-questions and down tag questions were found to deviate greatly from the English models, being most commonly uttered by Italian speakers with a rising tone. The class was thus guided to identify what elements the questions could have in common (e.g., questions 1b and 5b starting with a question word and ending with a falling tone) and to understand the role played by the context given in the choice of final pitch patterns. To do this, help was provided by the sheet called *Intonation + Context* (see Appendix H), in which intonation patterns (\nearrow \searrow) of each question type were linked to explanations of specific contexts of use. The last fill in the gaps exercise consisted in choosing the right question number by re-reading the contexts preceding the eight items (1b–8b) and finding the matching explanation.

Example of context matching:

Context of question 4b: you are standing in the student residence hallway when you hear two girls speaking a language that draws your attention. You walk up to them and are pretty sure to have recognized one of your favorite languages. For confirmation, you take courage and ask them: *You’re speaking Portuguese, aren’t you?*

Context of use linked to “not real” question tags: they are not real questions; who is asking these questions (using a final falling pitch ↘) is almost certain about the answer, they expect confirmation by their interlocutors.

Phase 3: Practice and Re-Recording

To put into practice what they had learnt until then, the students' task was now to carry out an audio and visual comparison between their initial recordings and the English speakers' productions, after which they would re-record their questions in an attempt to sound more native-like. To this purpose, the English speakers' audio files were uploaded on Google Classroom together with written guidelines and a new tutorial (see Appendix I). It is worth pointing out that this part of the activity refers uniquely to those pupils who had completed the first part of the recordings.

Following the instructions, learners were guided to import the files in Praat and simultaneously visualize both their productions and the corresponding native speaker's voice. Further directions included: comparing the questions duration; considering the number of parts forming the two waveforms and their relative intensity; paying careful attention to the final part of the pitch line. One main suggestion was to record and analyze as many times as necessary in order to obtain a good result.

At the end of phase 3, the number of students having completed the recordings was 6 for 2A class and 6 for 2B class, namely the actual subject participants to this experimental activity. Moreover, among these, only 3 subjects from 2A and 3 subjects from 2B carried out the task at home, hence autonomously (uploading the files on Google Classroom), while the rest preferred to do the exercise during school hours using the researcher's PC.

Phase 4: assessment of skills through dialogue reading

The final session centered on assessing whether training intonation with Praat on isolated sentences would transfer to discourse. To this end, students were asked to read two dialogues containing questions ending with either a rising or a falling tone, paying particular attention to the context of the discourse and the speakers' intentions (see Appendix L and M). To help participants guessing the proper intonation pattern, each sentence in bold had to be matched to one of the four question types by filling in the empty boxes with the right arrow symbol. Having done that, some learners volunteered

DATA ANALYSIS

This chapter consists of three sections. In the first section, the oral qualitative data gathered from the interviews will be analyzed; the interview transcripts will be interpreted holistically according to the major themes addressed and complemented by the quotation of some significant passages. In the second section, a comparative analysis of the twelve subjects' initial and final recordings will be conducted, in which the data collected will regard the number of accurate and inaccurate realizations of final pitch movements before and after students' introduction to Praat; these quantitative data will be elaborated through descriptive statistics in the form of percentages. In the third and final section, the numerical data obtained by the compilation of the questionnaire will be discussed and particularly relevant findings will be displayed in percentage bar graphs.

3.1 Interview Data

Of initial consideration were the questions concerning the *organization of English teaching throughout the school path and the role held by its components in the curriculum* (Q.3-4-5). In this regard, both Teacher A and Teacher B indicate that, during the first two years, English instruction should aim at taking pupils towards the B1 level of the CEFR, while the goal to be reached by the end of the fifth year should be the B2 level. In order to pursue these objectives, particular attention is initially placed on the language (e.g., grammar and morphosyntactic structures, listening and pronunciation) and oral questioning tends to focus on texts about current events, characterized by a generic form of English. Moving to the third year, English literature assumes a central role in the curriculum, becoming, among others, the object of oral questioning. Teacher A claims that not everyone likes this change of focus, considering it too specific and demanding. This, she underlines, is due to the fact that, although the B2 is the level expected to be reached by the end of the school path, learners' English proficiency may vary considerably: there might be students who, by the end of the second year, are already able to show B2 language competence, and others who, in the fifth year, are still struggling.

Teacher A: "Alla fine del triennio dovrebbero arrivare al livello B2. In realtà l'utenza che arriva ormai a scuola è molto diversificata, perché ci sono già al biennio dei ragazzi che sono già a un livello B2 o magari al triennio abbiamo dei ragazzi che sono in difficoltà, sono più impacciati."

Related to this, Teacher B argues that investing too much time on literature may imply leaving some pupils behind from a linguistic point of view. Accordingly, he usually reminds his students about the importance of measuring their English proficiency especially based on the richness of their vocabulary and their pronunciation accuracy, in that a more correct pronunciation may mean greater understanding.

Teacher B: “Io dico sempre che il loro inglese devono misurarlo, prima di tutto nel numero di vocaboli e nella qualità della pronuncia, poi nella capacità grammaticale. Per me la priorità è l’espansione lessicale e la correttezza nella fonologia, non a scopo cosmetico, ma perché la pronuncia corretta implica una comprensione corretta.”

A second major theme addressed was the *students’ attitude towards English learning and, more specifically, towards English pronunciation*, both with reference to the classroom in question (Q.6-7) and in general (Q.8-9-10). While Teacher A highlights an active participation during English lessons on the part of 2A class, Teacher B describes 2B class as collaborative and curious about his subject. Overall, the two classes seem to share a certain degree of intrinsic motivation towards the learning of English, which very often results in learners asking numerous questions about, for instance, informal expressions or slang terminology. This would happen because an ever-growing number of students are daily exposed to the FL outside the school environment. In fact, Teacher A asserts that her pupils are used to watching many videos in English, especially those about contemporary artists exhibiting their works (due to their course of study), as well as listening to songs and joining online chats where they write in English. They thus get to learn lots of current words, to the point that they themselves are often telling the teacher when some of the terms presented by the textbook as teen slang are already obsolete.

Teacher A: “A volte il libro di testo presenta delle parole che gli adulti, i *boomers*, non conoscono e i giovani sì, ma sono i ragazzi stessi a dirmi che anche questo libro è obsoleto perché queste parole non si usano già più.”

Compared to the past, Teacher A observes that, even in the absence of a proper educational background in English from middle school, many students are now showing improved listening comprehension skills from the start.

Teacher A: “Ci arrivano magari non con una preparazione scolastica tradizionale, però con molto orecchio e anche capacità di comprensione e di ascolto migliorate rispetto agli anni precedenti. Non per tutti, ma statisticamente parlando, una grossa percentuale adesso trova più facile l’ascolto e la comprensione.”

Similarly, Teacher B claims that his pupils are used to streaming movies and TV series through Netflix and YouTube, as well as joining online chat and videogames in English. He further points out that, thanks to the widespread availability of English input, the number of students in every classroom showing greater aptitude for foreign languages has increased considerably; he talks about four or five persons per class, which compared to the one or two of a few years ago, could make a big difference, since this appears to have a driving effect on the whole class, even on those learners who most lack interest or motivation.

Teacher B: “Mentre ricordo che all’inizio del mio insegnamento il problema della pronuncia era terrificante, non c’era verso di schiodarli, adesso comunque c’è l’effetto indotto — siccome gli altri pronunciano bene, allora non è una cosa secondaria, è importante, senno sono sempre io che pronuncio male, mentre gli altri pronunciano bene.”

In addition, Teacher B underlines how the perception towards English pronunciation has changed over time, as these days many learners seem to be more sensitive to possible social and image-related implications connected to good pronunciation skills.

Teacher B: “... la percezione migliorata dell’importanza della pronuncia anche come fattore sociale, di immagine; sono molto sensibili a questa cosa, è *trendy* e non è una noia in più.”

A final area of consideration regarded the didactic action per se and, in particular, the questions concerning the *influence of the curriculum guidelines on English pronunciation instruction* (Q.11-12), *the techniques and feedback used to teach pronunciation and prosody* (Q.13-14-15), and *the instruments and technologies at teachers’ disposal* (Q.16-17-18). Both interviewees report a gap in the curriculum in terms of pronunciation instruction. Due to the absence of precise guidelines, pronunciation is hence conceived as a language component to be only occasionally mentioned, such as when it needs correction. Connected to this, the total amount of time dedicated to pronunciation during one English class is estimated by both teachers to be around 8 minutes, intended as irregularly distributed throughout the lesson. Moreover, although Teacher A states that until a few years ago she used to introduce her pupils to the IPA from the outset, nowadays a phonetic approach would seem to be counterproductive, in that, no longer using paper dictionaries, learners prefer by far to listen to audio pronunciation provided by online dictionaries (*WordReference* being the most commonly used in class time). To meet this general preference for listening, Teacher A and Teacher B are constantly proposing listening and repetition activities to

their pupils, as it is also considered an effective way to learn pronunciation and prosody implicitly. This approach is made easier by the rich multimedia content offered by the textbook, which is regarded by both instructors as a valid and updated instrument, being replaced every three years.

Teacher B: “D’altra parte è importante sapere che i libri che adesso ci propongono ... fanno un’enorme differenza, perché il 65% della mia attività in classe è di *listening*, cioè le nostre attività partono sempre da un ascolto o un video. Questo significa un bel rafforzamento della capacità di ascolto e indirettamente di pronuncia.”

Finally, as for the type of feedback used, worth noting is that great attention is placed on preventing anyone to be hurt by the teacher’s judgement. Teacher A says that she prefers to wait until the students have finished talking before giving any pronunciation feedback, in order to let them gain in confidence.

Teacher A: “Come feedback generalmente lascio sempre che finiscano il loro discorso e alla fine intervengo. Ovviamente facendo così qualcosa mi sfugge, ma ho visto che l’interruzione costante ogni volta che uno sbaglia la pronuncia blocca il ragazzo, quindi alla fine, anche se non pronuncia bene, preferisco comunque che acquisisca sicurezza ... quando uno è interrogato è sempre un po’ mortificato dalla correzione, è sempre comunque una critica.”

On the other hand, Teacher B seems to intervene more assiduously on pronunciation errors, to the point that, during group activities, faulty pronunciation is, though tactfully, always corrected. In the case in which the same errors persist over time, he usually decides to dedicate part of the lesson focusing exclusively on them, writing examples on the whiteboard and doing some gags to draw students’ attention.

Teacher B: “Ogni volta che viene sbagliata una pronuncia, senza farlo pesare, la rileggo o la correggo e se dopo un mese viene ancora sbagliata, fermo tutto, mi metto alla lavagna, faccio un po’ di scena così rimane impresso e ho tutta la classe che si mette in moto. In questo modo vedo dei progressi nella pronuncia.”

When the feedback is personal, Teacher B first considers the pupil he is relating to; if the person is fragile, instead of giving immediate feedback, he makes sure that before the end of the lesson the same problematic sounds are tackled by the whole class within other contexts, so that personal feedback, though indirectly, is still provided.

Teacher B: “Se è una persona fragile, che ha già problemi, evito e magari aspetto entro la fine della lezione e trovo il modo di far ricapitare quelle due parole. I ragazzi questo l’hanno capito. Prima della fine della lezione, faccio una frase di grammatica e viene fuori di nuovo la stessa pronuncia, la correggiamo e vedi nuovamente gli sguardi che brillano come dire — ah non se ne era dimenticato, perché mi pareva che il mio compagno avesse sbagliato.”

3.2 Pronunciation Activity Data

In the collection of data stemming from the pronunciation activity, the researcher decided to narrow down the scope of the analysis only to *intonation contours*. This was due to the fact that, although variations in terms of *duration* and *rhythm* to sound more native-like were expected at the end of the practice and re-recording phase (as much as variations in final pitch patterns), the relatively low number of participants to the final recordings and the impatient attitude showed by some of the students completing the task in presence of the author, suggested that the subjects' effort would focus primarily on the activity most enjoyable topic, in other words intonation.

To perform the analysis, the *initial* and *final recordings* of the 8 questions (i.e., 1b, 2b, 3b, 4b, 5b, 6b, 7b, 8b) produced by 12 students were compared. The 12 subjects, who for anonymity will be referred to as Student 1–Student 12 (or S1–S12), were the only ones to have completed the whole task, meaning that, besides having recorded the questions in a spontaneous way during phase 1, they had attended the in-class guided analysis with Praat of phase 2, thus re-recording the questions in phase 3 in light of the lessons learnt. During this last passage, the subject-participants could count both on the audio-visual representations of native speakers' productions and the context provided for each question, which was now expected to give further clues on whether to choose a final rising or falling intonation pattern.

Worth noting is that only the researcher performed the evaluation. The improvements were measured by comparing the pitch contour displays of final recordings to the pitch contour displays of initial recordings (indicated in the appendices as Time 2, T2 and Time 1, T1 respectively) in relation to the native speakers' pitch trace. As showed in the Appendices O–Z, while in some cases, visual evaluation could be performed effortlessly and concerned much of the contour, in some other cases, detecting an improvement visually was not so immediate, which made the listening of audio files essential for the assessment.

These visual issues might have been caused by several reasons, such as:

- *differences in voice range*, whereby, given a particular frequency range, a more female pitch (generally higher) will appear further up compared to a more masculine pitch (notoriously lower); to be considered that, since the experiment in question involved eight female and four male students, the pitch range was set

between 40 and 400 Hz and only when needed (due to screen-space reasons) changed to 30–300Hz.

- *external contamination of the signal* and/or *software miscalculation*, causing sudden spikes in the pitch trace or brief sequences of much higher or lower contour that cannot be heard by the listener, which in turn may hinder the reading of intonation displays.

Against this background, the researcher opted to focus her assessment exclusively on the contours final portion, which for each question corresponded to (written in italics):

1b) What part of China *do you come from?*

2b) Do I need to book *the laundry room?*

3b) Northern Ireland is not part of Europe, *is it?*

4b) You're speaking Portuguese, *aren't you?*

5b) How many courses *are you taking?*

6b) You don't like green, *do you?*

7b) Are we allowed *to have parties?*

8b) The bus comes every thirty minutes, *doesn't it?*

Having said that, rather than providing a precise and unbiased account of a comparative evaluation of intonation contours, the following analysis aims at presenting pilot data possibly useful for similar pedagogical applications of speech analysis software.

The results are presented in four sections, each dedicated to a specific question type: the first section includes data from *wh-question* analysis; the second section shows the results of *yes/no question* analysis; the third and fourth sections concern data gathered from *up* and *down tag question* analysis respectively.

Each table, referring to one single question, displays: the number of required accurate realizations of the intonation pattern, the number of accurate and inaccurate realizations on the part of the subjects, before and after they were introduced to Praat, and its conversion to percentage.

Analysis of wh-questions

This section discusses the results emerging from the analysis of questions 1b and 5b. Based on the context given, the two sentences should end with a falling tone, as they assume the character of *finding out* questions (Hewings, 2007). The required number of accurate intonation pattern is 12, that is one accurate pronunciation for each student.

As seen in Table 1, the realization of accurate final tone for the question *What part of China do you come from?*, before the students' introduction to Praat, is only equivalent to 33% of the data, in other words 4 out of 12 recordings. After the in-class analysis and training with Praat, the number of accurate realizations increases to 10, that is 83% of the data.

	Wh-Question 1b			
	Before Praat		After Praat	
	Accurate	Inaccurate	Accurate	Inaccurate
Required	12	0	12	0
Realization	4	8	10	2
Percentage of realization	33%	67%	83%	17%

Table 1

Similarly, as shown in Table 2, the realization of accurate final tone for the question *How many courses are you taking?*, before the students' introduction to Praat, is equivalent to 42% of the data, that is 5 out of 12 recordings. After the in-class analysis and training with Praat, the number of accurate realizations increases to 9, corresponding to 75% of the data.

↙	Wh-Question 5b			
	Before Praat		After Praat	
	Accurate	Inaccurate	Accurate	Inaccurate
Required	12	0	12	0
Realization	5	7	9	3
Percentage of realization	42%	58%	75%	25%

Table 2

These results indicate that, given the initial tendency of many subjects to pronounce wh-questions following a more Italian-like final rising tone, practicing with Praat software proved to be beneficial to a considerable number of students, who eventually were able to produce the correct final falling intonation.

Analysis of yes/no questions

In this section, the data derived from the analysis of questions 2b and 7b are presented. Based on the context given, the two sentences should be pronounced with a final rising tone, as they assume the character of *making sure* questions (Hewings, 2007). The required number of accurate intonation pattern remains 12.

Table 3 reveals how the realization of accurate final tone for the question *Do I need to book the laundry room?*, before the students' introduction to Praat, already reaches high percentages (i.e., 67%), as 8 subjects out of 12 utter the sentence ending with a rising pitch. However, the training with Praat seems to allow this rate to increase up to 100%, since all of the 12 recordings, at this point, present a final rising intonation.

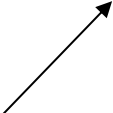
	Yes/No Question 2b			
	Before Praat		After Praat	
	Accurate	Inaccurate	Accurate	Inaccurate
Required	12	0	12	0
Realization	8	4	12	0
Percentage of realization	67%	33%	100%	0%

Table 3

Equally positively, Table 4 shows that, despite a lower starting percentage of accurate final tone, equivalent to 42% of the data, after the subjects' introduction to Praat, again this rate reaches 100%, as all of the 12 recordings end once more with a rising pitch.


	Yes/No Question 7b			
	Before Praat		After Praat	
	Accurate	Inaccurate	Accurate	Inaccurate
Required	12	0	12	0
Realization	5	7	12	0
Percentage of realization	42%	58%	100%	0%

Table 4

Arguably, in the case of yes/no questions, more positive outcomes are likely to have been facilitated by a familiar final rising tone, rather typical of Italian questions. The students' tendency to transfer their native language pitch patterns into English would explain the relatively high number of initial accurate realizations. Nevertheless, the total accuracy found after practicing with Praat would appear to support the effectiveness of audio-visual feedback in intonation enhancement, despite the presence of intonational similarities between the FL and the L1.

Analysis of tag questions - up tag

The term *up tag* refers to the intonational realization of a short question form added to the end of a statement. Based on the context associated with questions 3b and 8b, their tags should be pronounced with a rising tone, in that, like yes/no questions, they are meant to *clarify information* (Hewings, 2007). The required number of accurate intonation pattern remains 12.

Interestingly, Table 5 displays the exact same data as Table 3, in that 8 students out of 12 (i.e., 67%) utter the question *Northern Ireland is not part of Europe, is it?* with the correct final tone, even before being introduced to Praat. However, after the training, 100% of the recordings, now ending with a rising pitch, can be said to resembling the native speaker model.

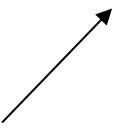
	Up Tag Question 3b			
	Before Praat		After Praat	
	Accurate	Inaccurate	Accurate	Inaccurate
Required	12	0	12	0
Realization	8	4	12	0
Percentage of realization	67%	33%	100%	0%

Table 5

In a similar fashion, Table 6 shows that, though starting from a slightly lower percentage of accurate realizations before Praat (i.e., 58%), noticeable progress are made in the attempt to achieve the target intonation of the question *The bus comes every thirty minutes, doesn't it?*. In fact, after practicing with Praat, the number of accurate realizations increases from 7 to 11, meaning that 92% of the sample is able to utter the question with the proper final rising pitch.

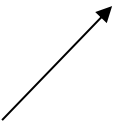
	Up Tag Question 8b			
	Before Praat		After Praat	
	Accurate	Inaccurate	Accurate	Inaccurate
Required	12	0	12	0
Realization	7	5	11	1
Percentage of realization	58%	42%	92%	8%

Table 6

As with yes/no questions, these results may reveal greater naturalness on the part of Italian subjects in the production of questions characterized by a final rising tone, in that, to a fairly large extent, up tags were pronounced accurately from the start and, after Praat training, they seem to have been corrected without major difficulties.

Analysis of tag questions - down tag

The term *down tag*, conversely, refers to the falling tone with which the short question forms added to the end of statements should be uttered. This would be the case of questions 4b and 6b, since, according to their context of reference, their aim is to *seek agreement with the hearer* about the statements (Hewings, 2007). The required number of accurate intonation pattern remains 12.

As illustrated in Table 7, the realization of accurate final tone for the question *You're speaking Portuguese, aren't you?*, before the students' introduction to Praat, regards only one single student, while inaccurate realization, i.e., produced with a final rising pitch, is equivalent to 92% of the sample. After practicing with Praat, the number of accurate realizations increases from 1 to 7, meaning that 58% of the recordings presents a correct final falling tone.


	Down Tag Question 4b			
	Before Praat		After Praat	
	Accurate	Inaccurate	Accurate	Inaccurate
Required	12	0	12	0
Realization	1	11	7	5
Percentage of realization	8%	92%	58%	42%

Table 7

On a more positive note, despite gathering the same initial data as for question 4b (i.e., 11 inaccurate final tone realizations out of 12), Table 8 displays better results for the question *You don't like green, do you?*, since, after the subjects are introduced to Praat, the number of accurate realizations increases from 1 to 9, meaning that 75% of the subjects are now able to pronounce the tag using a correct final falling intonation.

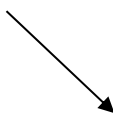
	Down Tag Question 6b			
	Before Praat		After Praat	
	Accurate	Inaccurate	Accurate	Inaccurate
Required	12	0	12	0
Realization	1	11	9	3
Percentage of realization	8%	92%	75%	25%

Table 8

The results of this last section reveal that the intonation of down tag questions might be a little more challenging than the intonation of wh-questions, the only other category ending with a falling tone. This could be due to the fact that, besides trying to mimic a less familiar pitch pattern, the subjects are dealing with a question type, which, if not completely missing, assumes rather different forms in Italian.

Taken together, these findings highlight that the 12 students faced more difficulties in pronouncing questions characterized by a final falling pitch, compared to questions ending with a rising pitch. In fact, besides scoring much lower rates of accurate final tone realization before the training, these questions never reached 100% of intonation accuracy after the practice with Praat. In contrast, pronunciation of yes/no questions and up tag questions seemed to be more straightforward, as 3 out of 4 sentences reached 100% of intonation accuracy after Praat training. As mentioned above, such score difference may be linked to language interference, which in the case of Italian learners would facilitate the mimicry of final rising tones in questions, while hindering that of final falling tones. Nevertheless, working with Praat can be argued to have led to an overall enhancement of intonation accuracy, as even the question with the lowest final score (i.e., down tag question 4b) saw a 50% improvement in intonation accuracy, thus confirming the efficacy of audio-visual feedback in pronunciation training.

3.3 Questionnaire Data

The final goal of this research was to determine whether the pronunciation activity as a whole and Praat use had been welcomed favourably by those learners who had taken part in all steps of the study, despite their young age.

For this purpose, an 18-item anonymous questionnaire was issued on line, in which the 12 respondents were asked to express their degree of satisfaction on four major areas of analysis. All responses were presented on a scale from 1 to 5, with 1 signifying *not at all*, 2 *a little*, 3 *rather*, 4 *a lot* and 5 *very much*.

The questionnaire was submitted by all 12 students, however, the answers of two of them were missing, indicating perhaps a misunderstanding of the instruction. Therefore, the number of actual respondents was 10. Questionnaire results are presented in four sections, each referring to one area of analysis.

Overall perceived benefits of the pronunciation activity

A first theme investigated concerns the pronunciation activity as a whole and the degree of perceived benefit deriving from it. It includes item 1, 2, 3, 4, and 15 of the questionnaire, which, compared to the other three item categories, can be said to have received the highest assessment.

Given the way in which questionnaire statements were formulated, positive responses are to be considered those assigning a value ranging from 3 to 5 (i.e., 3 matching a *rather positive* feedback, 4 a *positive* feedback and 5 a *very positive* feedback).

According to this data interpretation, three out of four items obtained one negative evaluation (i.e., below 3), while most of the positive evaluations ranged from 4 to 5, such as for item 1 (*Overall, the pronunciation activity was useful*), item 2 (*Overall, the pronunciation activity made me reflect on my voice potential*) and item 3 (*Overall, the pronunciation activity made me reflect on my English pronunciation*). However, particularly noteworthy is that, on the one hand, respondents were more cautious in judgment regarding item 4 (*Learn more about rhythm and intonation was useful to reflect on my English pronunciation*), assigning an equal number of responses for each of the three positive scale values (see Chart 1), while, on the other hand, they showed greater consensus about item 15 (*Overall, I enjoyed participating in the pronunciation*

activity), which received the highest number of very positive feedback, namely 70% of the responses (see Chart 2).

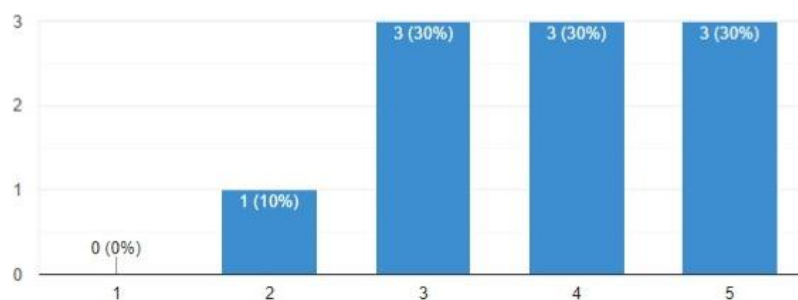


Chart 1. Responses to item 4

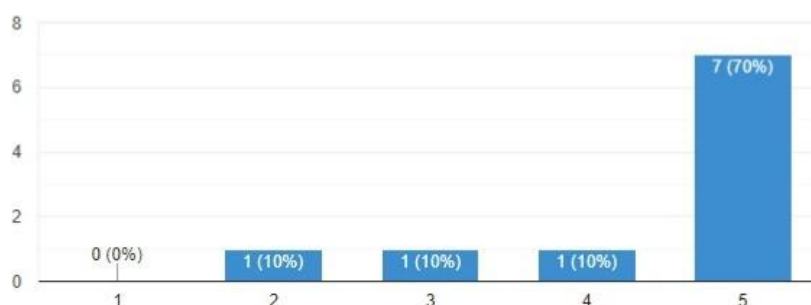


Chart 2. Responses to item 15

Taken together, these results indicate that students perceived the activity as useful, having served as a source of reflection both on their own voice potential and their English pronunciation. The specific role of rhythm and intonation in pronunciation learning remained probably unclear to a certain extent, being aspects of the FL rarely addressed during classes. Nevertheless, the majority of respondents seem to have enjoyed this experience.

Overall gains from the use of Praat in pronunciation learning

The second area of analysis focuses specifically on the potential advantages of using Praat as a tool for learning English pronunciation and includes item 5, 6, 7 and 8 of the questionnaire.

Even in this case, the overall assessment may be regarded as positive, since three out of four statements received one single negative evaluation (i.e., below 3). However, the number of responses for each of the three positive values are more equally distributed, such as for item 5 (*Overall, the use of Praat software was useful*) and item 6 (*The use of*

Praat software helped me reflect on my voice features). Standing out are item 7 and 8: the first (*The use of Praat software helped me reflect on my English pronunciation*) receiving a slightly higher number of very positive responses (see Chart 3), the second (*Praat functions displaying soundwaves and intonation contours helped me improving my English pronunciation during the activity*) obtaining, instead, a high percentage of rather positive responses (see Chart 4).

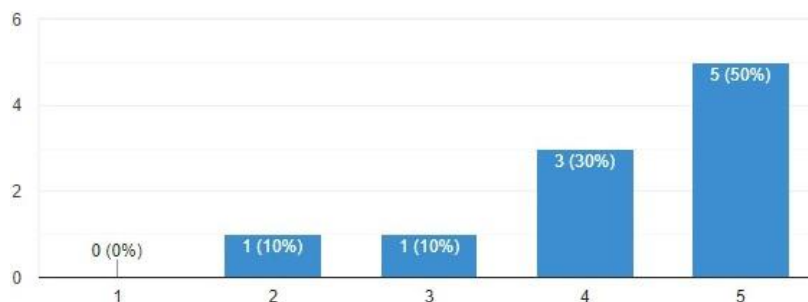


Chart 3. Responses to item 7

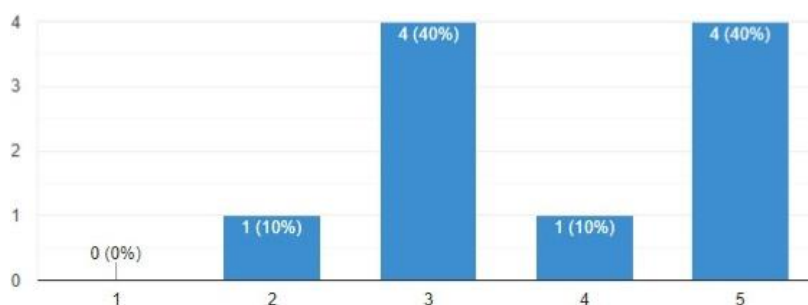


Chart 4. Responses to item 8

Collectively, these data reveal an overall understanding of the potential advantages that Praat can offer to students in the scope of FL pronunciation learning. The fact of being able to count on visual feedback, in addition to audio feedback, was probably considered an effective method by the respondents to conceptualize pronunciation. However, specific Praat functions allowing learners to see their voice still seemed to arouse some perplexity, perhaps due to reasons related to both the novelty of the method and the software complexity (addressed in further detail below).

User-friendliness of the Praat software

A third theme of interest regards the modality in which the software interfaces with the user. It includes items 9, 10, 11, 12, 13 and 14 of the questionnaire, which, compared to

the other three item categories, can be said to have received the lowest assessment. In fact, among the statements collecting negative feedback (five out of six), three showed only one negative evaluation, while two statements received two negative evaluations, namely item 9 (*Downloading Praat software was easy*), as seen in Chart 5, and item 11 (*Visualizing soundwaves and intonation contours with Praat was easy*), as seen in Chart 6.

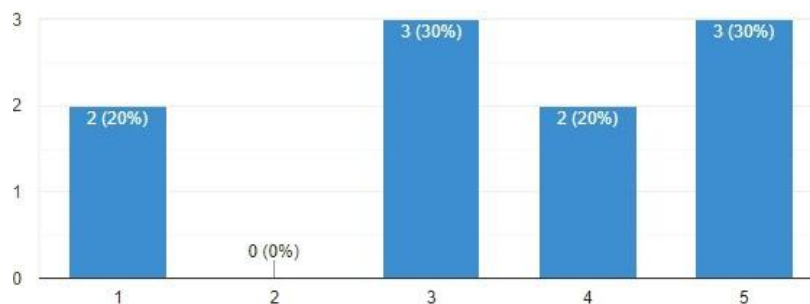


Chart 5. Responses to item 9

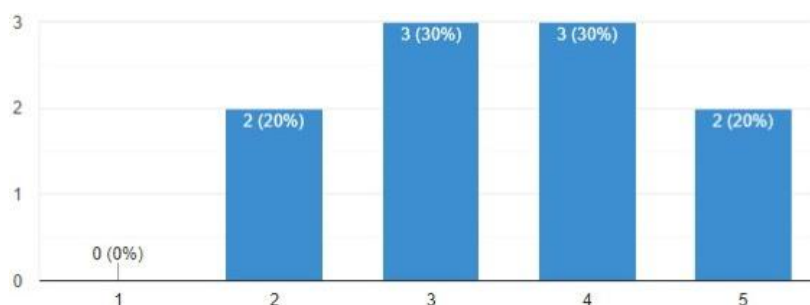


Chart 6. Responses to item 11

As for the number of responses for each of the three positive values, a fairly even distribution may be noted, such as in the case of item 12 (*Comparing my own pronunciation to the native speakers' pronunciation by means of Praat visualizations was easy*) and item 14 (*Overall, using Praat was easy*). Exceptions can be considered item 10 and 13, since the first one (*Recording my voice with Praat was easy*), showed a slightly higher number of rather positive values (see Chart 7), while the second one (*The tutorials illustrating the use of Praat were useful*) obtained a clear prevalence of very positive values (see Chart 8).

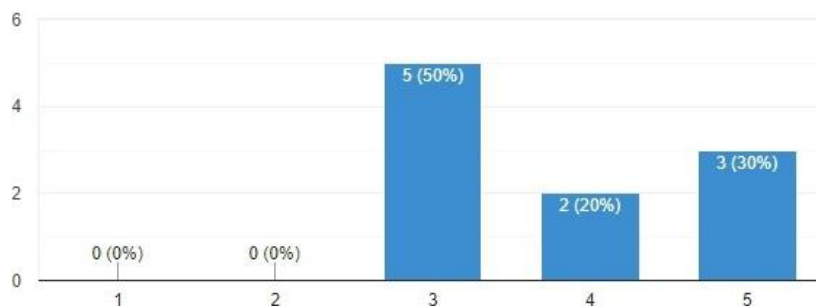


Chart 7. Responses to item 10

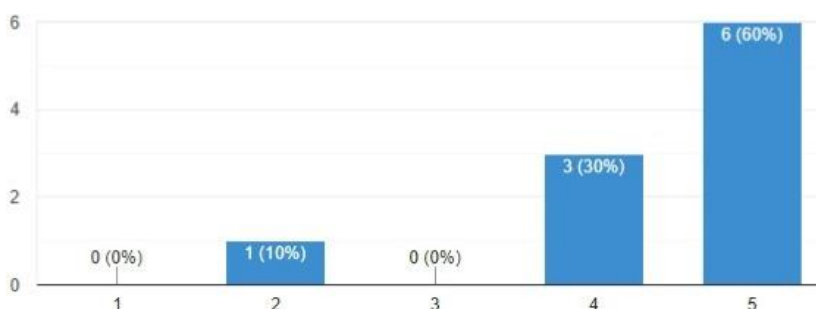


Chart 8. Responses to item 13

Taken as a whole, these results confirm that respondents were not particularly intimidated by the supposed user-unfriendliness of the software, as the majority of them were able to carry out the various tasks. If presented with due precautions, it seems that Praat could rather effortlessly be used even by younger students. Nevertheless, some technical problems were reported, specifically with regard to both the software download and visualization functions. Perhaps unsurprisingly, the importance of guiding tutorials emerged, highlighting that written guidelines alone might have been insufficient to understand the functioning of a program indeed complex in some respects.

Future outlook on pronunciation learning and Praat use

The fourth and last area of analysis concerns the extent to which learners wish to deepen their knowledge about what they have learnt and whether that could be taken into consideration for possible future applications. It includes items 16 (*Tackling topics during class concerning English rhythm and intonation more frequently would help me improve my pronunciation*), 17 (*I wish I could use Praat again during class as a tool for English pronunciation analysis*) and 18 (*I wish I could learn more about Praat*

functioning, in order to use it autonomously every time I want to assess my English pronunciation). Results are illustrated in the following bar charts (Chart 9, 10 and 11 respectively).

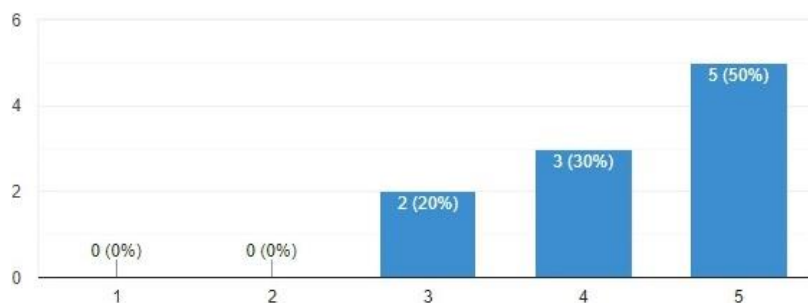


Chart 9. Responses to item 16

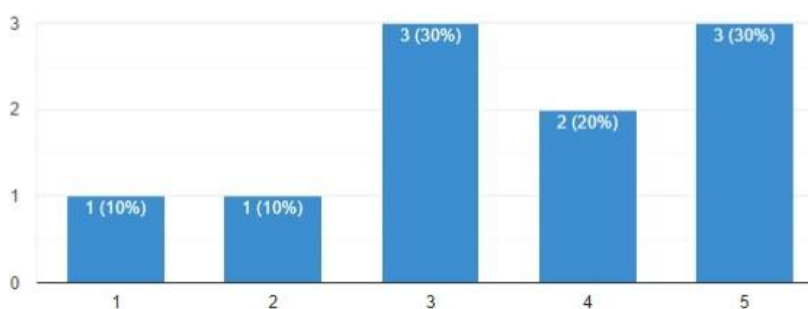


Chart 10. Responses to item 17

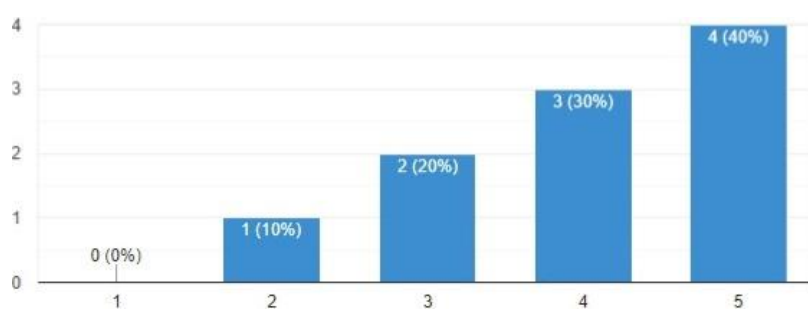


Chart 11. Responses to item 18

As noted from the graphs, statement 16 is the only one reporting no negative feedback and at the same time a high percentage of very positive evaluations. In this regard, it should be reminded that students' assessment also involved the in-class presentation, which seem to have been much appreciated by the audience. Thus, at least on a theoretical level, the respondents would be in favour of tackling these topic more

frequently, considering it helpful for improving their accent. As far as future applications of Praat are concerned, learners appeared more cautious in judgment, especially with statement 17. Again, the absolute novelty of the method and the unexpected obstacles encountered might have had a bearing on the outcomes.

DISCUSSION OF RESULTS

The final chapter of this thesis offers a discussion of the findings in relation to the three research questions. It examines in what way and to which extent the data gathered from each of the three research phases may answer the queries and at the same time it describes the obstacles and limitations faced by the author in carrying out this work. As conclusion, a summary of the main issues addressed in the dissertation will be presented, along with suggestions for future research in this area of study.

4.1 Research Question 1

The first research question explored whether the space dedicated to pronunciation instruction within the two classes being investigated was enough to meet the English language requirements outlined in the school syllabus. Furthermore, it inquired whether pronunciation instruction was favourably welcomed by students.

On the first point, the hypothesis formulated, according to which pronunciation teaching time would be regarded as inadequate by the instructors in relation to the syllabus requirements, has been confirmed.

Interview data reveal that, although the national curriculum of upper secondary education for English language does not include precise indications regarding pronunciation instruction, the school syllabus for this subject features the achievement of level B1 of the CEFR by the end of the second school year. In terms of speech skills, this would imply that “pronunciation is generally intelligible; [the learner] can approximate intonation and stress at both utterance and word levels; however, accent is usually influenced by other language(s) he/she speaks”, quoting the words of the CEFR’s scale for overall phonological control (Council of Europe, 2018, p.134). On this matter, what the two teachers claim is that three hours of English per week are indeed insufficient to address pronunciation exhaustively, given the number of topics to be covered. As a result (at least as far as pronunciation is concerned), only a few learners succeed in achieving the required level.

Nevertheless, both interview and observation findings show that Teachers A and B are very attentive to their pupils’ speaking skills, as they try every class to discuss pronunciation issues despite time constraints. With the aid of the textbook and a good deal of personal initiative, these brief insights appear to have positive effects on pupils’

intelligibility, especially of those who would normally fall behind because less motivated.

These considerations can be said to reflect what several authors have been arguing in recent articles on L2 pronunciation instruction, namely that pronunciation can no longer be considered the underdog or, to use the same metaphor, the Cinderella of the language teaching family (Derwing, 2018; Levis, 2019; Busà, 2021). Classroom observation confirms that the instructors seem to feel at ease when dealing with pronunciation, a picture emerging simultaneously from the interview transcripts. Unlike what previously asserted in the literature (see for e.g., MacDonald, 2002; Busà, 2008a; Derwing, 2010), the interviewees like dwelling on the topic, whenever possible, and have no problem adapting their didactic methods to the preferences of newer generations of learners. However, at the same time they admit the need of additional teacher training regarding some specific issues, such as suprasegmental phonology (Busà, 2008b), as well as further improvements in curriculum development, particularly in the formulation of more detailed criteria and descriptors (MacDonald, 2002).

Conversely, on the second point, the hypothesis for which students would manifest a certain reluctance towards the subject matter due to a sense of inadequacy, has *not* been confirmed.

Field notes reveal that learners show overall good speech skills during class, given their age and the educational context. In addition, the final role-play that took place in 2A class and the active engagement exhibited by 2B class during the teacher's explanation of problematic sounds convey anything but a reluctant attitude towards English pronunciation. Interview transcripts indicate the same state of affairs, in that both instructors describe their pupils as driven by a greater intrinsic motivation compared to past generations, most probably stemming from the massive amount of English input to which they are, even passively, exposed on a daily basis due to the evolution of technology.

In light of this, assertions such as that by Setter and Jenkins (2005), who claimed that “pronunciation is universally considered to be a difficult aspect of an L2 to teach and learn — and possibly the most difficult, for various reasons” (p.1), are nowadays more likely to be called into question. As reported in the current literature (Busà, 2021), there is a growing awareness on the part of L2 speakers of the importance of pronunciation in

speech, both for communication and image-related purposes. In the same vein, the students in question seem to pay much attention to their English pronunciation, to the point that many of them show to be trying their best to sound more native-like.

Finally, it is worth pointing out that these findings are far from being generalizable. The time devoted to classroom observation was undoubtedly limited; the same applied to the number of subjects observed and instructors interviewed. The best option would have been to examine more classrooms of different grades over a longer time period and interview all teachers of English working at the school (originally the author's intention), in order to obtain a more complete picture. However, despite the evident limitations, these results may be deemed suggestive, as they allude to a possible change of course (seemingly already happening) in the teaching priorities of lower-level, general skills language classrooms.

4.2 Research Question 2

The second research question addressed the potential of Praat audio-visual feedback to help learners enhance their awareness on English suprasegmentals as well as improve their prosodic abilities while practicing different types of English questions.

The hypothesis formulated, according to which the use of Praat software would prove its effectiveness both as an awareness raising tool and as a learning aid for improving the prosody of English questions may be considered to have been overall confirmed.

Taken in their entirety, the findings resulting from the comparison of the 12 subjects' productions before and after they were introduced to Praat show substantial improvements in the accuracy of prosody, especially with regard to intonation patterns. Pitch contour visualizations seem particularly useful in helping learners approximate the native English models of those questions characterized by a final falling tone, generally approached with greater difficulty by Italian speakers. Moreover, despite not being analyzed in detail for the reasons outlined in section 3.2, pitch traces reveal, in many cases, improvements in terms of rhythm, identifiable with tone peaks matching word and/or sentence stresses of native productions.

These results would also confirm that the software can help L2 speakers develop sensitivity towards suprasegmentals, notoriously harder than segmental features to recognize even for native speakers. In fact, as Busà (2008b) points out, the visual component of feedback “allows learners to grow an awareness of their distance from the

target language and stimulates self-monitoring and auto-correction” (p.124). Interestingly, as observed by the researcher in those subjects completing the recordings using her PC, audio-visual feedback proved to be important even in case of failure. When unable, to their surprise, to reproduce the correct final pitch pattern even after multiple attempts, these learners were aided by visual representations to acknowledge their perception’s possible unreliability and also the challenge behind an apparently simple task, such as modifying their own voice melody. It may additionally be argued that Praat served its awareness raising function also during the in-class analysis, when most students showed great interest in understanding the link between tone variations and pitch trace movements as well as the difference between voiced and unvoiced sounds identifiable in waveform patterns.

On the whole, the gains made by the 12 subject participants may be considered in line with the findings of numerous previous investigations concerning the same issues. Similarly, for example, to Le and Brook (2011), Hamlaoui and Bengraït (2016) and Larassati et al. (2022), this small-scale study provides evidence on the benefits that speech visualization technology, and Praat in particular, can offer to learners in order to improve their intonation of different types of English questions. Furthermore, the promising results, obtained in spite of the sample’s particular age group and school grade, seem to confirm what claimed by Weltens and De Bot (1984) in their groundbreaking study on visualization of intonation, namely that “the feedback system is a useful aid for teaching intonation to a relatively wide range of learners”, since also in the case in question, students’ proficiency level and age (two of the researcher’s major concerns) “did not interfere with the effectiveness of the audio-visual feedback of pitch contours” (p.162).

As a final point, it has to be highlighted that the data interpretation proposed may be more properly regarded as speculative, due to the many limitations characterizing the research. Among these, the small sample size and the particular time constraints were shown to represent two major drawbacks. While the first inevitably affected the study’s statistical power (O’Brien et al., 2018), the second allowed the author to provide only very limited training in the software use, an aspect considered central in most studies of this kind (see for e.g., Lord, 2005; Osatananda & Thinchana, 2021). It is reasonable to assume that, with more time available, the results of this research could have been strengthened; if on one hand, learners’ curiosity, and hence participation, would have

increased, on the other hand many technical inconveniences encountered with audio recordings, thus with the quality of pitch contour displays, could have been avoided. Ultimately, as a further consequence of time restrictions, the gathered data, while providing insights into the short-term effects of audio-visual feedback, give no indication as to whether improvement could be sustained in the long run. Again, presenting the same activity over a longer time frame could have allowed the creation of multiple pre- and post-tests useful for a better assessment of students' progress.

4.3 Research Question 3

The third research question investigated students' response to the proposed pronunciation activity and examined whether the perception of Praat could be such to lead to a possible integration of the software in the scope of pronunciation teaching.

The hypothesis for which students would respond positively to the proposed exercises and would be motivated, due to the interest aroused by the new technological tool, to adopt Praat for further pronunciation activities has been *partially* confirmed.

As illustrated by questionnaire findings, the overall response to the project seems to be very positive. Given that this judgment includes both the presentation and the final reading-of-dialogue sessions as well, what emerges is that addressing cumbersome pronunciation topics, such as suprasegmental phonology, may be deemed interesting and beneficial even by lower-level upper-secondary language classrooms. Supporting this, when asked for their opinion on the usefulness of tackling more often English rhythm and intonation during class (item 16), respondents showed great approval. As underlined by Derwing (2018), although "sometimes teachers think that pronunciation is too difficult to explain or introduce with beginner level students ... it's never too early" (p.15).

On the other hand, the evaluation of speech analysis software Praat can be said to be more varied. From the analysis of students' ratings, it appears that, while the instructional value of the software visual feedback system was acknowledged, most of the problems are of technical nature. Perhaps unsurprisingly then, the step-by-step tutorials proved to be very appreciated by respondents (item 13).

Based on the evidence, it could be argued that the adaptation of the visual feedback paradigm (VFP) to the specific classroom setting resulted to be essential to the achievement of the study's, though limited, successful outcomes; a strictly student-

centered approach such as that implemented by Olson (2014a), aiming at fostering above all participants' self-monitoring and self-assessment by means of a guided inductive method, would not be conceivable in this particular case. The fact of dealing with non-trained young learners, both in terms of English phonology and speech visualization technology, made it necessary for the instructor (i.e., the researcher) to maintain a central role in guiding the students throughout the activity; consider, for example, the explicit phonological explanations provided in the presentation as well as the insights on Praat functioning during the in-class analysis session.

However, despite these efforts, only a very limited number of students completed the entire task and, among them, only half decided to do it at home, following the instructions autonomously. Besides the already mentioned time issues, such relatively low participation could be due not so much to learners' inability to understand Praat basic functions as to its unappealing interface and use mode (i.e., being downloadable on PC only). Had there been a mobile app version of the software, possibly with a more game-like appearance, a greater participation would have most likely been observed (Fouz-González, 2015).

To conclude, it may be stated that questionnaire responses reveal a certain degree of interest deriving from Praat use. Nonetheless, further adaptations of the experimental design seem to be necessary in order to make the software more appealing for younger learners. As pointed out by Olson (2014a), "the creativity of language instructors" is a key aspect to "incorporating speech analysis technology into the language classroom at every level, especially at the beginning levels of instruction" (p.59–60).

4.4 Summary and Conclusions

Communicating in a foreign language is one of the most common situations in which individuals, often misunderstood by native speakers, may start wondering about their voice nature. In the attempt to reproduce difficult target sounds, L2 speakers might realize that the act of speech involves much more than air flowing out of their mouth. L2 pronunciation instruction deals with the multifaceted nature of human voice, in that the teaching of a language segmentals and suprasegmentals must necessarily refer to the articulatory and phonatory aspects of speech.

If traditionally, L2 pronunciation teaching approaches (e.g., audio-lingual) aimed at the nativelikeness of L2 speakers by means of a systematic practice of sound

discrimination, articulation and memorization, more recent communicative methods, influenced by theories, such as the critical period hypothesis, recognized the importance of intelligible pronunciation over perfect accuracy in order to foster communication and gradually highlighted the relevance of suprasegmental aspects in L2 perception and production.

This new emphasis on a sometimes neglected component of language teaching, (especially with reference to prosodic features) found some L2 instructors unprepared and didactic materials inadequate in terms of contents. However, recent technological advances have given a strong boost to the development of CAPT systems, which are now able to compensate for the lack of authentic input (particularly in FL contexts) and constitute, in many cases, a sort of extension of teacher feedback.

Among these systems, speech visualization technology has proven to be a helpful tool in the teaching of prosody (as well as phonemes), in that functions such as waveform and pitch contour displays provide a form of ongoing feedback (though not of a corrective type) on numerous speech features, such as duration, pauses, accents and intonation. Despite being initially designed for acoustic research purposes, speech analysis software, and Praat in particular, have been introduced, due to their versatility, first at more advanced levels of tertiary education and then gradually at lower levels of instruction. Thus, this research sought to understand whether, the implementation of analysis software Praat could be effective even among young teenage learners, in improving their English prosodic abilities.

In this regard, the initial general assumption of this work, according to which students attending a Venetian high school would benefit from Praat use in enhancing their English pronunciation skills, can be considered confirmed.

This favourable outcome points at what previous studies have demonstrated, namely that, by means of ad hoc implementation methods, speech visualization technology may become a valid and versatile CAPT tool for both teachers and students at any level, including beginner-level teenage EFL learners.

In addition to the adaptations applied to the VFP in order to meet the needs of the specific classroom setting, it could be argued that pupils' enthusiasm has played a central role in the experiment's overall success. As emerged from the preliminary, exploratory phase of the research, there appears to be a strong contrast between today's

FL learners and past generations as to the degree of curiosity towards English pronunciation issues as well as to the importance given to *how they sound like* while speaking in the FL; the general tendency is thus that of wanting to improve their own speech skills or at least to not lag behind.

Consequently, the explicit introductory explanations on phonological differences between Italian and English (with specific focus on rhythm and intonation) and their articulatory as well as phonatory implications were very appreciated by both classes, to the point that the researcher was eventually asked by the subject coordinator to hold the same presentation in an extra second year class. Large participation was also detected during the second meeting, when pupils, working in pairs, did not seem to be intimidated by Praat voice analysis, but rather showed interest in finding out through the software display about intonation differences between their own productions and the native speaker models.

Similarly, the twelve most daring students, that is the ones completing all tasks, not only were they helped by audio-visual feedback to become more aware of pitch patterns variations, but, after practicing with Praat, they showed in most cases to be able to improve their pronunciation of English questions. Besides some technical issues, these learners managed to use the software basic functions successfully and gave the activity an overall positive evaluation.

As a final comment, far from attempting to present Praat as the ideal aid to L2/FL pronunciation practice, this investigation emphasized the potential of speech analysis technology in allowing learners, by means of acoustic visualization, to conceptualize their voice and hence to ask themselves questions about its mechanisms. This, in turn, is expected to favour FL pronunciation enhancement, since, once the link between the acoustic patterns and the fragments of sound uttered is understood, speakers should be able to *tune their vocal instrument* to the target sounds. Moreover, the pilot data provided illustrated how, with the necessary precautions, Praat audio-visual feedback could represent a useful aid to pronunciation self-analysis even for upper-secondary beginner students of English, especially concerning prosody, this way filling, though only partly, the lack of information about possible pedagogical applications of speech visualization technology in the scope of L2/FL instruction at the upper-secondary level of education.

At this stage, further research is needed to confirm these findings. First of all, longitudinal studies should be conducted, in order to determine whether and how long improvement fostered by Praat may last with younger learners. Second, given that several unsolved issues regarding the experimental design remain to be addressed, more structured approaches are suggested, in which, for example, a control group is paired up with an experimental group and in which, in addition to questions in isolation, emphasis is placed on the use of Praat with more authentic discourse; in this respect, as shown by Osatananda and Thinchon (2021), one way would be to let learners practicing on voice models chosen on their own from Internet sources. Lastly, similar pedagogical implementations of Praat should be conducted on a wider variety of languages to determine whether the software could be equally effective as with English.

REFERENCES

- Behr, N. S. (2022). English diphthong characteristics produced by Thai EFL learners: Individual practice using Praat. *Computer Assisted Language Learning Electronic Journal* 23(1), 401–424. <http://callej.org/journal/23-1/Behr2022.pdf>
- Boersma, P., & Heuven, V. V. (2001). Speak and unSpeak with PRAAT. *Glott International* 5(9), 341–347.
https://www.fon.hum.uva.nl/paul/papers/speakUnspeakPraat_glot2001.pdf
- Boersma, P., & Weenink, D. (2023). Praat: doing phonetics by computer. Version 6.3.12 [Computer software]. Retrieved from: <https://www.fon.hum.uva.nl/praat/>
- Brett, D. (2004). Computer generated feedback on vowel production by learners of English as a second language. *ReCALL*, 16(1), 103–113.
<https://doi.org/10.1017/S0958344004000813>
- Brown, A. (2015). Syllable structure. In M. Reed, & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 85–105). Wiley Blackwell.
<https://doi.org/10.1002/9781118346952.ch5>
- Busà, M. G. (1995). *L'inglese degli italiani. L'acquisizione delle vocali*. Unipress.
- Busà, M. G. (1999). The Italian accent in English. Results from an experimental study and implications for language teaching. In G. Azzaro, & M. Ulrych (Eds.), *Transiti letterari e culturali. Volume II* (pp. 57–73). EUT Edizioni Università di Trieste.
<https://www.openstarts.units.it/server/api/core/bitstreams/d5b55c0b-b43c-4575-aec0-e9d539714638/content>
- Busà, M. G. (2008a). New perspectives in teaching pronunciation. In A. Baldry, M. Pavesi, C. Taylor Toresello, & C. Taylor (Eds.), *From didactas to ecolingua: An ongoing research project on translation and corpus linguistics* (pp. 165–182). EUT Edizioni Università di Trieste.
<https://www.openstarts.units.it/server/api/core/bitstreams/0f4106df-52f7-4ad7-bcf7-0bd0a440000c/content>
- Busà, M. G. (2008b). Teaching prosody to Italian learners of English: Working towards a new approach. In C. Taylor (Ed.), *Ecolingua. The role of E-corpora in translation and language learning* (pp. 113–126). EUT Edizioni Università di Trieste.

<https://www.openstarts.units.it/server/api/core/bitstreams/0bc189c4-ccaa-4b41-ae6c-446446951e97/content>

- Busà, M. G. (2021). Cenerentola entra a palazzo: Il nuovo ruolo della pronuncia nell'insegnamento linguistico. *ELLE 10*(3), 435–455.
<https://doi.org/10.30687/ELLE/2280-6792/2021/03/005>
- Council of Europe (2018). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume*. Retrieved from:
<https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>
- Cutler, A. (2015). Lexical stress in English pronunciation. In M. Reed, & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 106–124). Wiley Blackwell.
<https://doi.org/10.1002/9781118346952.ch6>
- De Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech 26*(4), 331–350.
<https://doi.org/10.1177/002383098302600402>
- De Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly 16*(1), 71–77. <https://doi.org/10.2307/3586564>
- Denes, P. B., & Pinson, E. N. (2015). *The speech chain: The physics and biology of spoken language* (2nd ed.). Waveland Press, Inc.
- Derwing, T. M. (2010). Utopian goals for pronunciation teaching. In J. Levis, & K. LeVelle (Eds.), *Proceedings from the 1st conference of pronunciation in second language learning and teaching* (pp. 24–37). Iowa State University Digital Press
<https://www.iastatedigitalpress.com/psllt/article/15147/galley/13776/view/>
- Derwing, T. M. (2018). Putting an accent on the positive: New directions for L2 pronunciation research and instruction. *ISAPh 2018 International Symposium on Applied Phonetics*, Aizuwakamatsu, Japan, 19-21 September 2018 (pp. 12–18).
https://isca-speech.org/archive_v0/ISAPh_2018/pdfs/03.pdf
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly 39*(3), 379–397.
<https://doi.org/10.2307/3588486>
- Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favour of a broad

- framework for pronunciation instruction. *Language Learning* 48(3), 393–410.
<https://doi.org/10.1111/0023-8333.00047>
- Deterding, D. (2015). Segmentals. In M. Reed, & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 69–84). Wiley Blackwell.
<https://doi.org/10.1002/9781118346952.ch4>
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford Applied Linguistics.
- Fouz-González, J. (2015). Trends and directions in computer-assisted pronunciation training. In J. A. Mompean, & J. Fouz-González (Eds.), *Investigating English Pronunciation: Trends and Directions* (pp. 314–342). Palgrave Macmillan.
https://doi.org/10.1057/9781137509437_14
- Gorjian, B., & Arvand, E. (2022). Computer-mediated pronunciation technique in developing high school language learners: The case of Praat. *Journal of Global Research in Education and Social Science* 16(3), 1–9.
<https://doi.org/10.56557/jogress/2022/v16i37861>
- Graffi, G., & Scalise, S. (2013). *Le lingue e il linguaggio: Introduzione alla linguistica* (terza edizione). Il Mulino.
- Hamlaoui, N., & Bengrait, N. (2016). Using BetterAccent Tutor and Praat for learning English intonation. *Arab World English Journal AWEJ, Special Issue on CALL*(3), 99–112. <https://doi.org/10.2139/ssrn.2822981>
- Hewings, M. (2007). *English pronunciation in use: Advanced. Book with answers, 5 audio CDs and CD-ROM*. Cambridge University Press.
- Hincks, R. (2015). Technology and learning pronunciation. In M. Reed, & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 505–519). Wiley Blackwell.
<https://doi.org/10.1002/9781118346952.ch28>
- Larassati, A., Setyaningsih, N., Suryaningtyas, V. W., & Cahyono, S. P. (2022). Using Praat for EFL English pronunciation class: Defining the errors of question tags intonation. *Language Circle: Journal of Language and Literature* 16(2), 245–254.
<https://doi.org/10.15294/lc.v16i2.34393>
- Laver, J. (1994). *Principles of phonetics*. Cambridge University Press.

- Le, H. T., & Brook, J. (2011). Using Praat to teach intonation to ESL students. *TESOL Working Paper Series*, 9(1-2), 2–15. https://www.hpu.edu/research-publications/tesol-working-papers/2011/new-with-metadata/9_1-2_le-brook.pdf
- Lepage, A., & Busà, M. G. (2014). Intelligibility of English L2: The effects of incorrect word stress placement and incorrect vowel reduction in the speech of French and Italian learners of English. In *Proceedings of the international symposium on the acquisition of second language speech* (pp. 387–400). Concordia Working Papers in Applied Linguistics.
http://doe.concordia.ca/copal/documents/27_Lepage_GraziaBusa_Vol5.pdf
- Levin, T. C., & Edgerton, M. E. (1999). The throat singers of Tuva. *Scientific American* 281(3), 80–87. <https://doi.org/10.1038/scientificamerican0999-80>
- Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly* 33(1), 37–63. <https://doi.org/10.2307/3588190>
- Levis, J. M. (2019). Cinderella no more: Leaving victimhood behind. *Speak Out* 60, 7–14. https://www.academia.edu/38362208/Cinderella_No_More_pdf
- Levis, J. M., & Wichmann, A. (2015). English intonation – Form and meaning. In M. Reed, & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 139–155). Wiley Blackwell. <https://doi.org/10.1002/9781118346952.ch8>
- Lord, G. (2005). (How) Can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, 88(3), 557–567.
<https://doi.org/10.2307/20063159>
- Low, E. L. (2015). The rhythmic patterning of English(es): Implications for pronunciation teaching. In M. Reed, & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 125–138). Wiley Blackwell.
<https://doi.org/10.1002/9781118346952.ch7>
- Macdonald, S. (2002). Pronunciation – Views and practices of reluctant teachers. *Prospect* 17(3), 3–18.
http://www.ameprc.mq.edu.au/__data/assets/pdf_file/0012/230061/17_3_1_MacDonald.pdf
- Mihás, P., & Odum Institute (2019). *Learn to use an exploratory sequential mixed*

- method design for instrument development*. SAGE Publications, Limited.
<https://parsmodir.com/wp-content/uploads/2020/10/exploratory-method.pdf>
- Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. Hansen Edwards, & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 193–218). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.36.10mun>
- Munro, M. J., & Derwing, T. M. (2015). Intelligibility in research and practice: Teaching priorities. In M. Reed, & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 377–396). Wiley Blackwell.
<https://doi.org/10.1002/9781118346952.ch21>
- Murphy, J. M., & Baker, A. A. (2015). History of ESL pronunciation teaching. In M. Reed, & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 36–65). Wiley Blackwell. <https://doi.org/10.1002/9781118346952.ch3>
- Nordquist, R. (2020, June 12). *Suprasegmental definition and examples*. ThoughtCo.
<https://www.thoughtco.com/suprasegmental-speech-1692008>
- O'Brien, M. G., Derwing, T. M., Chucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R. I., Stirk, H., Levis, J. M., Munro, M. J., Foote, J. A. & Levis, G. M. (2018). Directions for the future of technology in pronunciation research and teaching. *Journal of Second Language Pronunciation* 4(2), 182–207.
<https://doi.org/10.1075/jslp.17001.obr>
- Olson, D. J. (2014a). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania*, 97(1), 47–68. <https://doi.org/10.1353/hpn.2014.0030>
- Olson, D. J. (2014b). Benefits of visual feedback on segmental production in the L2 classroom. *Language Learning & Technology* 18(3), 173–192.
<https://doi.org/10.125/44389>
- 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/article/view/253273>
- Oxford English Dictionary (n.d.). *Pronunciation*. Retrieved April 27, 2023, from <https://www.oed.com/search/dictionary/?scope=Entries&q=pronunciation>

- Pokrivcáková, S. (2015). CALL and teaching pronunciation. In *CALL and Foreign Language Education: e-textbook for foreign language teachers* (pp. 29–37).
https://www.academia.edu/19338241/CALL_and_teaching_pronunciation
- Piske, T., Flege, J. E., MacKay, I. R., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica* 59(1), 49–71.
<https://doi.org/10.1159/000056205>
- Revell-Rogerson, P. M. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal* 52(1), 189–205.
<https://doi.org/10.1177/0033688220977406>
- Saito, K. (2007). The influence of explicit phonetic instruction on pronunciation in EFL settings: The case of English vowels and Japanese learners of English. *The Linguistics Journal* 3(3), 16–40.
https://www.researchgate.net/publication/237538878_The_Influence_of_Explicit_Phonetic_Instruction_on_Pronunciation_Teaching_in_EFL_settings_The_Case_of_English_Vowels_and_Japanese_Learners_of_English
- Setter, J., & Jenkins, J. (2005). State-of-the-art review article. *Language Teaching* 38(1), 1–17. <https://doi.org/10.1017/S026144480500251X>
- Titze, I. R. (2008). The human instrument. *Scientific American* 298(1), 94–101.
<https://doi.org/10.1038/scientificamerican0108-94>
- Warren, P. (2013). *Introducing psycholinguistics*. Cambridge University Press.
- Weltens, B. & De Bot, K. (1984). The visualisation of pitch contours: Some aspects of its effectiveness in teaching foreign intonation. *Speech Communication* 3(2), 157–163. [https://doi.org/10.1016/0167-6393\(84\)90037-2](https://doi.org/10.1016/0167-6393(84)90037-2)
- Wheelock, A. (2016). Phonological difficulties encountered by Italian learners of English: An error analysis. *Hawai Pacific University TESOL Working Paper Series 14*, 41–61. <https://www.hpu.edu/research-publications/tesol-working-papers/2016/04AnneWheelock.pdf>
- Wilson, I. (2008). Using Praat and Moodle for teaching segmental and suprasegmental pronunciation. In T. Koyama, J. Noguchi, Y. Yoshinari, & A. Iwasaki (Eds.), *Proceedings of the WorldCALL 2008 conference: CALL bridges the world* (pp. 112–115). The Japan Association for Language Education and Technology (LET).

https://www.academia.edu/4007777/Using_Praat_and_Moodle_for_Teaching_Segmental_and_Suprasegmental_Pronunciation

Zielinski, B. (2015). The segmental/suprasegmental debate. In M. Reed, & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 397–412). Wiley Blackwell.
<https://doi.org/10.1002/9781118346952.ch22>

Appendix A

Interview Questions

- 1) Da quanti anni insegni inglese alle scuole superiori?
E in questo liceo in particolare?
- 2) Al momento, quali classi stai seguendo?
- 3) Al nostro primo incontro hai accennato al fatto che all'interno dell'Istituto, l'inglese è l'unica lingua straniera insegnata, puoi dirmi qualcosa di più in merito?
- 4) Mi hai anche parlato della differenza tra biennio e triennio, per cui la scuola propone al biennio un curriculum uguale per tutti e dal terzo anno in poi gli studenti devono scegliere quale indirizzo intraprendere. Per quanto riguarda la tua materia: come si articola l'insegnamento dell'inglese tra biennio e triennio?
- 5) Rispetto al tema della *pronuncia*, esiste un momento lungo il percorso scolastico in cui a livello curricolare gli si dedica maggiore attenzione?
- 6) Passiamo alla classe 2A (o 2B) che ho conosciuto pochi giorni fa, nella quale, appunto, verrà svolta l'attività didattica sulla pronuncia inglese:
 - come la descriveresti?
 - com'è complessivamente il clima all'interno della classe durante le tue lezioni?
(con te e tra compagni)
- 7) Per quanto riguarda l'interesse e la motivazione nei confronti della tua materia, come si pone la classe? L'inglese è una materia generalmente apprezzata?
- 8) Parlando nello specifico di pronuncia inglese, una delle prime cose che mi hai detto al nostro primo incontro è stata che negli ultimi anni hai notato un generale miglioramento nella pronuncia inglese degli studenti, puoi parlarmi più approfonditamente di questo aspetto? A cosa è dovuto, secondo te?
- 9) Come descriveresti, in generale, il livello di pronuncia degli studenti della 2A (o 2B)?
- 10) Dunque, in un contesto in cui emerge un quadro di generale miglioramento, che importanza assume oggi, secondo te, un insegnamento di tipo esplicito della pronuncia inglese?

- 11) In quale misura viene incoraggiato l'insegnamento della pronuncia a livello curricolare? Vengono indicati obiettivi specifici oppure lo si affronta in modo generico?
- 12) Mediamente quanto tempo dedichi alla pronuncia durante una lezione?
- 13) Quali sono le tecniche per insegnare la pronuncia che utilizzi più di frequente?
- 14) In quale misura e in che modo sei solito correggere gli errori di pronuncia dei tuoi studenti?
- 15) Che cosa mi puoi dire della prosodia, in particolare del ritmo e dell'intonazione? Sono temi presi in considerazione? Pensi che un'attenzione maggiore verso questo tema potrebbe avere effetti positivi sul livello di pronuncia degli studenti?
- 16) Come valuti il libro di testo in dotazione per quanto riguarda gli aspetti legati alla pronuncia?
- 17) Escludendo le risorse multimediali offerte dal libro di testo, quali altre tecnologie vengono utilizzate in classe per approfondire elementi di pronuncia?
- 18) Hai mai avuto occasione di utilizzare nell'ambito dell'insegnamento della pronuncia un software che analizza acusticamente le produzioni orali?
- 19) Ora facciamo un'ipotesi: se venisse stabilito di dedicare maggiore attenzione a livello curricolare al tema della pronuncia sia nel biennio che nel triennio, sia riguardo agli aspetti segmentali che soprasegmentali, ti riterresti sufficientemente preparato nel proporre l'argomento ai tuoi studenti in modo più approfondito? In quel caso, riterresti di disporre di adeguate risorse didattiche?
- 20) C'è qualcos'altro, rispetto ai temi trattati lungo l'intervista, che può essermi sfuggito e che vorresti aggiungere?

Appendix B



ATTIVITÀ SULLA PRONUNCIA INGLESE: CALENDARIO

Hi guys!

Per prima cosa vorrei ringraziarvi per il tempo che dedicherete a questa attività nelle prossime settimane. Mi auguro possiate trovare spunti interessanti tra gli argomenti che andremo ad affrontare. Ho pensato di fornirvi un calendario per capire meglio come si articolerà l'intera attività. A presto! *Alice*

CALENDARIO		
	IN CLASSE	A CASA
01.03.'23	<p>Presentazione Attività:</p> <p>Con l'aiuto di un <i>PowerPoint</i> parleremo di:</p> <ul style="list-style-type: none"> - perché è difficile imparare la pronuncia di una lingua straniera - quali sono gli elementi principali che compongono la pronuncia di una lingua - l'utilizzo del software <i>Praat</i> per analizzare la nostra pronuncia inglese. 	
da 01.03.'23 a 06.03.'23		<p>Registrazioni File Audio con software <i>Praat</i></p> <p>(upload su <i>Google Classroom</i> entro 06.03.'23)</p> <p>Troverete su <i>Google Classroom</i>:</p> <ul style="list-style-type: none"> - il <i>PowerPoint</i> della presentazione - due tutorial per scaricare e utilizzare <i>Praat</i> - le fotocopie che vi sono state consegnate in classe, in caso di vostra assenza o se doveste perderle, che servono da supporto ai due tutorial - il file con le parole e le frasi da registrare con <i>Praat</i> <p>N.B. La richiesta di caricare i vostri file audio due giorni prima del secondo incontro è indispensabile per poter organizzare il materiale per la prossima lezione.</p>

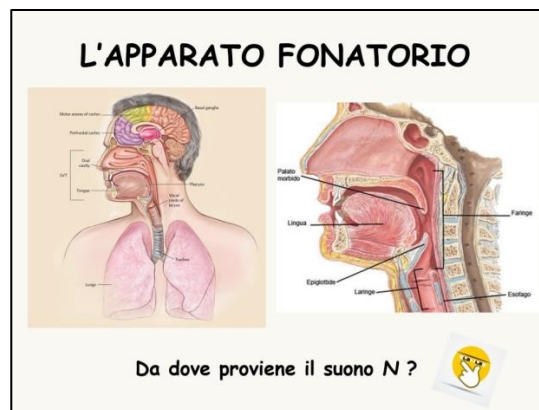
<p>08.03.'23</p>	<p>Analisi RegISTRAZIONI ed Esercizi Pratici</p> <p>Vi porterò le fotocopie delle rappresentazioni grafiche (ottenute con <i>Praat</i>) delle vostre registrazioni. Le analizzeremo insieme e faremo un confronto con le rappresentazioni grafiche ottenute da parlanti madrelingua inglese.</p>	
<p>da 08.03.'23 a 13.03.'23</p>		<p>Registrazioni File Audio con software <i>Praat</i> (upload su <i>Google Classroom</i> entro 13.03.'23)</p> <p>Questa volta l'esercizio consiste nel registrare nuovamente le stesse parole e frasi, ma avendo a disposizione un modello di riferimento, quello dei parlanti madrelingua inglese.</p> <p>Fate sempre attenzione alla data entro la quale caricare i file!</p>
<p>15.03.'23</p>	<p>Confronto conclusivo: dubbi/domande</p> <p>Vi porterò le fotocopie delle nuove rappresentazioni grafiche tratte dalle vostre ultime registrazioni. Sarà così possibile vedere se e in che modo si differenziano dalle prime registrazioni.</p>	
<p>da 15.03.'23 a 20.03.'23</p>		<p>Questionario di "gradimento" (online)</p> <p>Si tratta di un questionario online in cui vi chiederò di dare dei voti all'attività svolta, per capire se è stata facile o difficile, coinvolgente o noiosa ecc.</p>

Appendix C

Images drawn from the PowerPoint slideshow

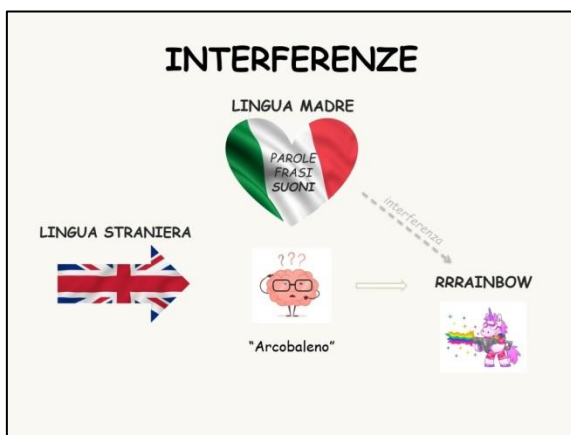


A Xhosa Language Speaker

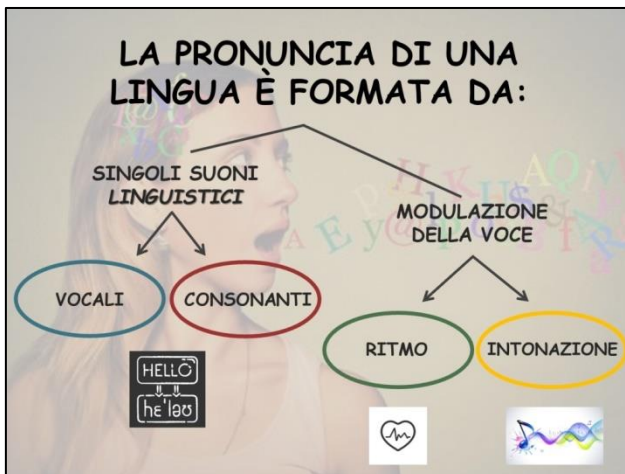


The Human Vocal Apparatus.

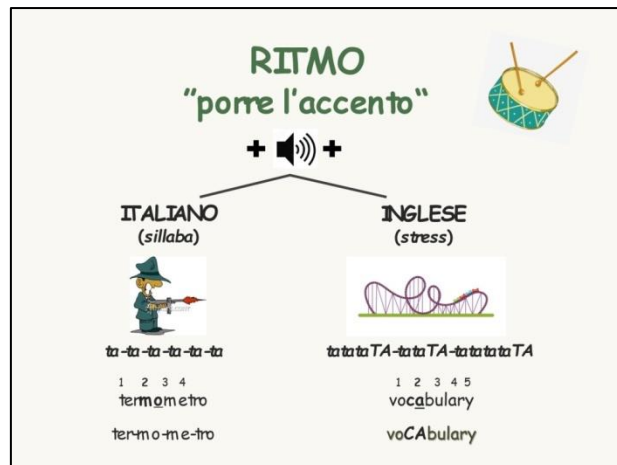
Where does the sound [n] come from? (Try to pronounce [n] by holding your nose)



Interference from L1 to L2



Pronunciation's Main Components



Italian Rhythm vs. English Rhythm

INTONAZIONE
"frequenza della voce"

Hz FREQUENZA: quante volte le **CORDE VOCALI** si aprono e si chiudono in **1 SECONDO**

n.b. la frequenza si misura in **Hertz (Hz)**

Es. Vocale di 200HZ = le corde vocali vibrano 200 volte in 1 secondo!!!

Intonation and Vocal Folds' Vibration

Appendix D

Item list to be recorded (words and questions)

PAROLE & FRASI DA REGISTRARE

(8 parole + 8 frasi = 16 file audio da registrare)



Contesto generale:



ti trovi in Irlanda per un soggiorno-studio e risiedi presso uno studentato assieme ad altri studenti provenienti da tutto il mondo. Durante la settimana frequenti diversi corsi presso il college della città.

1a) *China*

Contesto: ti trovi all'interno del college e stai facendo amicizia con i tuoi compagni di corso. C'è un ragazzo/a cinese che ti sta particolarmente simpatico/a. Per conoscerlo/a meglio gli/le chiedi:

1b) *What part of China do you come from?*

2a) *laundry*

Contesto: ti trovi allo studentato in cui risiedi e hai bisogno di fare una lavatrice perché non hai più un vestito pulito! Hai visto che c'è una lavanderia comune al piano terra, ma non sai bene come funziona. A questo punto ti rivolgi ad un ragazzo che abita lì già da un po' di tempo e gli chiedi:

2b) *Do I need to book the laundry room?*

3a) *Northern*

Contesto: un amico ti chiede se nel weekend vuoi andare a visitare Belfast, in Irlanda del Nord, ma ti viene un sospetto; se devi varcare il confine forse hai bisogno del passaporto, ma con te hai solo la carta di identità. Cercando informazioni in merito, gli chiedi:

3b) *Northern Ireland is not part of Europe, is it?*

4a) *Portuguese*

Contesto: nei corridoi dello studentato senti due ragazze parlare una lingua che attira la tua attenzione. Ti avvicini e finalmente riconosci una delle tue lingue preferite; per averne conferma, prendi coraggio e chiedi loro:

4b) *You're speaking Portuguese, aren't you?*

5a) courses

Contesto: hai appena finito una lezione e alle macchinette incontri un ragazzo/a straniero/a che come te è lì da poco. Iniziate a parlare di scuola e vuoi saperne di più sul suo percorso di studi all'interno del college, per questo gli/le chiedi:

5b) How many courses are you taking?**6a) green**

Contesto: è un giorno di festa nazionale in Irlanda, è il 17 marzo, il giorno di San Patrizio. Per festeggiare decidete di vestirvi e truccarvi di verde, proprio come fanno gli irlandesi. Uno dei tuoi amici però non vuole né vestirsi, né truccarsi e cerchi di capire quale sia il motivo. Alla fine sei quasi certo che il problema sia proprio il colore tipico della festa e gli chiedi:

6b) You don't like green, do you?**7a) parties**

Contesto: allo studentato hai conosciuto un sacco di ragazzi e ragazze da tutto il mondo e vorreste organizzare una festa con della musica, ma non siete sicuri se sia permesso dal regolamento. Nel dubbio, decidi di andare dal responsabile dello stabile e gli chiedi:

7b) Are we allowed to have parties?**8a) thirty**

Contesto: un gruppo di voi è andato a visitare un tipico castello irlandese che dista circa una ventina di chilometri dalla città in cui risiedete. Si è fatto tardi e dovete rientrare, ma non siete sicuri che l'autobus passi con la stessa frequenza del mattino. Per avere informazioni in merito, ti rivolgi a una signora del posto che come voi sta aspettando alla fermata e le chiedi:

8b) The bus comes every thirty minutes, doesn't it?

Appendix E

Written Guidelines: How to download *Praat*

COME SCARICARE PRAAT



(Praat Icon)

1. Digita www.praat.org , verrai reindirizzato al sito <https://www.fon.hum.uva.nl/praat/>
2. In alto a sinistra, sotto **Download Praat**, scegli la versione per il tuo computer (Mac, Windows, Linux, ecc.)
3. Se scegli **Windows**, appariranno due versioni (64 e 32-bit edition) scegli la **64-bit edition** cliccando su *praat6306_win64.zip*
4. Attendi che il download sia completato. Vai nella cartella *Downloads* e apri lo *zip file*.
5. Seleziona **Praat.exe** e trascinalo direttamente sul *Desktop* del tuo computer.
6. Apparirà l'icona di *Praat* formata da una bocca e un orecchio. Ora sei pronto per iniziare!

N.B. Se scegli **Macintosh**, apparirà un'unica versione. Clicca su *praat6306_mac.dmg*

Accetta il download -> Clicca l'icona di Download in alto a destra  e apri il file.

Si aprirà una finestra con l'icona di *Praat*. Trascina direttamente l'icona sul Desktop.

Appendix F


Written Guidelines: how to record, visualize and read your voice with *Praat*, and how to save your audio files

ESERCIZIO PER CASA

- come registrare la propria voce con *Praat*
- come visualizzare e leggere la propria voce con *Praat*
- come salvare i file audio

Come registrare la propria voce con *Praat*



1. Sul desktop del tuo computer clicca sull'icona di *Praat* , si apriranno due finestre *Praat Objects* e *Praat Picture*. Chiudi pure *Praat Picture* perché per il nostro esercizio useremo solo *Praat Objects*.
2. Clicca su **New** -> **Record Mono Sound**. Si aprirà la finestra **Sound Recorder**.
3. In basso a destra vai nello spazio preceduto da **Name**. Cancella la scritta "untitled" e dai un nome alla tua registrazione. Come nome utilizza **il numero e la lettera** con cui sono indicate le parole e le frasi nella *lista Parole & Frasi da Registrare*, seguite da un trattino _ e il vostro **cognome** (per es. 1a_Percivalli) È importante perché così saprò a chi appartiene la voce della registrazione!
4. Clicca su **Record** e registra la parola o la frase, quando hai terminato clicca **Stop**.
Puoi subito riascoltare l'audio cliccando su **Play**.
Se non sei soddisfatto, puoi registrare un nuovo audio sopra quello precedente cliccando di nuovo Record -> (registra parola o frase) -> Stop. (Se senti un eccesso di brusio nell'audio puoi provare ad abbassare il microfono del computer, ma non troppo!)

Ripeti questa operazione fino a quando sarai soddisfatto della tua registrazione, dopodiché clicca su **Save to list & Close**. In questa fase della nostra attività registra pure le parole e le frasi nel modo più spontaneo possibile.☺

5. Il file audio apparirà, con il nome che gli hai dato, nello spazio **Objects**, all'interno della finestra *Praat Objects*. Ora sei pronto per visualizzare la tua voce!

Come visualizzare e leggere la propria voce con *Praat*

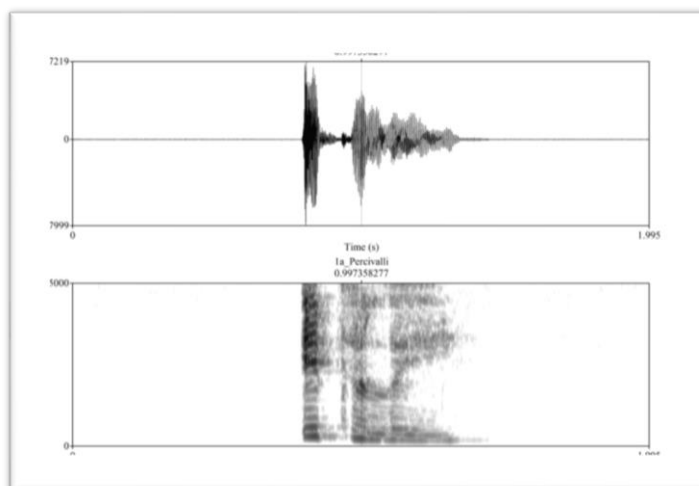


Prima di iniziare, fai attenzione!

- *Praat* è molto sensibile e registra ogni piccolo mutamento nella qualità della voce. Per questo motivo, non otterrai mai una rappresentazione grafica della stessa parola o frase identica a quella precedente.
- La stessa parola o frase potrebbe avere una rappresentazione grafica un po' diversa se pronunciata da una voce più maschile o più femminile.
- Nelle spiegazioni che seguono prenderò come esempio la registrazione della mia voce che pronuncia la parola **Italy**, la stessa che abbiamo visto durante la presentazione in classe.

Dal menù sulla destra di *Praat Objects*, clicca su **View & Edit**.

Appariranno due diverse rappresentazioni grafiche della nostra voce: l'**onda sonora** in alto e lo **spettrogramma** in basso.



N.B. L'**onda sonora** registra le pulsazioni delle particelle d'aria che colpiscono il microfono. Quando emettiamo un suono dalla bocca o dal naso "facciamo pressione" sull'aria circostante. In questo modo mettiamo in moto un meccanismo per cui le particelle d'aria cominciano a vibrare avanti e indietro come delle molle. Grazie a questo movimento il suono (e quindi la *vibrazione*) viene trasportato fino all'orecchio di chi ascolta o fino al microfono, come in questo caso.

Porta l'attenzione all'onda sonora...

- l'*asse orizzontale* dell'onda rappresenta il **tempo**, ovvero quanto dura la registrazione
- l'*asse verticale* rappresenta l'**intensità** della nostra voce. Lo zero, a metà dell'asse verticale, indica l'assenza di suono, il silenzio

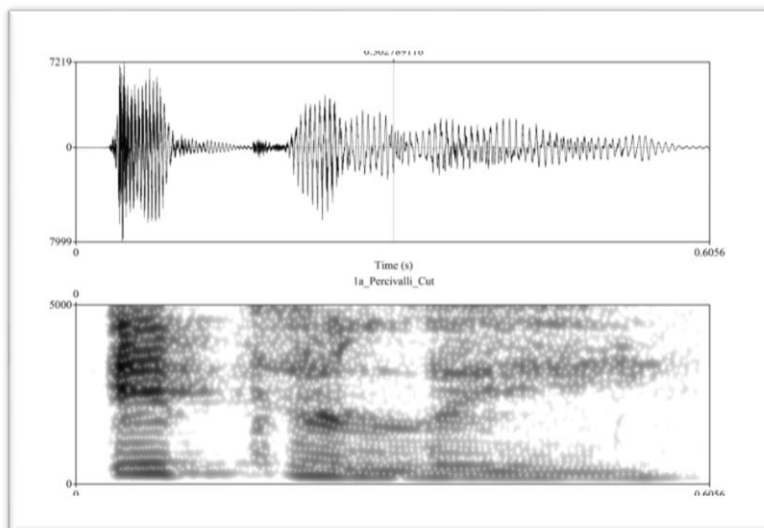
Tra le parole e le frasi che stai registrando, soffermati su una parola e una frase e prova ad eseguire le seguenti operazioni sull'onda sonora corrispondente:

1. Ascolta la voce cliccando su una delle due barre in basso: **Visible Part** o **Total Duration**.
2. Sposta il *cursore*, che inizialmente si trova al centro, e ascolta la prima parte e la seconda parte dell'onda, cliccando sulla barra che si trova sopra a *Visible Part*.
3. Ritaglia le parti di *silenzio* in eccesso, per avere un'onda sonora più chiara. Per farlo, seleziona con il cursore la parte che precede l'onda, clicca su **Edit** in alto a

sinistra e poi su **Cut**. Puoi fare lo stesso con la parte che segue l'onda.

Riascolta la tua voce ad ogni "ritaglio" cliccando su *Visible Part* per assicurarti di non aver cancellato anche la tua voce. Se inavvertitamente hai ritagliato troppo, puoi tornare indietro cliccando di nuovo su *Edit*, in alto a sinistra e su **Undo Cut**.

Ecco il risultato:



4. Porta l'attenzione alla forma dell'onda sonora, in particolare al **CONTORNO**.
Come mai, secondo te, in alcuni punti l'onda quasi scompare e in altri si allarga molto verso l'alto e verso il basso? In quali punti la voce è più intensa?
5. Prova ad individuare, nel caso della singola parola, i diversi suoni linguistici (vocali e consonanti) che compongono la parola oppure, nel caso della frase, ad individuare le diverse parole che compongono la frase. In che modo la forma dell'onda ti aiuta a distinguere i suoni o a distinguere le parole?

Per aiutarti, seleziona le parti dell'onda sonora che ti interessano e ascolta quelle parti cliccando sulla barra che si trova sopra a *Visible Part*. Quella stessa barra ti dice anche quanto dura il frammento di suono.
6. Se vuoi puoi zoomare sul frammento di suono che hai selezionato cliccando su **sel** in basso a sinistra e lo puoi ascoltare cliccando *Visible Part*.

Per tornare alla forma iniziale basta premere **all** sempre in basso a sinistra.

Come salvare i file audio



Quando hai terminato le operazioni sull'onda sonora e sei soddisfatto del tuo file audio, chiudi la finestra di *View & Edit*, seleziona il file che vuoi salvare da *Objects*, clicca su **Save** in alto e seleziona **Save as WAV file**.

Per non perdere i file che poi dovrai caricare su *Google Classroom*, crea una *cartella* dove salvarli tutti. Puoi chiamarla con la data della prossima lezione, ad es.

Praat_08.03.23

Una volta che hai salvato tutti i file in una cartella a parte, se desideri riascoltarli e vederne la rappresentazione grafica con *Praat*, basterà riaprire il software e in ***Praat Objects***, cliccare **Open** in alto a sinistra e **Read from file**.

Si aprirà la finestra **Read Objects from file**, da qui puoi selezionare la cartella e il file che ti interessa. Sempre all'interno della stessa finestra, cliccando **Open** in basso a destra, il file apparirà in *Praat* nella sezione **Objects**. Potrai nuovamente aprirlo, visualizzarlo, ascoltarlo, modificarne la lunghezza ecc. In caso di modifiche, se vuoi mantenere la nuova versione, dovrai salvarne una copia con la stessa procedura usata precedentemente (Save+Save as WAV file).




Ricordati di caricare i file audio

su ***Google Classroom!***



Appendix G

Guess the Intonation

INDOVINA L'INTONAZIONE ↗ ↘ →			
	Your name:	?  (Ipotesi)	 (Conferma)
			
1b) What part of China do you come from?			
2b) Do I need to book the laundry room?			
3b) Northern Ireland is not part of Europe, is it?			
4b) You're speaking Portuguese, aren't you?			
5b) How many courses are you taking?			
6b) You don't like green, do you?			
7b) Are we allowed to have parties?			
8b) The bus comes every thirty minutes, doesn't it?			

Appendix H

Intonation + Context

INTONAZIONE + CONTESTO			
QUESTION TAGS		OPEN QUESTIONS ↘	CLOSED QUESTIONS ↗
Real Questions ↗	Not Real Questions ↘		
Sono domande vere e proprie fatte per chiedere informazioni.	Sono delle finte domande. Chi la fa è abbastanza convinto di quello che sta dicendo e si aspetta una conferma dal suo interlocutore.	Sono domande fatte per ottenere un'informazione specifica. L'interlocutore potrebbe rispondere con una lunga serie di informazioni. - Sono anche dette: Finding Out Questions - Iniziano con una Question Word	Sono domande fatte per accertarsi di qualcosa di cui non si è sicuri. L'interlocutore può rispondere anche solo con un <i>si</i> o con un <i>no</i> . - Sono anche dette: Yes/No Questions oppure Making Sure Questions
Leggi il contesto di ogni domanda prima di rispondere ...			
_ b e _ b	_ b e _ b	_ b e _ b	_ b e _ b

Appendix I

Written Guidelines: how to conduct a comparison between the subject's initial recordings and the native speaker models

ESERCIZIO PER CASA

Dopo aver aperto il nostro audio e quello del parlante inglese, come mostrato nel Tutorial 1, proviamo a fare attenzione ai seguenti aspetti:

1) CONFRONTO DURATA

Quanto dura la mia domanda rispetto a quella del parlante inglese?

Una pronuncia simile a quella di un parlante madrelingua viene data anche da quanto facciamo durare i suoni/le parole della frase che stiamo pronunciando.

Nell'audio definitivo:

proviamo ad avvicinarci al tempo dell'audio del modello inglese.

2) CONFRONTO ONDE SONORE

In che modo si differenzia la nostra onda sonora da quella del parlante inglese?

Da quante parti è formata?

Abbiamo suddiviso le parole della frase come il parlante inglese, oppure abbiamo accorpato delle parole pronunciandole più in fretta, oppure ancora abbiamo aggiunto delle pause?

Nell'audio definitivo:

proviamo a far sì che il numero di parti di cui si compone la nostra onda sonora sia lo stesso di quello del modello inglese.

Che tipo di intensità (volume della voce) abbiamo dato ad ogni parte dell'onda?

Il pezzetto d'onda con maggiore intensità, quindi quello su cui abbiamo fatto cadere l'accento, corrisponde a quello del parlante inglese?

Nell'audio definitivo:

proviamo ad alzare il volume della voce negli stessi punti del modello inglese.

3) CONFRONTO LINEA INTONAZIONE

Qual è l'andamento della nostra intonazione rispetto a quella del parlante inglese?

Nell'audio definitivo:

proviamo a seguire l'andamento dell'intonazione del modello inglese, in particolare nella PARTE FINALE DELLA DOMANDA.

Appendix L

Intonation in Discourse: Dialogue 1

DIALOGO 1				
INTONATION ↗ ↘ →	QUESTION TAGS		OPEN QUESTIONS	CLOSED QUESTIONS
	Real Questions	Not Real Questions		
A: Hi Lucy! You look tired. <i>What did you do last night?</i>				
B: Hi Paul! Yes, I'm a bit tired. I watched the entire first season of Wednesday until 2a.m. It's really good! <i>You've watched it too, haven't you?</i>				
A: Actually, I haven't yet. I've got too much to study these days...you know, the final Spanish exam is getting closer. <i>Do you feel ready for it?</i>				
B: I feel pretty much ready. Come on, it won't be that hard to pass... By the way, <i>we are allowed to bring a dictionary, aren't we?</i>				
A: Yep, the teacher said we're allowed to. Good for us, it would be a disaster without a dictionary.				

Appendix M

Intonation in Discourse: Dialogue 2

DIALOGO 2				
<i>INTONATION</i> ↗ ↘ →	QUESTION TAGS		OPEN QUESTIONS	CLOSED QUESTIONS
	Real Questions	Not Real Questions		
A: Hi Jane! <i>How was school today?</i>				
B: Hi mom...mmm, as usual. <i>Is lunch ready yet?</i>				
A: Almost ready. Please Jane, before you go upstairs, check the mailbox outside.				
B: I don't have time mom, I'm going to call Russel now. <i>You can check yourself, can't you?</i>				
A: Don't talk to me like that. I've been doing the housework all morning, but <i>you don't care do you?</i>				

Appendix N
List of Questionnaire Items

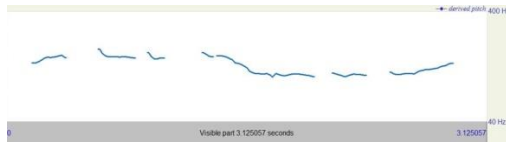
- 1) L'attività sulla pronuncia dell'inglese nel suo complesso è stata utile.
- 2) L'attività sulla pronuncia dell'inglese nel suo complesso mi ha fatto riflettere sulle potenzialità della mia voce.
- 3) L'attività sulla pronuncia dell'inglese nel suo complesso mi ha fatto riflettere sulla mia pronuncia in inglese.
- 4) Approfondire il tema del ritmo e dell'intonazione è stato utile per riflettere sulla mia pronuncia in inglese.
- 5) L'utilizzo del software Praat nel suo complesso è stato utile.
- 6) L'utilizzo del software Praat mi ha aiutato a riflettere sulle caratteristiche della mia voce.
- 7) L'utilizzo del software Praat mi ha aiutato a riflettere sulla mia pronuncia in inglese.
- 8) La funzione di Praat che permette di vedere sullo schermo la mia voce attraverso l'onda sonora e la linea dell'intonazione è stata di aiuto nel migliorare la mia pronuncia in inglese durante l'attività.
- 9) Scaricare il software Praat è stato facile.
- 10) Registrare la mia voce con Praat è stato facile.
- 11) Visualizzare l'onda sonora e la linea dell'intonazione con Praat è stato facile.
- 12) Confrontare attraverso la rappresentazione grafica di Praat la mia pronuncia inglese e quella del parlante madrelingua è stato facile.
- 13) I tutorial con le spiegazioni sull'utilizzo di Praat sono stati utili.
- 14) L'utilizzo del software Praat nel suo complesso è stato facile.
- 15) L'attività sulla pronuncia dell'inglese nel suo complesso mi è piaciuta.
- 16) Affrontare più spesso in classe i temi del ritmo e dell'intonazione inglesi mi aiuterebbe a migliorare la mia pronuncia.

- 17) Mi piacerebbe utilizzare di nuovo Praat durante le lezioni in classe come aiuto per analizzare la pronuncia inglese.
- 18) Mi piacerebbe capire meglio il funzionamento di Praat per poterlo utilizzare in autonomia ogni volta che voglio valutare la mia pronuncia inglese.

Appendix O: Intonation Contours by Student 1



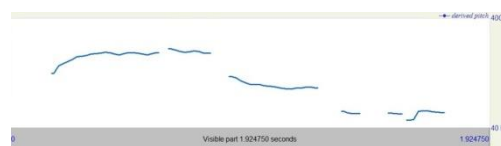
Eng. Model Q.1b_Wh-question



Q.1b_S1_T1



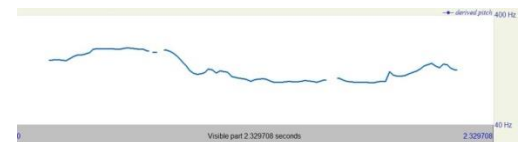
Q.1b_S1_T2



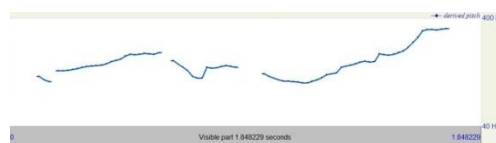
Eng. Model Q.5b_Wh-question



Q.5b_S1_T1



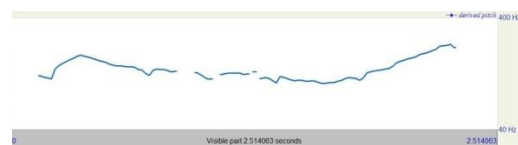
Q.5b_S1_T2



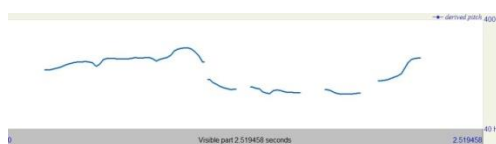
Eng. Model Q.2b_yes/no question



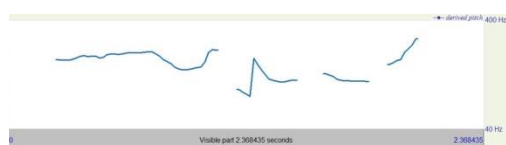
Q.2b_S1_T1



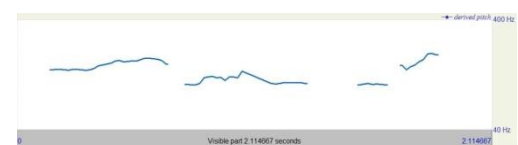
Q.2b_S1_T2



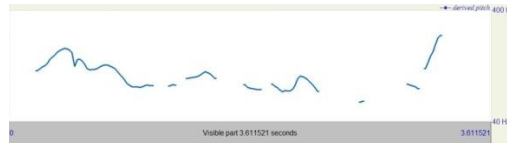
Eng. Model Q.7b_yes/no question



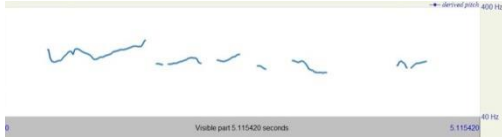
Q.7b_S1_T1



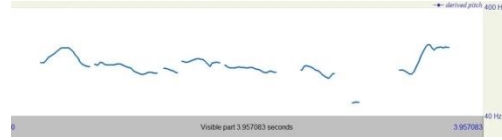
Q.7b_S1_T2



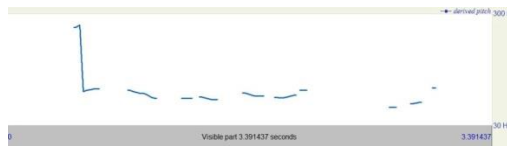
Eng. Model Q.3b_tag question – up tag



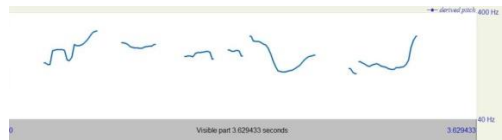
Q.3b_S1_T1



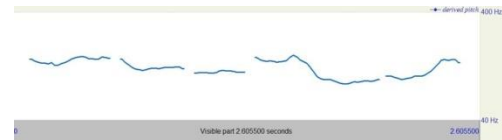
Q.3b_S1_T2



Eng. Model Q.8b_tag question – up tag



Q.8b_S1_T1



Q.8b_S1_T2



Eng. Model Q.4b_tag question – down tag



Q.4b_S1_T1



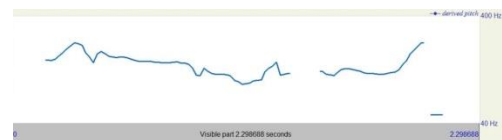
Q.4b_S1_T2



Eng. Model Q.6b_tag question – down tag



Q.6b_S1_T1

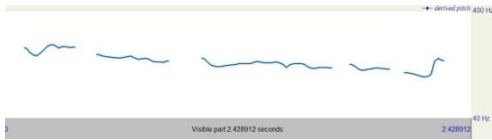


Q.6b_S1_T2

Appendix P: Intonation Contours by Student 2



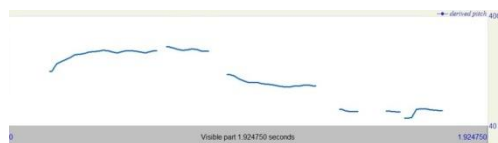
Eng. Model Q.1b_Wh-question



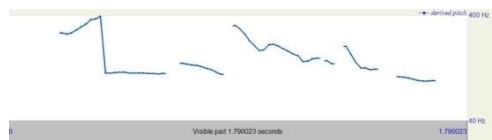
Q.1b_S2_T1



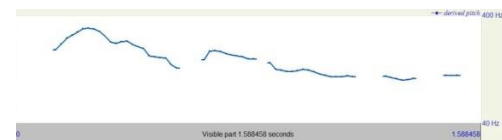
Q.1b_S2_T2



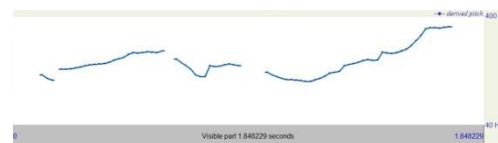
Eng. Model Q.5b_Wh-question



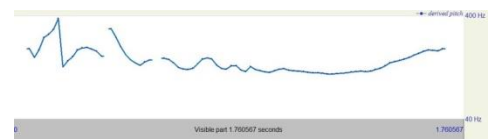
Q.5b_S2_T1



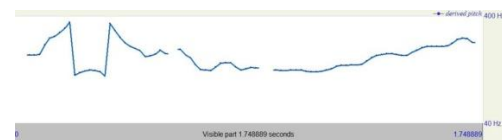
Q.5b_S2_T2



Eng. Model Q.2b_yes/no question



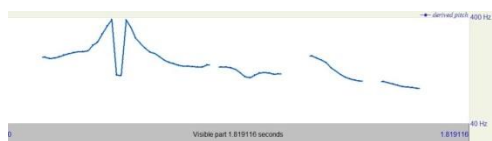
Q.2b_S2_T1



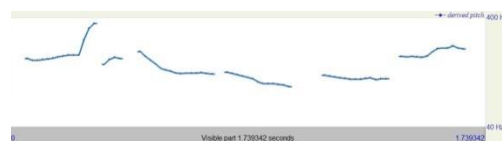
Q.2b_S2_T2



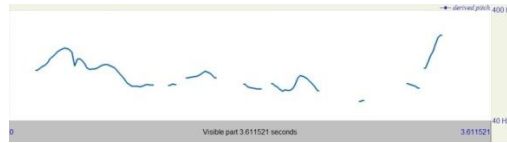
Eng. Model Q.7b_yes/no question



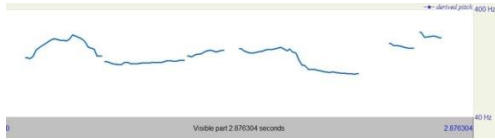
Q.7b_S2_T1



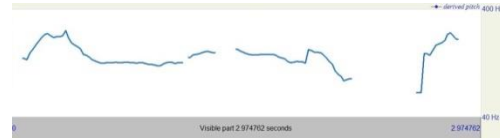
Q.7b_S2_T2



Eng. Model Q.3b_tag question – up tag



Q.3b_S2_T1



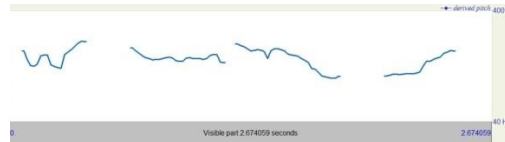
Q.3b_S2_T2



Eng. Model Q.8b_tag question – up tag



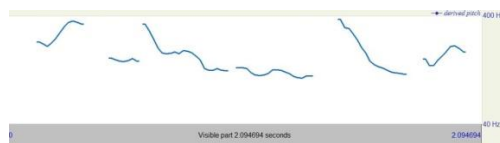
Q.8b_S2_T1



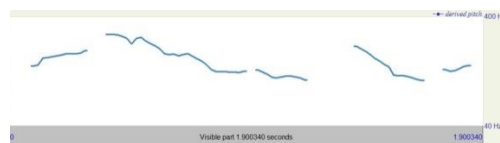
Q.8b_S2_T2



Eng. Model Q.4b_tag question – down tag



Q.4b_S2_T1



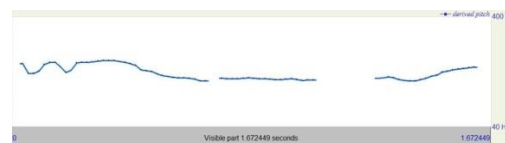
Q.4b_S2_T2



Eng. Model Q.6b_tag question – down tag



Q.6b_S2_T1



Q.6b_S2_T2

Appendix Q: Intonation Contours by Student 3



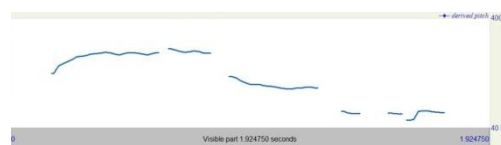
Eng. Model Q.1b_Wh-question



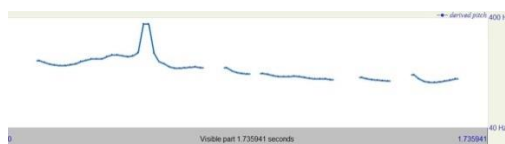
Q.1b_S3_T1



Q.1b_S3_T2



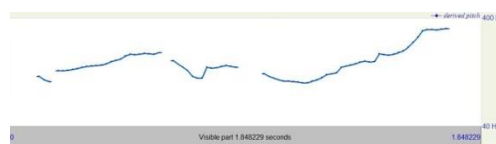
Eng. Model Q.5b_Wh-question



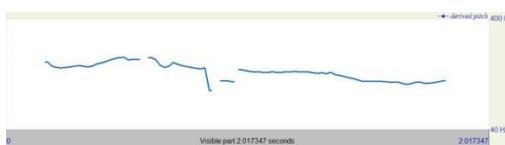
Q.5b_S3_T1



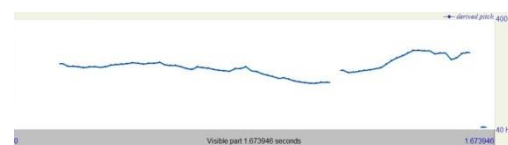
Q.5b_S3_T2



Eng. Model Q.2b_yes/no question



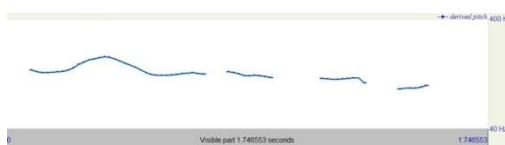
Q.2b_S3_T1



Q.2b_S3_T2



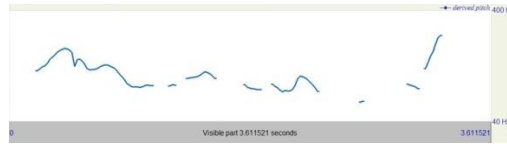
Eng. Model Q.7b_yes/no question



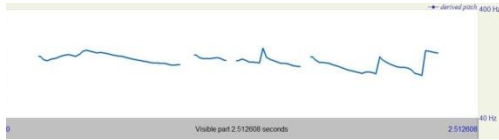
Q.7b_S3_T1



Q.7b_S3_T2



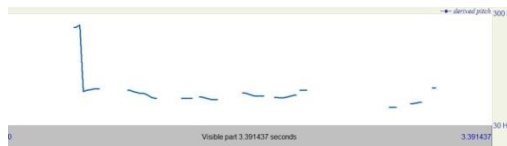
Eng. Model Q.3b_tag question – up tag



Q.3b_S3_T1



Q.3b_S3_T2



Eng. Model Q.8b_tag question – up tag



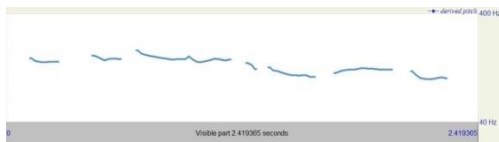
Q.8b_S3_T1



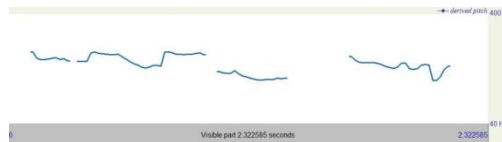
Q.8b_S3_T2



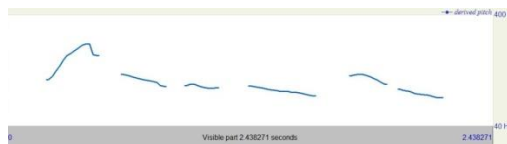
Eng. Model Q.4b_tag question – down tag



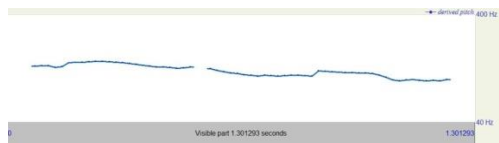
Q.4b_S3_T1



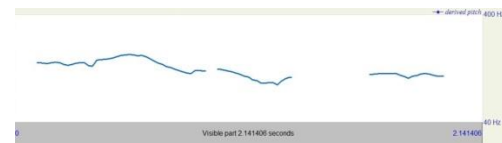
Q.4b_S3_T2



Eng. Model Q.6b_tag question – down tag



Q.6b_S3_T1

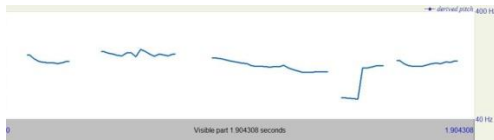


Q.6b_S3_T2

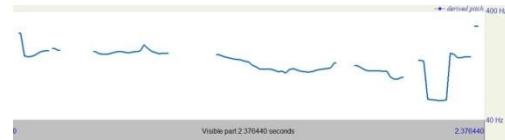
Appendix R: Intonation Contours by Student 4



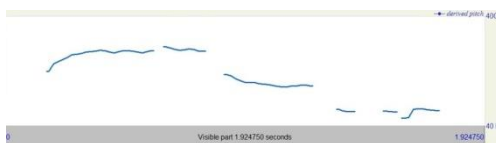
Eng. Model Q.1b_Wh-question



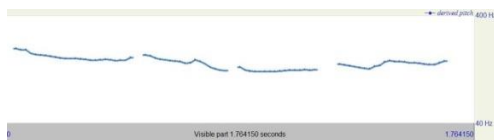
Q.1b_S4_T1



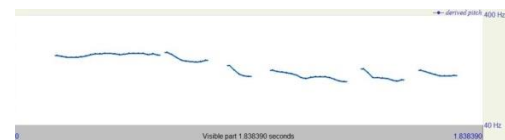
Q.1b_S4_T2



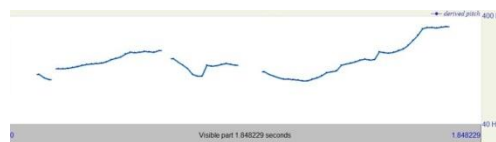
Eng. Model Q.5b_Wh-question



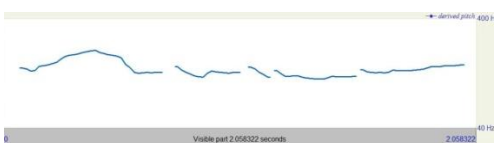
Q.5b_S4_T1



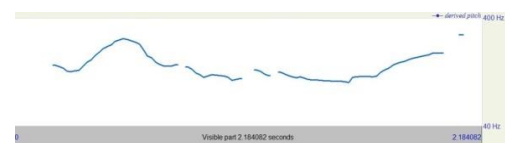
Q.5b_S4_T2



Eng. Model Q.2b_yes/no question



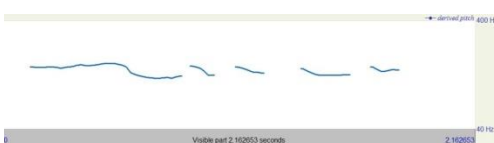
Q.2b_S4_T1



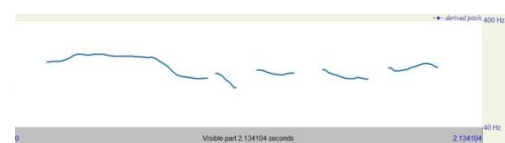
Q.2b_S4_T2



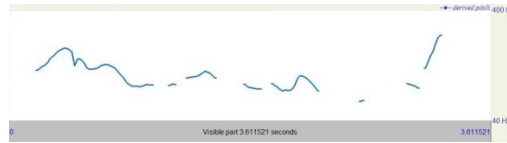
Eng. Model Q.7b_yes/no question



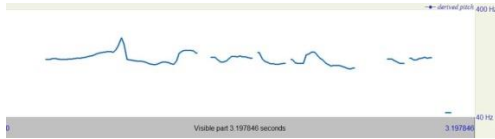
Q.7b_S4_T1



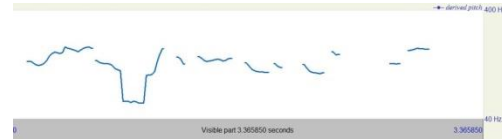
Q.7b_S4_T2



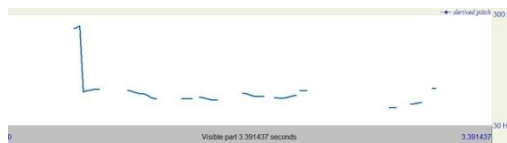
Eng. Model Q.3b_tag question – up tag



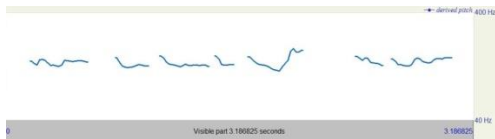
Q.3b_S4_T1



Q.3b_S4_T2



Eng. Model Q.8b_tag question – up tag



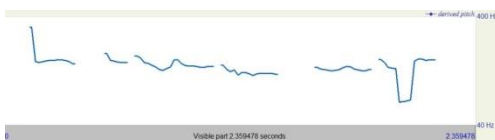
Q.8b_S4_T1



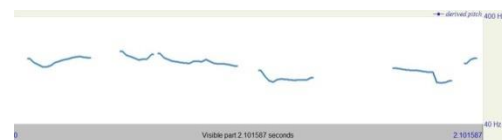
Q.8b_S4_T2



Eng. Model Q.4b_tag question – down tag



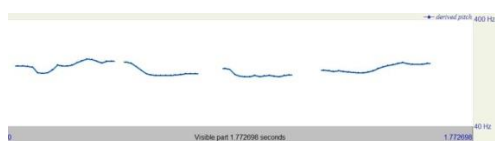
Q.4b_S4_T1



Q.4b_S4_T2



Eng. Model Q.6b_tag question – down tag



Q.6b_S4_T1

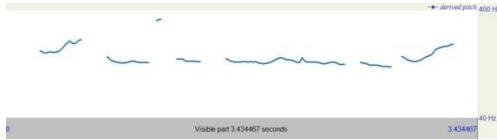


Q.6b_S4_T2

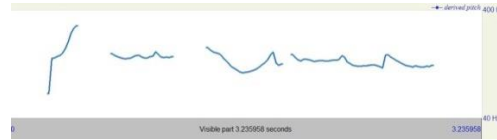
Appendix S: Intonation Contours by Student 5



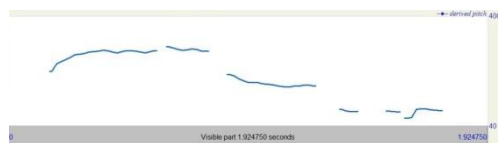
Eng. Model Q.1b_Wh-question



Q.1b_S5_T1



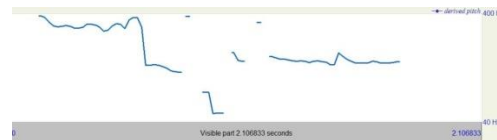
Q.1b_S5_T2



Eng. Model Q.5b_Wh-question



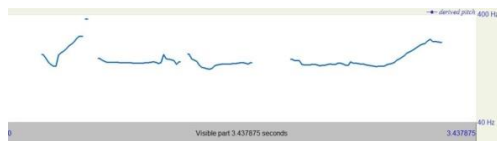
Q.5b_S5_T1



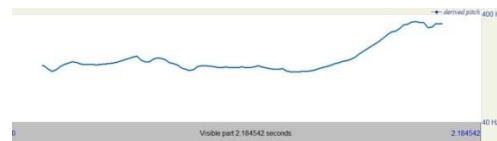
Q.5b_S5_T2



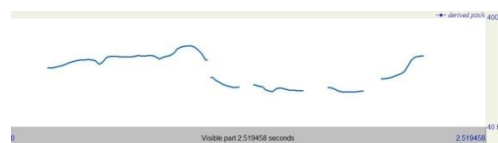
Eng. Model Q.2b_yes/no question



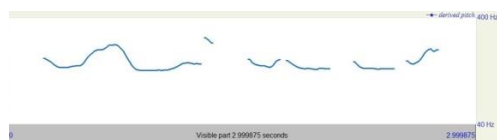
Q.2b_S5_T1



Q.2b_S5_T2



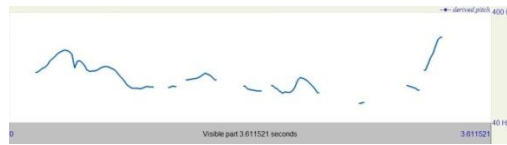
Eng. Model Q.7b_yes/no question



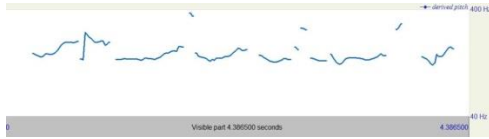
Q.7b_S5_T1



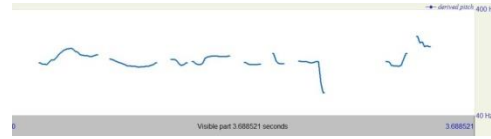
Q.7b_S5_T2



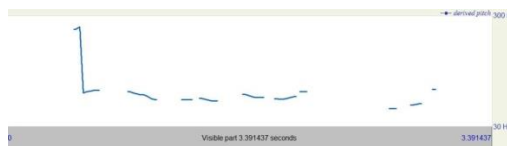
Eng. Model Q.3b_tag question – up tag



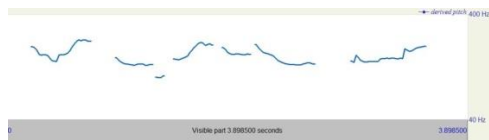
Q.3b_S5_T1



Q.3b_S5_T2



Eng. Model Q.8b_tag question – up tag



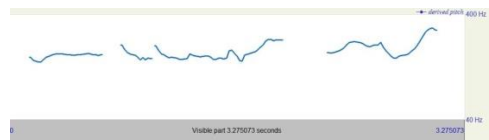
Q.8b_S5_T1



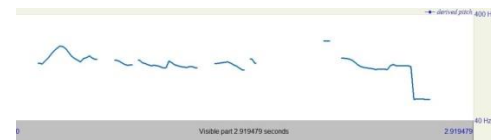
Q.8b_S5_T2



Eng. Model Q.4b_tag question – down tag



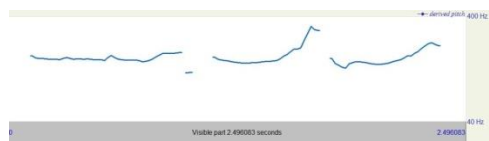
Q.4b_S5_T1



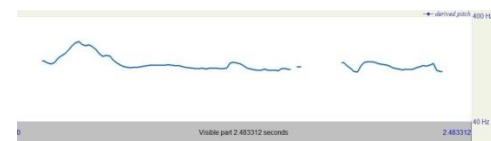
Q.4b_S5_T2



Eng. Model Q.6b_tag question – down tag



Q.6b_S5_T1

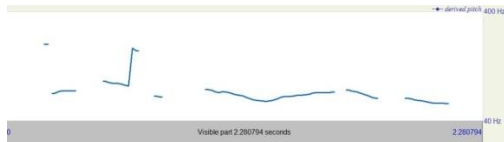


Q.6b_S5_T2

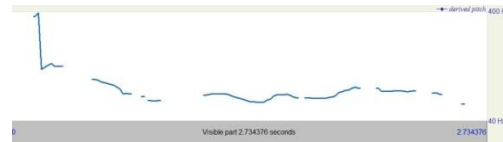
Appendix T: Intonation Contours by Student 6



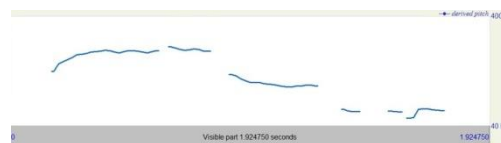
Eng. Model Q.1b_Wh-question



Q.1b_S6_T1



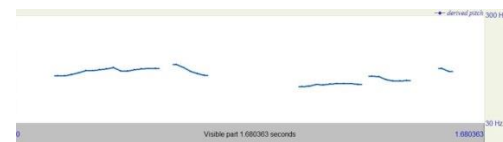
Q.1b_S6_T2



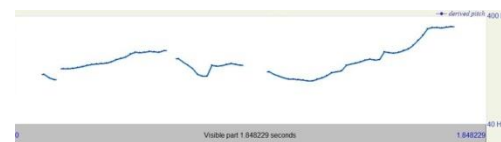
Eng. Model Q.5b_Wh-question



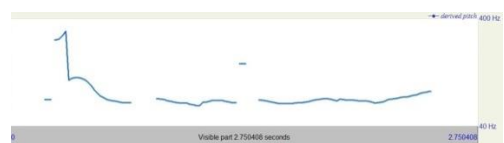
Q.5b_S6_T1



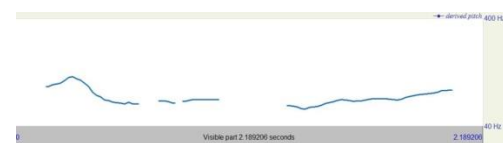
Q.5b_S6_T2



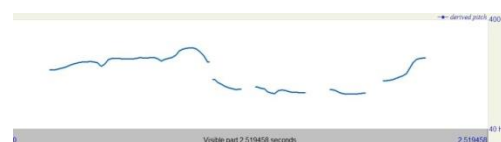
Eng. Model Q.2b_yes/no question



Q.2b_S6_T1



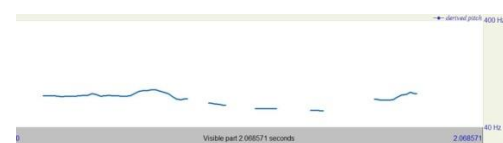
Q.2b_S6_T2



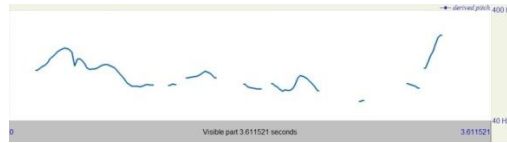
Eng. Model Q.7b_yes/no question



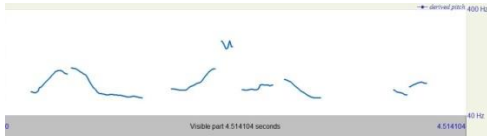
Q.7b_S6_T1



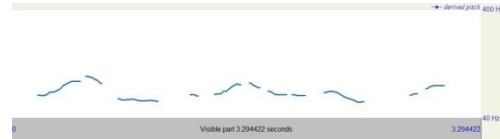
Q.7b_S6_T2



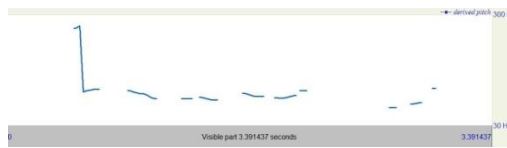
Eng. Model Q.3b_tag question – up tag



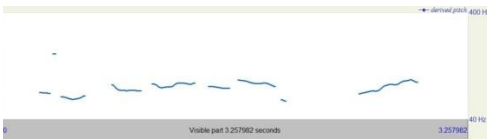
Q.3b_S6_T1



Q.3b_S6_T2



Eng. Model Q.8b_tag question – up tag



Q.8b_S6_T1



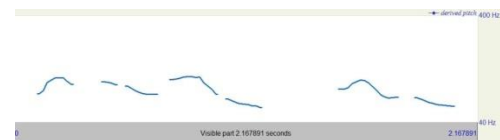
Q.8b_S6_T2



Eng. Model Q.4b_tag question – down tag



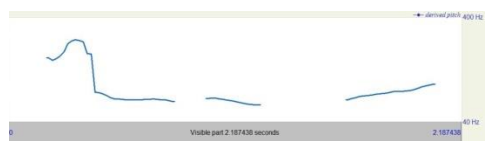
Q.4b_S6_T1



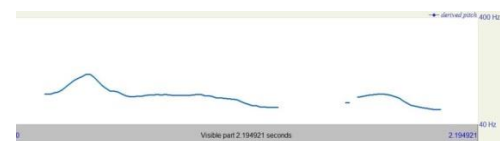
Q.4b_S6_T2



Eng. Model Q.6b_tag question – down tag



Q.6b_S6_T1

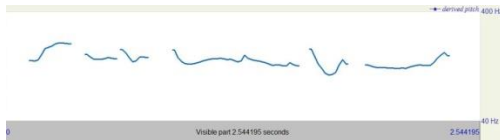


Q.6b_S6_T2

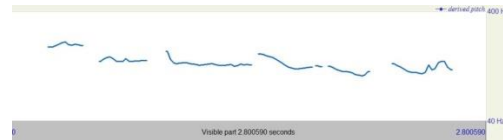
Appendix U: Intonation Contours by Student 7



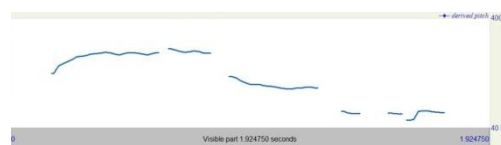
Eng. Model Q.1b_Wh-question



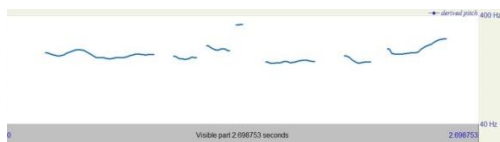
Q.1b_S7_T1



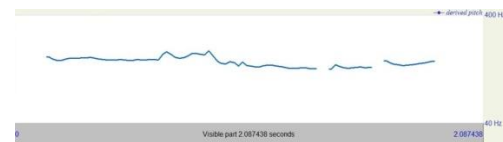
Q.1b_S7_T2



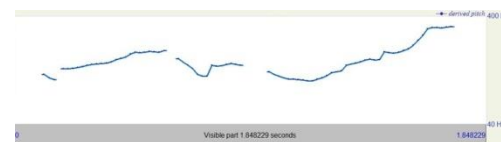
Eng. Model Q.5b_Wh-question



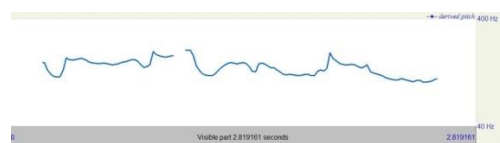
Q.5b_S7_T1



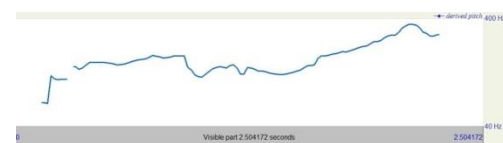
Q.5b_S7_T2



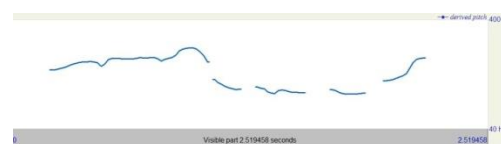
Eng. Model Q.2b_yes/no question



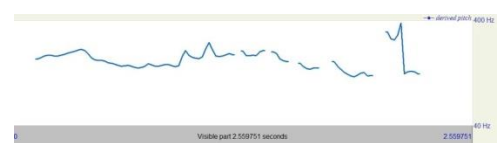
Q.2b_S7_T1



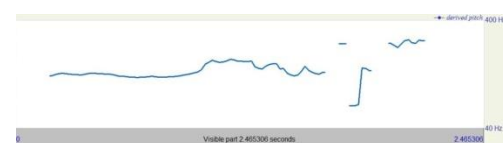
Q.2b_S7_T2



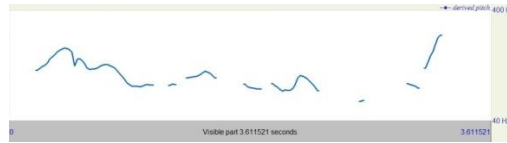
Eng. Model Q.7b_yes/no question



Q.7b_S7_T1



Q.7b_S7_T2



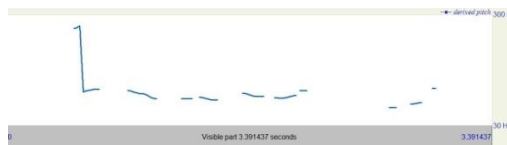
Eng. Model Q.3b_tag question – up tag



Q.3b_S7_T1



Q.3b_S7_T2



Eng. Model Q.8b_tag question – up tag



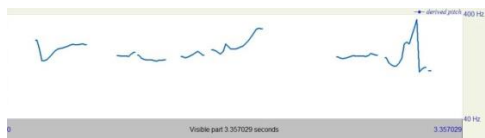
Q.8b_S7_T1



Q.8b_S7_T2



Eng. Model Q.4b_tag question – down tag



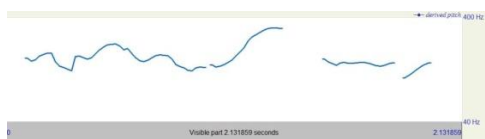
Q.4b_S7_T1



Q.4b_S7_T2



Eng. Model Q.6b_tag question – down tag



Q.6b_S7_T1

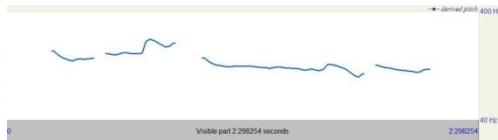


Q.6b_S7_T2

Appendix V: Intonation Contours by Student 8



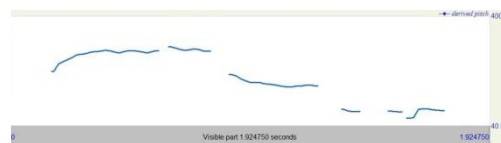
Eng. Model Q.1b_Wh-question



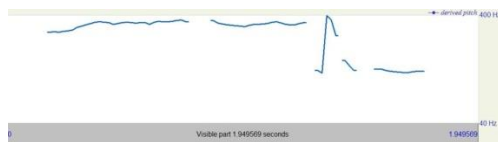
Q.1b_S8_T1



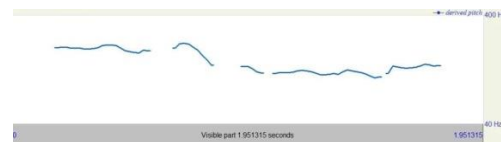
Q.1b_S8_T2



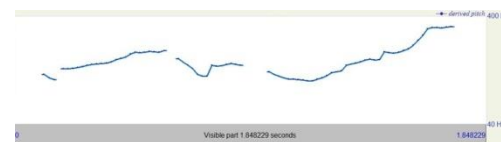
Eng. Model Q.5b_Wh-question



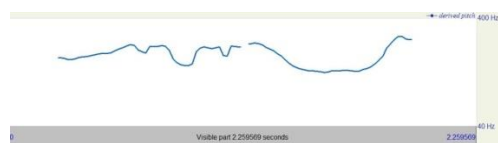
Q.5b_S8_T1



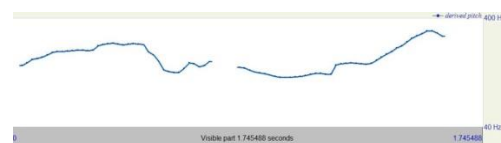
Q.5b_S8_T2



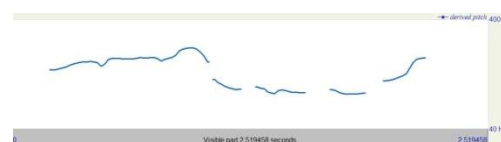
Eng. Model Q.2b_yes/no question



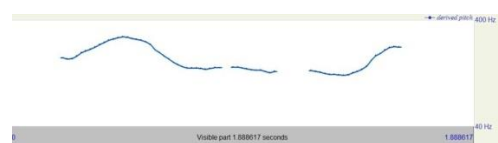
Q.2b_S8_T1



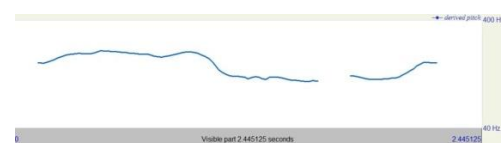
Q.2b_S8_T2



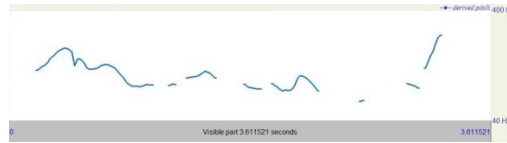
Eng. Model Q.7b_yes/no question



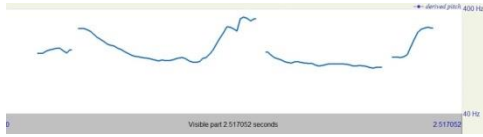
Q.7b_S8_T1



Q.7b_S8_T2



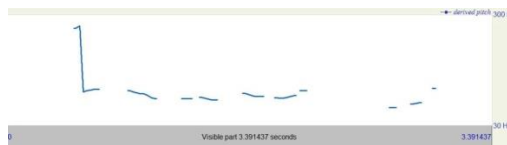
Eng. Model Q.3b_tag question – up tag



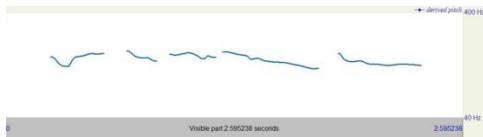
Q.3b_S8_T1



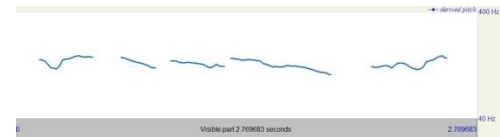
Q.3b_S8_T2



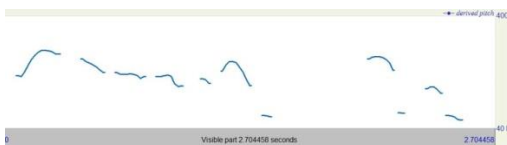
Eng. Model Q.8b_tag question – up tag



Q.8b_S8_T1



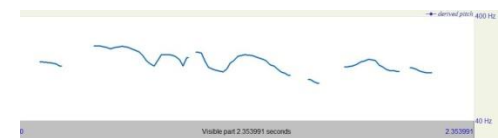
Q.8b_S8_T2



Eng. Model Q.4b_tag question – down tag



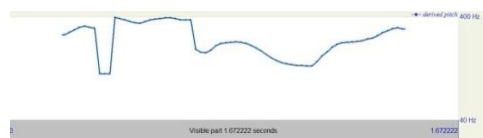
Q.4b_S8_T1



Q.4b_S8_T2



Eng. Model Q.6b_tag question – down tag



Q.6b_S8_T1

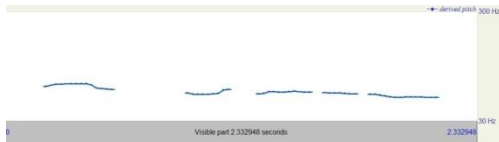


Q.6b_S8_T2

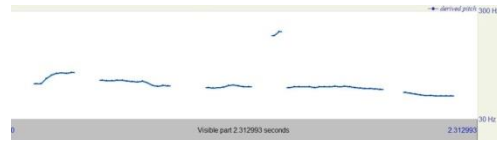
Appendix W: Intonation Contours by Student 9



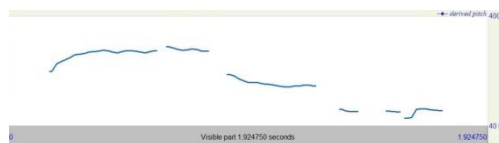
Eng. Model Q.1b_Wh-question



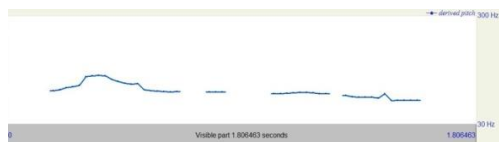
Q.1b_S9_T1



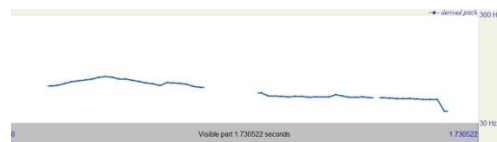
Q.1b_S9_T2



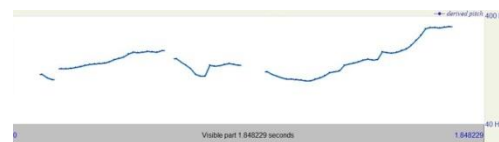
Eng. Model Q.5b_Wh-question



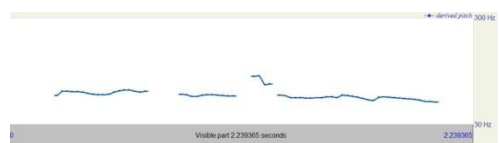
Q.5b_S9_T1



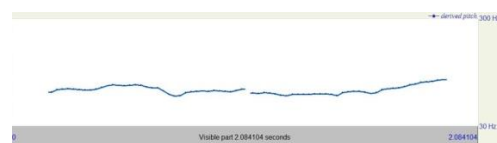
Q.5b_S9_T2



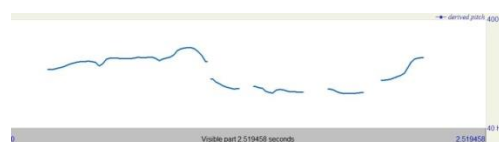
Eng. Model Q.2b_yes/no question



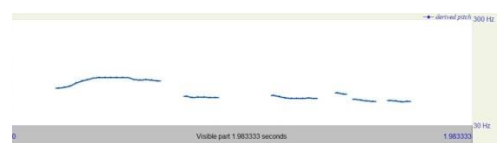
Q.2b_S9_T1



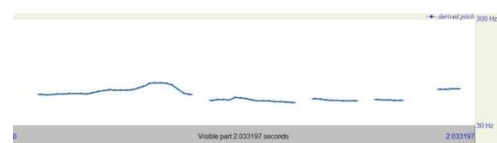
Q.2b_S9_T2



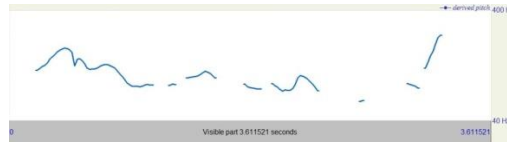
Eng. Model Q.7b_yes/no question



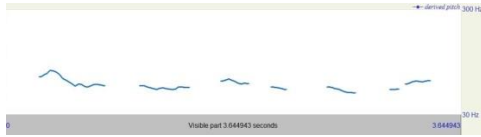
Q.7b_S9_T1



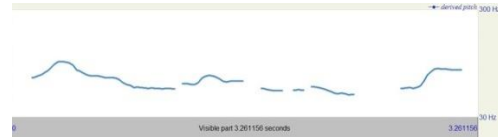
Q.7b_S9_T2



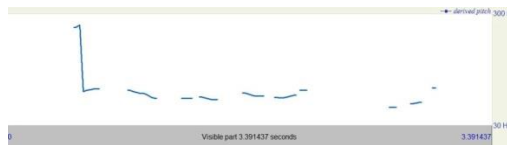
Eng. Model Q.3b_tag question – up tag



Q.3b_S9_T1



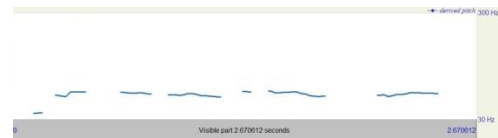
Q.3b_S9_T2



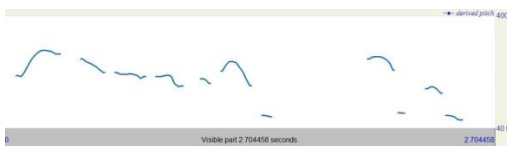
Eng. Model Q.8b_tag question – up tag



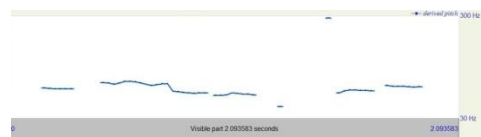
Q.8b_S9_T1



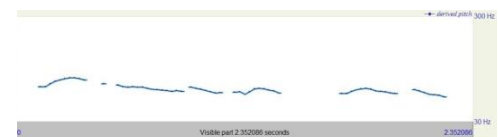
Q.8b_S9_T2



Eng. Model Q.4b_tag question – down tag



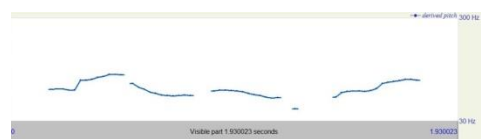
Q.4b_S9_T1



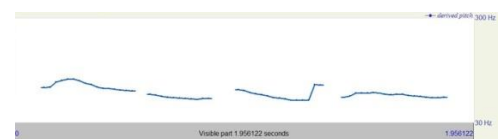
Q.4b_S9_T2



Eng. Model Q.6b_tag question – down tag



Q.6b_S9_T1



Q.6b_S9_T2

Appendix X: Intonation Contours by Student 10



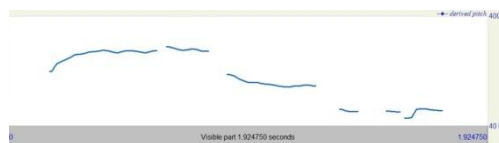
Eng. Model Q.1b_Wh-question



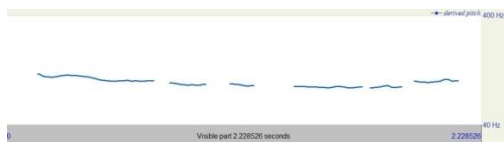
Q.1b_S10_T1



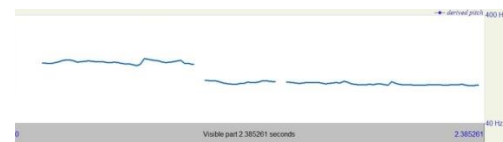
Q.1b_S10_T2



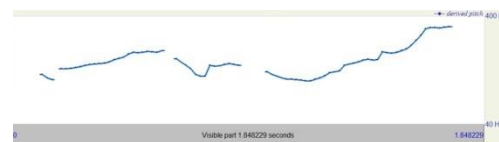
Eng. Model Q.5b_Wh-question



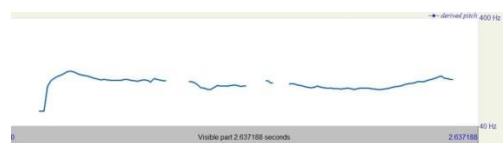
Q.5b_S10_T1



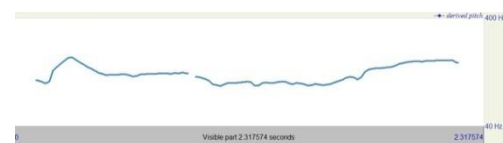
Q.5b_S10_T2



Eng. Model Q.2b_yes/no question



Q.2b_S10_T1



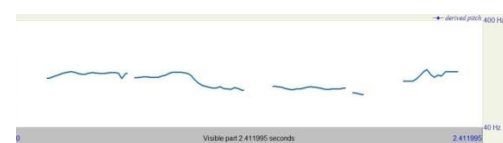
Q.2b_S10_T2



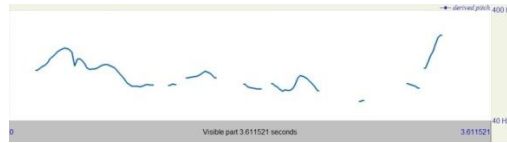
Eng. Model Q.7b_yes/no question



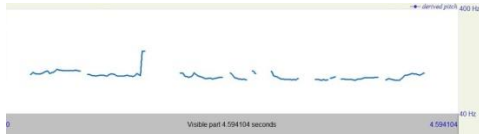
Q.7b_S10_T1



Q.7b_S10_T2



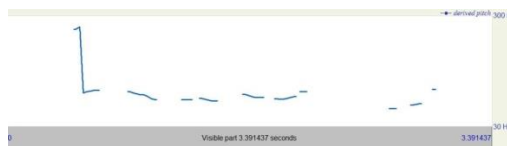
Eng. Model Q.3b_tag question – up tag



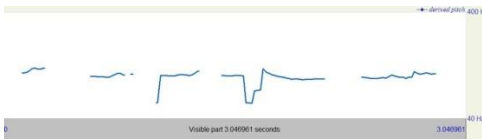
Q.3b_S10_T1



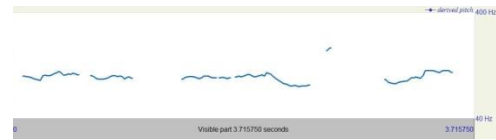
Q.3b_S10_T2



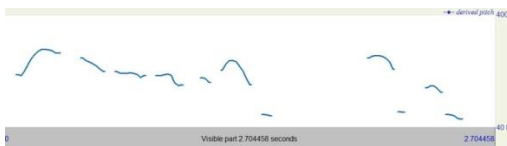
Eng. Model Q.8b_tag question – up tag



Q.8b_S10_T1



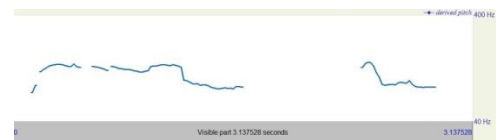
Q.8b_S10_T2



Eng. Model Q.4b_tag question – down tag



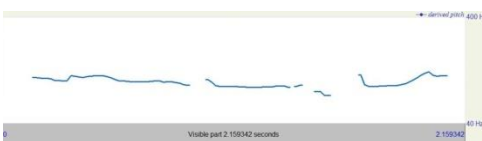
Q.4b_S10_T1



Q.4b_S10_T2



Eng. Model Q.6b_tag question – down tag



Q.6b_S10_T1



Q.6b_S10_T2

Appendix Y: Intonation Contours by Student 11



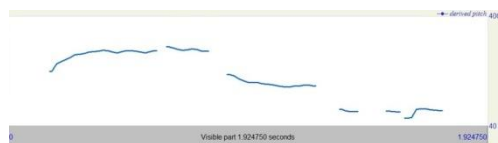
Eng. Model Q.1b_Wh-question



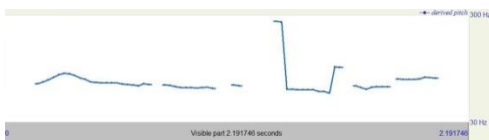
Q.1b_S11_T1



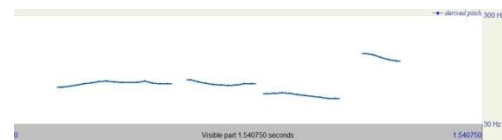
Q.1b_S11_T2



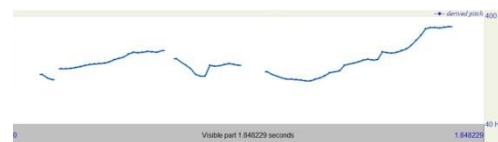
Eng. Model Q.5b_Wh-question



Q.5b_S11_T1



Q.5b_S11_T2



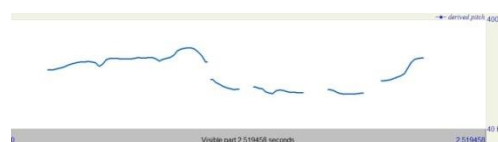
Eng. Model Q.2b_yes/no question



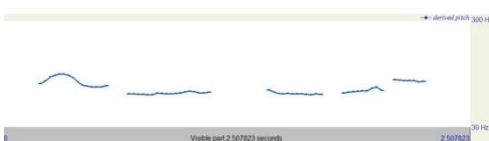
Q.2b_S11_T1



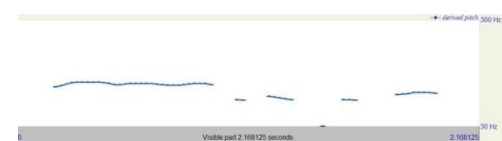
Q.2b_S11_T2



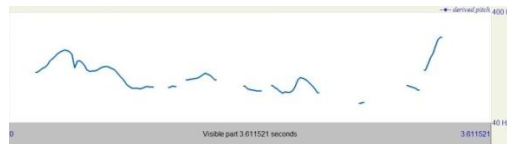
Eng. Model Q.7b_yes/no question



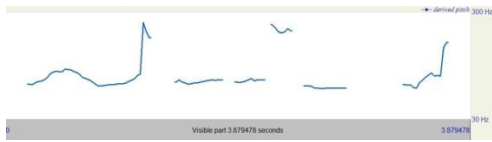
Q.7b_S11_T1



Q.7b_S11_T2



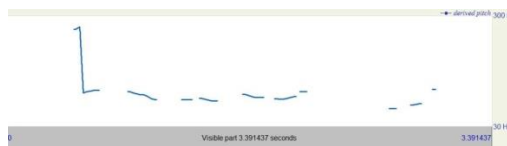
Eng. Model Q.3b_tag question – up tag



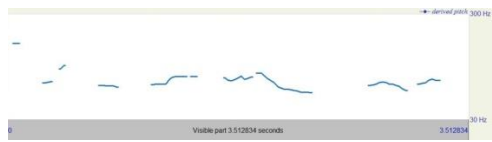
Q.3b_S11_T1



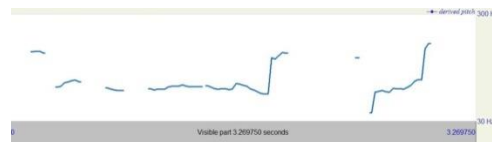
Q.3b_S11_T2



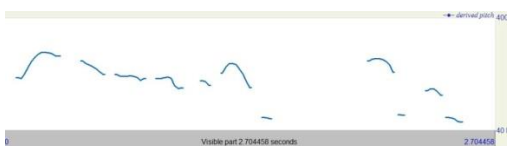
Eng. Model Q.8b_tag question – up tag



Q.8b_S11_T1



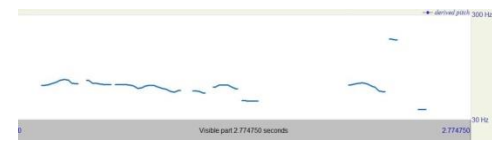
Q.8b_S11_T2



Eng. Model Q.4b_tag question – down tag



Q.4b_S11_T1



Q.4b_S11_T2



Eng. Model Q.6b_tag question – down tag



Q.6b_S11_T1

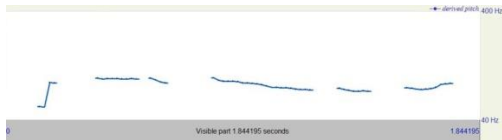


Q.6b_S11_T2

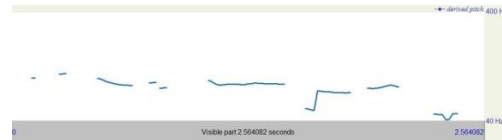
Appendix Z: Intonation Contours by Student 12



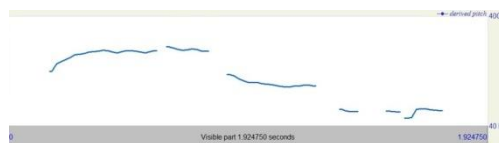
Eng. Model Q.1b_Wh-question



Q.1b_S12_T1



Q.1b_S12_T2



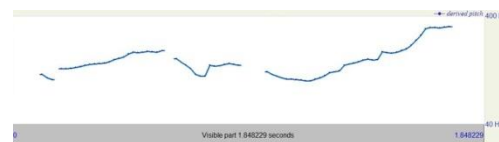
Eng. Model Q.5b_Wh-question



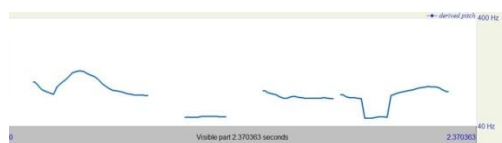
Q.5b_S12_T1



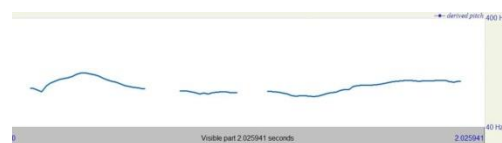
Q.5b_S12_T2



Eng. Model Q.2b_yes/no question



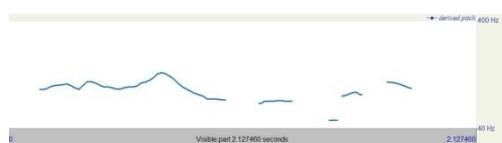
Q.2b_S12_T1



Q.2b_S12_T2



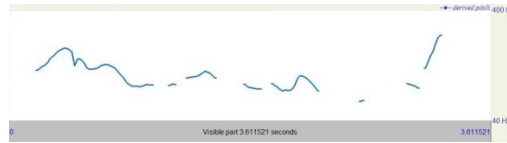
Eng. Model Q.7b_yes/no question



Q.7b_S12_T1



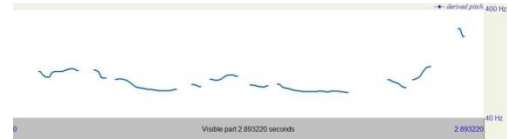
Q.7b_S12_T2



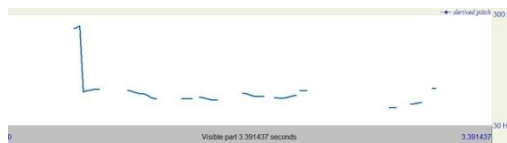
Eng. Model Q.3b_tag question – up tag



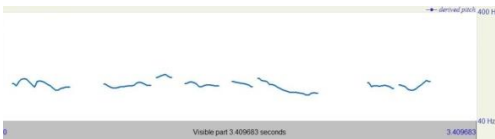
Q.3b_S12_T1



Q.3b_S12_T2



Eng. Model Q.8b_tag question – up tag



Q.8b_S12_T1



Q.8b_S12_T2



Eng. Model Q.4b_tag question – down tag



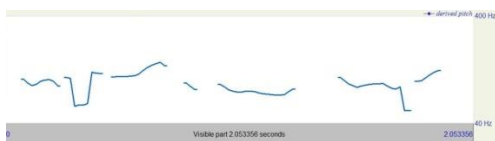
Q.4b_S12_T1



Q.4b_S12_T2



Eng. Model Q.6b_tag question – down tag



Q.6b_S12_T1



Q.6b_S12_T2