Claudia Roberta Combei

SPEAKING ITALIAN WITH A TWIST

A Corpus Study of Perceived Foreign Accent

Materiali Linguistici. Collana a cura del Dipartimento di Studi Umanistici – Sezione di Linguistica Teorica e Applicata – dell'Università di Pavia

Materiali Linguistici è una collana fondata nel 1990, che pubblica studi su vari aspetti delle lingue naturali e della loro evoluzione storica. La collana riflette la molteplicità di punti di vista teorici e metodologici che la linguistica oggi comprende, con riferimento in particolare ai seguenti temi: descrizione dei sistemi linguistici, tipologia, sociolinguistica, semantica, pragmatica e apprendimento linguistico.
Ogni volume è sottoposto a processo di peer review.

Materiali Linguistici. *Book Series edited by the Department of Humanities – Section of Theoretical and Applied Linguistics – University of Pavia*

Materiali Linguistici *is a peer-reviewed series founded in 1990. It publishes studies on various aspects of natural languages and their historical evolution. The series reflects the wide range of methodological and theoretical approaches of contemporary linguistics, with particular reference to the following themes: description of linguistic systems, typology, sociolinguistics, semantics, pragmatics and language learning.*

Claudia Roberta Combei

SPEAKING ITALIAN WITH A TWIST

A Corpus Study
of Perceived Foreign Accent

FrancoAngeli

OPEN ACCESS

Table of contents

Introduction

Languages are highly complex systems consisting of a combination of both shared and unique traits that collectively shape how they look and sound. When learning a non-native language, learners possess varying degrees of awareness regarding its characteristics and structure. While some learners may easily grasp these aspects, others might struggle to fully comprehend them. Nevertheless, learners usually make a conscious effort to acquire them. There are, however, obstacles on the path of achieving native-like language proficiency, with phonology being one of the most challenging aspects in second language acquisition[1]. Mastering the phonetics and phonology of a non-native language requires retuning one's perception and production mechanisms to account for entirely new patterns or contrasts, alongside familiar cues (Flege, 2007; Kuhl et al., 2008). In the early stages of learning a new language, learners may indeed apply the phonetics and phonology of their native language or other languages they learned before the target language. Through the ears of native speakers these phenomena are perceived as deviant, or foreign, resulting in what is commonly known as 'foreign accent'.

In the globalized world we live in, where many people are active speakers of at least one non-native language, it is highly likely that foreign-accented speech is heard on a daily basis. However, this does not imply that attitudes towards foreign accents are always positive or neutral. On the contrary, native speakers may have varying reactions to non-native speech. For instance, when the Netflix series Narcos was first aired, several controversial headlines appeared, such as: "*Narcos is a hit for Netflix but iffy accents grate on Colombian ears*"[2], "*Netflix's Narcos: Pablo Escobar's accent has*

---

1. It is worth noting that phonology is typically mastered in the first language during school age (Klein, 1996).
2. Available at: https://www.theguardian.com/world/2015/sep/17/narcos-netflix-colombian-accents (accessed on the 31st of January 2023).

annoyed Colombians. Wagner Moura speaks Spanish with a Brazilian accent"[3], "*Narcos en Netflix: un 'ménage' de acentos*"[4],"*Pare che in Colombia si continui a ridere ascoltando l'accento brasiliano del (grandioso) Wagner Moura*"[5]. These examples are clear evidence that accent matters, not only to the Colombian public in this specific case, but also to a more general audience.

The social implication of speaking with a foreign accent is just one of the reasons why this phenomenon has evolved from initially being treated like a "theoretical puzzle" to a topic of "great public interest" (Moyer, 2013: 3). Speaking with a foreign accent has been linked to unintelligibility, which affects communication, but it was also reported to have personal, social, communicative, and professional implications. It may lead to difficulties in interacting and integrating with the community of native speakers, as well as accessing job opportunities. Moreover, in recent years, foreign accents have presented challenges to Automatic Speech Recognition systems. The significant number of studies on foreign accents across various fields, including theoretical and applied linguistics, as well as speech technology, indicates that the topic is of great interest to the scientific community.

Building on this discussion, the present work aims to explore non-native Italian speech by focusing on six foreign accents, Russian, English, German, French, Romanian, and Spanish, that are compared to native varieties of Italian. The main hypothesis of this study is that the perception of foreign accent is influenced by various factors related to the speaker, the listener, and the style. At the same time, it is expected that certain non-native Italian accents would exhibit higher levels of recognisability compared to others, even when all other variables remain constant. I hypothesise that this distinction may be largely attributed to the presence of specific segmental and suprasegmental features that are perceived as particularly salient. Such noticeable cues contribute to the recognition of foreign accents, both by automated classification systems and by native speakers. I will test these hypotheses through a large-scale empirical study that deploys a mixed-method approach. The methodology involves a combination of

---

3. Available at: https://www.independent.co.uk/arts-entertainment/tv/news/netflixs-narcos-pablo-escobars-accent-has-annoyed-colombians-10505706.html (accessed on the 31st of January 2023).

4. Available at: https://www.elnuevoherald.com/entretenimiento/revista-viernes/article36473328.html (accessed on the 31st of January 2023). Translation into English: "Narcos on Netflix: a 'ménage' of accents".

5. Available at: https://www.wired.it/play/televisione/2016/09/15/narcos-figlio-escobar/ (accessed on the 31st of January 2023). Translation into English: "Apparently, people in Colombia still laugh when listening to the Brazilian accent of the (great) Wagner Moura".

quantitative and qualitative analyses to investigate accent perception, as well as several machine learning accent classification experiments.

Prior to presenting the results of the analyses, the first chapters of this monograph will explore several key theoretical and practical topics on foreign-accented speech. Chapter 1 will discuss some relevant theories on the acquisition of non-native speech while, Chapter 2 will provide a comprehensive exploration of the concept of foreign accent and its reflection in society. This includes the definition of foreign accent and an in-depth survey of selected research studies on perceived foreign-accented Italian. Chapter 3 will discuss the implications and challenges of non-native accent in speech technology applications. In addition, it will examine the availability of spoken learner corpora and techniques for accent recognition.

An original contribution of this work is the collection of a non-native speech database for Italian. The corpus comprises more than 8 hours of both read and spontaneous speech uttered by 122 young adult speakers. The detailed description of the corpus will be presented in Chapter 4, which marks the beginning of the experimental section of this book. In this respect, guided by the hypotheses of this study, in Chapters 5 and 6, I will seek to answer several research questions. As mentioned above, one of the main objectives of this work is to investigate how the perception of foreign accent is influenced by a variety of factors, operationalised as listener-, style-, and speaker-related variables. The analyses will, thus, account for the impact of sociolinguistic characteristics on the perceived accentedness level. Additionally, I will attempt to determine whether certain accents are more recognisable than others, even when all other variables are held constant. A significant part of this study will be dedicated to the investigation of the segmental and suprasegmental features that are perceived as the most salient for the varieties of foreign-accented Italian considered. This work will also examine the capability of baseline machine learning classification systems to distinguish between different accents, comparing their performance to that of native speakers of Italian.

Chapter 5 will provide a detailed account of the experiment conducted to measure foreign accent perception through the ears of native speakers of Italian. The study stands out for its inclusion of a vast array of accents, with a total of six different varieties that are analysed and compared to native Italian speech. Additionally, the experiment features a large number of stimuli, speakers, and listeners, all of which are analysed using advanced quantitative methods. Chapter 6 will go in a slightly different direction, as it will explore the role of segmental and suprasegmental features in accent perception and in automated accent classification tasks. The objective is to identify which of these features play a more significant role in recognising

specific accents, while also evaluating whether an automated classification system can perform at par with native Italian speakers in foreign accent identification.

This monograph will, therefore, aim to address several critical research questions concerning foreign-accented Italian by means of empirical analyses and computational techniques. The summary of its findings, along with the concluding remarks, will be presented in the Conclusions.

1. Second language speech

1.1 The beginnings of second language acquisition research

This section briefly introduces some of the concepts that marked the beginnings and the development of second language acquisition research. Detailed reviews may be found in Ellis (1994, 1997), Doughty and Long (2003), Gass and Selinker (2008), Van Patten and Benati (2015).

Second language acquisition (henceforth, 'SLA') has established itself as a scientific discipline in the last sixty years. Prior to that, topics related to SLA had been approached by scholars in the field of language teaching or in other areas of linguistics. The main aim of SLA is to characterise and account for the learner's productions, by collecting and investigating authentic samples of the language being learned[1]. Ever since this discipline came into existence, the impact of the learner's native language (L1) has been investigated in detail. The influence of L1 on the acquisition of the second language[2] (L2) has been commonly known as 'language transfer'[3], with 'transfer' being one of the key concepts in the behaviourist paradigm (Postman, 1971).

In SLA, the notion of 'transfer', which remains a topic of significant interest today (Peukert, 2015), is often associated with Lado's Contrastive Analysis Hypothesis (Lado, 1957). This hypothesis involves making comparisons at all levels of language between L1 and L2, with the purpose of identifying similarities and differences. According to Lado (1957), this type of analysis would be able to predict the ease or the difficulty of acquiring specific L2 structures. In particular, he argues that due to the significant

---

1. In this chapter 'to learn' and 'to acquire' are used without differences in meaning.

2. In this book second language (L2) is used to refer to any language (i.e. second, third, fourth, etc.) different from the native language (L1).

3. A more recent definition of 'language transfer' considers the role of the first language and/or that of other languages previously learned before learning the target language (Edwards et al. 2008); however, 'interlanguage transfer' has also been used to refer to the influence of other second languages over the language being learned (Gass and Selinker, 2008).

differences in form, meaning, and distribution of grammatical structures between even closely related languages like German and English, the habits of the learner's L1 structure tend to transfer to the L2, resulting in either difficulty or ease in learning the structure of L2. The transfer of similar structures will be easy to learn and function well in L2, while the transfer of different structures will be difficult because they will not function effectively in L2 and will require alteration.

A decade later, in 1967, Corder published *The significance of learners' errors* which may be considered the actual starting point of SLA research. This work has shifted the view on non-native productions, suggesting that deviations from the standard target language should no longer be seen as 'annoying' or 'distracting'. On the contrary, if thoroughly investigated, errors could glean valuable insights on the mechanisms and strategies of acquiring the L2, and consequently, the teaching-learning process could be tailored on the learner's built-in syllabus. Later on, Selinker (1972), who draws on Weinreich's (1953) 'interlingual identifications' and Lenneberg's (1967) 'latent language structure', formulates the concept of 'interlanguage'[4] (henceforth, IL). He describes it as the "existence of a separate linguistic system based on the observable output which results from the learner's attempted production of a target language norm" (Selinker, 1972: 214). Hence, it can be stated that IL is not only linked to the language being acquired but also to the learners' L1, as it is shaped by their cognitive processes and reflections on the L2. Independently of their L1, learners may display some universal patterns of acquisition and gradual transition stages (Colantoni et al., 2015). A notion frequently associated to IL is 'fossilization', referring to the stage at which the language learning process stabilises or ends (Adjemian, 1976; Gass and Selinker, 2008).

The Italian research community has a long-standing tradition in the field of SLA, with several scholars devoting their attention to the acquisition of L2 Italian. The limited space of this monograph may not allow a comprehensive list of all the works produced by the Italian scholars on this topic. I will mention some of the works that besides having made a significant contribution to the field of SLA are also relevant for this study. Chini's (2005) book is one such work, which provides a reflection on the various modes and challenges associated with language acquisition. The work discusses the basics of the SLA field, lays out the main theoretical models, and examines linguistic and non-linguistic factors that affect the learning process. Bernini's (2015) paper also provides a general overview of language acqui-

---

4. The concept of 'interlanguage' is extensively discussed in Larsen-Freeman and Long (1991), Brown (1994), and Gass and Selinker (2008).

sition and so does the volume edited by Giacalone Ramat (2003) that highlights the idea that individuals with different L1s can develop competence in Italian gradually and without formal instruction. The volume also suggests that despite their diversity, languages share universal properties, and the study of how they are learned can offer insights into how human language capacity works. On the same note, Giacalone Ramat's (2009) work focuses on the relationship between typological universals and second language acquisition, discussing how knowledge of linguistic universals can aid in understanding the L2 acquisition. Vedovelli's (1991) study also explores L2 acquisition in the context of immigration, while another study by Vedovelli (1994) discusses the phenomenon of fossilization, crystallization, and L2 competence in spontaneous acquisition. As far as non-native speech is concerned, Bernini's (1988) work focuses specifically on phonology in L2 Italian, while another study (Bernini, 2018) investigates the phonetics of early L2 varieties, with a focus on regularity and instability. Another important work on non-native speech is the volume edited by Chini (2015) that focuses on prosodic, pragmatic, and interactional aspects in IL and L2. Finally, Sisinni's (2016) book goes in a similar direction, as it reviews some theoretical models and methods of analysis related to the phonetics and phonology of L2 speech. Her work covers both spontaneous acquisition and formal learning, relying on original data.

The concepts mentioned in this section laid the foundations of SLA and they have continued to influence many studies in this field. Yet, as Ellis (1994) points out, research in SLA has always stood at the cross-roads: on the one hand, in the form of an established autonomous discipline with its own research community and methodology; on the other hand, in the form of various sub-fields, sometimes even subordinated to other disciplines. These circumstances, together with the fact the language is analysed at distinct levels of analysis may be the reasons why theoretical models of L2 acquisition have generally evolved to deal with specific aspects of language. In fact, it can be noted that the models proposed by the SLA community so far do not always manage to fully encompass the learner's acquisition at all linguistic levels.

1.2 The acquisition of second language speech

The SLA community has long been interested in exploring the differences between native and non-native speech. Scholars have sought to identify and explain the various phenomena associated with the perception and production of the language being acquired. Despite the widespread interest

for the topic, models explaining the acquisition of non-native phonology have started to emerge relatively late. Furthermore, while many SLA scholars would now agree that speech perception exerts a strong effect on the oral production of the L2, the first theories of non-native speech did not give perception particular weight. It was only later, with the increase of interdisciplinarity, that models of L2 phonology based on perception started to circulate. As described below, many of these latter theories postulate that the speaker's difficulty or ease with the phonology of a foreign language are, indeed, due to perceptual constraints. It is important to note that, as various studies have shown, learners overcome the perceptual difficulties with L2 sounds faster than their production difficulties (Escudero, 2007; Colantoni et al., 2015).

The following subsections will delineate some of the most relevant theoretical models that have aimed to explain the perception of non-native speech and the acquisition of L2 phonology. Instead of presenting them in a purely chronological order, the classification proposed in Gut (2009) will be adopted. It differentiates between models that interpret L2 speech in light of L1 influence and those built around universal principles. Various studies managed to outline these and other theories (reviews are available in Archibald, 1998; Van Patten and Benati, 2015; etc.) and therefore may provide further details that cannot be included in this monograph.

1.2.1 Models based on the influence of L1

The impact of the speaker's L1 in perceiving and producing speech in the target L2 has legitimately been among the most studied aspects in SLA. As mentioned in §1.1, one of the first studies aiming to elucidate L2 speech is Lado's (1957) Contrastive Analysis Hypothesis. According to this theory, the differences between L1 and L2 phonology may foresee the learner's acquisition difficulties (e.g. L2 phonemic contrasts which may represent only allophonic differences in the speaker's L1). However, Lado's assumptions with respect to L2 speech are confined to issues occurring at the segmental level. At the same, it is important to note that predicting and accounting for speech production challenges are demanding tasks that cannot be fully solved by contrastive analyses between L1 and L2 in the form advanced by Lado (1957).

Several scholars have focused on the impact of L1 on L2 speech, and they have proposed models of phonological acquisition that are centred around speech perception. Three of the most prominent models will be described below, namely Native Language Magnet (NLM and NLM-e) by Kuhl and colleagues (1992, 1994, 2008), Perceptual Assimilation Model

(PAM and PAM-L2) by Best and colleagues (1995, 2007), and Flege's (1995, 2003, 2007) Speech Learning Model (SLM). These models assume that the speaker's L1 forges her/his perception of L2 speech, but, as it will be shown in the next paragraphs, the models differ with respect to their scope, the assimilation mechanisms they consider, as well as their predictions on the learner's new phonetic categories.

Kuhl's (1992, 1994) NLM is a model of perceptual assimilation mostly interested in explaining the initial states of L1 and L2 speech perception, but it also hints at developmental and final states, from childhood through adulthood. Both NLM (Kuhl, 1992, 1994) and NLM-e (Kuhl et al., 2008) are presented, the latter being a revised model that proposes five new principles. Essentially, NLM suggests that, during early infancy, human beings establish phonetic prototypes that will eventually function as 'magnets' when non-native phonemic categories are encountered.

The first versions of NLM aim to investigate the "infants' native phonetic categories and how they could be structured through ambient language experience" (Kuhl et al., 2008: 982). According to this model, there are three developmental stages. In the first phase, due to innate auditory sensitivity, healthy infants can discriminate among the sounds of all languages. During the second phase, because of experience, the infants' sensitivity to the distributional properties of language generates phonetic representations. In time, before the speech is categorised into phonemic units, these representations become prototypes and start to act as 'perceptual magnets' for other elements of that category, hence the name 'Native Magnet Model'. The perceptual sensitivity decreases and whenever the speaker is confronted with an L2 sound, the prototype exerts an attraction force over it, so that the new sound is perceived within the space of the prototype. The settlement of the third phase causes a 'facilitation' for L1 and a 'reduction' for L2 phonetic abilities. The infants' speech perception abilities start to change after the first year of life: first, their non-native speech perception weakens, while at the same time, the perception of the L1 improves due to exposure experience.

NLM-e (Kuhl et al., 2008) derives from NLM (Kuhl, 1992, 1994) and includes five basic principles. The first principle designates two agents of change that prompt the transition from a universal to a language specific phonetic perception: the healthy infants' ability of recognising 'distributional frequencies' in patterns of phonetic units; and the 'motherese', namely the language mothers or caregivers speak in 'infant-directed speech' (ID) – this type of language is generally characterised by the use of exaggerated acoustic cues.

The second principle refers to 'native language neural commitment' (NLNC), a concept introduced by Kuhl (2000, 2004) to explain the fact that an early L1 coding will eventually influence the human beings' abilities of acquiring L2 speech. Kuhl (2008: 985) says that language exposure at an early stage leads to changes in the neural tissue and circuitry that reflect the statistical and perceptual properties of the language input. The neural networks reinforce the detection of higher-order patterns in language that rely on learned phonetic patterns, making the perception of these patterns more efficient while reducing sensitivity to alternative phonetic schemes. Therefore, native language phonetic perception reflects neural commitment that has developed over time, whereas non-native phonetic abilities reflect a less committed circuitry, which is still in an immature state.

The third principle of NLM-e is inspired by a foreign language intervention experiment conducted by Kuhl et al. (2003) on a group of L1 English-speaking infants that were exposed to L2 Mandarin Chinese speech; parallelly, under similar experimental conditions, a control group was exposed to the same number of stimuli of English speech. The outcomes of their study revealed that children exposed to Mandarin Chinese performed better than the control group on a Mandarin syllables test (i.e. a non-phonemic affricate–fricative contrast in English), suggesting that phonetic acquisition from first-time exposure could occur at 9 months. At the same time, the experiment tested the role of social interaction in language acquisition. Infants exposed to Mandarin speech delivered by a disembodied source (i.e. television or audiotape) differed significantly from the live-exposure group but not from the control group who were not exposed to Mandarin at all. These results suggest that, for complex language learning tasks, a 'social tutor' may be required.

Next, the fourth principle describes a link between perceptual representations formed because of experience with language and 'vocal imitation'. This prediction is in line with the motor theory of speech perception (Liberman et al., 1967), the direct realism (Fowler, 1986) and the mirror neuron system (Rizzolatti and Craighero, 2004). Still, Kuhl et al. (2008) consider that the link is developmental, and it is due to perceptual experience and to the mapping established between perception and production. As infants try to imitate the sounds they hear, they are guided by the degree of similarity between the sounds they produce and those stored in their memory.

The last principle of NLM-e states that the early perception of L1 phonetic units relates to later language development. The studies described by Kuhl et al. (2005) suggest that better L1 phonetic perception at 7 months predicts accelerate language growth from 14 through 30 months, while bet-

ter L2 performance predicts slower language advancement. This prediction is tested by Kuhl et al. (2008) on event-related-potential (ERP) data and conclude that the 7.5 months old infants' L1 and L2 phonetic perception of consonant contrasts would be able to predict language growth, as better native phonetic abilities are linked to faster linguistic advancement, while better non-native phonetic abilities are associated with slower progress. Specifically, they claim that the phonetic perception of consonant contrasts in both L1 and L2 by 7.5-month-old infants can predict various aspects of language growth, including the number of words produced, degree of sentence complexity, and mean length of utterance. Thus, the authors conclude that early phonetic perception plays a crucial role in language development.

Regarding the phases proposed by the NLM-e, their division differs slightly from the one presented in the first versions of NLM. Kuhl et al. (2008) suggest that in the first phase of NLM-e, corresponding to early infancy, children can discriminate all phonetic units of all languages; still, according to previous studies, performance varies across phonetic contrasts (Kuhl, 1980; Nittrouer, 2001). During the second phase, the infant's sensitivity to the distributional patterns (Kuhl et al. 1992) and the exaggerated acoustic cues of ID speech (Liu et al., 2003) affect phonetic learning; therefore, by the end of this phase, the infants' speech perception changes. On the one hand, the recognition of L1 phonetic cues is strengthened, while, on the other hand, the identification of phonetic patterns of unfamiliar languages is weakened. During the third phase, infants improve three skills: the identification of phonotactic patterns (Mattys et al., 1999); the identification of transitional probabilities between segments and syllables (Newport and Aslin, 2004); and the association between sound patterns and objects (Werker and Tees, 2002; Ballem and Plunkett, 2005). Finally, by the time the fourth phase has been reached, neural representations have started to stabilise. However, according to Kuhl et al. (2008), since neural networks are not completely formed, infants are able to learn multiple languages; for adults, these representations are stable, so exposure to an unfamiliar language will not necessarily create a new neural structure.

All in all, NLM offers an interesting proposal for interpreting the perception of L2 speech in initial phases and criticism to this model regards especially its first version. An important limitation of Kuhl's (1992, 1994) NLM regards the fact that it does not account for the way in which the separation of perceptual mappings for the L1 and L2 and the activation of overlapping cerebral regions may be influenced by L2 proficiency. Moreover, Escudero (2005) remarks that even if NLM theorises that L2 learners may create new perceptual mappings for L2 sounds, the model does not clearly explain whether the establishment of these mappings is achieved

19

through the same mechanisms used for L1. In fact, Kuhl (2000) says that the formation of mappings for the perception of L2 sounds is indeed different, but NLM does not propose an alternative acquisition mechanism.

A similar model is Perceptual Assimilation Model (PAM) described in Best (1995) and later in Best and Tyler (2007). PAM aims to explain how non-native contrasts are handled by speech perception and why some L2 sounds are easier to discriminate than others. Best (1995) claims that non-native speakers perceive unfamiliar phones on the basis of articulatory similarities and dissimilarities to L1 phonemes and contrasts. Best's (1995) model identifies three possible categorizations of L2 sounds. The first categorization involves L2 sounds being assimilated to an L1 category, cluster, or string, which may be perceived in three ways: as a good example of that category; as an acceptable but not perfectly fitting example; as a clearly deviant example. The second categorization involves L2 sounds being assimilated as an uncategorizable speech sound, even if it falls within the L1 phonological space. The third categorization involves L2 sounds not being assimilated into any speech sound, but rather heard as a non-speech sound.

Therefore, the premise is that in early infancy, human beings establish perceptual categories for L1 sounds and they will eventually learn how to utter them. As reported by Best (1995), subsequently to the establishment of the L1 phonemic inventory, unfamiliar and L2 phonemic categories are assimilated into L1 categories based on their articulatory similarities. Accordingly, on the one hand, it should be less difficult for learners to perceive and acquire L2 sounds that can be absorbed by L1 categories; while on the other hand, in case of an L2 phonemic contrast in which both elements are perceived as a single L1 sound, assigning different categories may be challenging.

Furthermore, PAM hypothesises that the deterioration of non-native speech perception begins at around 6-12 months of age. This process is explained in terms of the children's skills of recognising articulatory gestures underlying speech: their difficulties in distinguishing L2 contrasts depend on the articulatory similarity between specific L1 and L2 categories. In the same direction, Best and McRoberts (2003) and Best et al. (2009) aim to characterise this articulatory similarity by proposing Studdert-Kennedy and Goldstein's (2003) and Goldstein and Fowler's (2003) Articulatory Organ Hypothesis (AO). Best et al. (2009: 2758) start from the basic assumption that infants can break down the oral-facial system into separate articulatory organs, such as lips, tongue tip, and tongue dorsum, and recognise their roles in creating vocal tract constrictions. As a result, detecting differences between organs should be a relatively straightforward process for infants and adults, while detecting differences within a specific organ must be at-

tuned to the range of constriction locations and types used by that organ in the input speech. Essentially, it is assumed that infants possess the capability to differentiate between these articulatory organs, allowing for the perception and learning of inter-organ contrasts. This implies that L2 discrimination skills decrease in case of phonetic contrasts that involve the same articulatory organ (e.g. /θ–ð/), unlike those whose organ of articulation is different (e.g. /k–p/).

Although support for Best's (1995) PAM has been found in various studies concerning non-native perception (Best and Strange, 1992; Best and McRoberts, 2003), and although this model aims to account for patterns of sound assimilation in the acquisition of L2 speech, its first versions focus exclusively on inexperienced informants, described as 'functional monolinguals', namely people that are not actively learning or using an L2 and who are linguistically naïve to the target language (Best and Tyler, 2007). To overcome this limitation and therefore take into account experienced L2 learners, Best and Tyler (2007: 28-30) update PAM, presenting PAM-L2. This final model predicts the challenges of perceiving non-native sounds by analysing L2 minimal contrasts and defining four possible situations. The first is that only one L2 phonological category is perceived as equivalent to a specific L1 phonological category. The second refers to the possibility that both L2 phonological categories are perceived as equivalent to the same L1 phonological category, but one is perceived as more deviant than the other. The third situation involves both L2 phonological categories being perceived as equivalent to the same L1 phonological category, but with equal or poor instances of that category. Finally, the fourth possibility is that there is no L1-L2 phonological assimilation.

Despite not providing explicit predictions concerning the challenges involved in L2 production, both Best's (1995) PAM and Best and Tyler's (2007) PAM-L2 have proven effective in empirical investigations of how non-native listeners, including both naïve and experienced L2 learners, perceive L2 contrasts (Tyler et al., 2014). These studies also aim to determine the factors that contribute to the ease or difficulty of discriminating these contrasts. Some recent implementations of PAM in empirical research are Chen et al.'s (2020, 2023) studies on non-native lexical tones. The results suggest that the principles of PAM are applicable to non-native tone assimilations and perception by native listeners of other tone languages. However, the role of native phonological and phonetic contributions in non-native speech perception varies between categorization and discrimination tasks, as evident from the distinct modulation effects of memory load and stimulus variability.

Some initial criticism of PAM and PAM-L2 regarded the fact that they were exclusively interested in segments, so no prediction could be made regarding the acquisition of non-native prosodic features. Additional criticisms of PAM have been raised regarding its insufficient consideration of the L2 end state and the potential impact of L2 development on L1 perception, as noted by Escudero (2005).

In 1995, simultaneously with Best's (1995) PAM, Flege proposes SLM, a model of non-native speech acquisition that has had a major impact in the field of L2 phonology. SLM is based on previous work by Flege (1988a, 1988b, 1992), but the version presented in 1995 has evolved and undergone variations over time, as new empirical data were obtained by Flege and his colleagues (Flege et al., 1999; Piske et al., 2001; Flege, 2003, 2007). SLM, just like Best's (1995) PAM, assumes that the learner's performance in perceiving and producing non-native speech sounds is conditioned by the extent of the perceived similarity between L1 and L2 sounds. More specifically, SLM aims to investigate age-related changes, if any, in non-native speech learning. Hence, the model is mostly interested in the ultimate attainment of L2 phonology, but it also touches upon initial and developmental states.

Before presenting the premises of Flege's (1995, 2003, 2007) SLM, one needs to recall Lenneberg's (1967) Critical Period Hypothesis (CPH), which exploits the homonymous concept coined by Penfield and Roberts (1959). CPH suggests that due to neuroplasticity, after puberty, language acquisition is less effective, and therefore native-like proficiency may be difficult or even impossible to achieve. Consequently, regarding the level of phonology, according to Lenneberg, this would imply that there is a rather limited period of time (i.e. from age two until puberty, a period of time allegedly corresponding to the lateralization – a set of processes and mechanisms that leads to the specialization of the dominant brain hemisphere for language-related functions) for the accurate acquisition of segments[5]. He states that "[f]oreign accents cannot be overcome easily after puberty" (Lenneberg, 1967: 176).

---

5. Although early acquisition processes may be to some extent both desirable and beneficial, empirical research failed to support the predictions of Lenneberg's (1967) CPH (Flege et al., 1999; Singleton, 2005; Ortega, 2009; Muñoz and Singleton, 2011). As suggested by Scovel (2000), Flege and Liu (2001), Moyer (2004, 2007, 2013), and Flege (2018), the quality and the quantity of L2 input (i.e. experience and training), as well as the learner's attitude towards the L2 may be the factors that mostly influence the rate of success in reaching native-like performance in L2 speech (see Chapter 2 for a detailed description of these factors).

While Lenneberg (1967) suggests that there is a biologically determined age limit for successful second language acquisition, Flege's research posits the opposite, namely that individuals of all ages who learn a L2 a second language are capable of auditorily detecting phonetic differences across languages (Flege, 1999). Furthermore, he contends that L2 learners retain the same cognitive capacities used during their first language acquisition, such as the ability to establish new representations and convert secondary-based perceptual information into articulation (Flege, 2008: 177).

Flege's (1995) SLM is built around four postulates (P1-P4) and seven hypotheses (H1-H7). In short, according to (P1), when one learns an L2, she/he makes use of the same cognitive mechanisms and processes used for L1. Then, (P2) states that the speaker is endowed with mental representations of the phonetic categories of her/his L1. When other languages are acquired, the existing categories may undergo variations (P3), while, at the same time, new ones may be created (P1). However, according to (P4), learners try to keep L1 and L2 phonetic categories separated. The seven hypotheses predict the process of phoneme acquisition. Flege (1995) assumes that speakers compare L1 and L2 sounds at the allophonic level (H1). According to (H2), to create new phonetic categories, speakers should be able to perceive at least some of the phonetic differences between a given pair of L1 and L2 sounds. Additionally, the likelihood of establishing a new phonetic category increases as a function of the perceived phonetic dissimilarity between L1 and L2 (H3). Based on (H4), the chance that L1-L2 phonetic differences will be perceived decreases with the age of learning (AOL) of the L2. Next, (H5) and (H6) predict the bidirectionality of the cross-language phonetic interference. Also, in case of an 'equivalence classification', meaning that the learner does not perceive any phonetic difference between two sounds, the creation of a category for an L2 sound is blocked, so the speaker gets the representation of a single phonetic category for both sounds. Equally important, according to (H6), the newly established categories by L2 learners for L2 sounds may differ from the categories of native speakers. Finally, (H7) states that the L2 sound will eventually be uttered according to the properties represented in its phonetic category.

Flege (1995, 2003, 2007) distinguishes between three types of sounds: 'new', 'similar', and 'same'. Those sounds that are perceived as the same in L1 and L2 are the easiest to produce, since they do not need to be learned. In the SLM framework, the phonetic category created during childhood for an L1 sound evolves to embody the L2 sound. On the other hand, new categories are established when the L2 sound is perceived as distant from the closest L1 sound (i.e. the L2 sound is not present in the learner's L1 inven-

tory), because the mechanisms required to create new sounds remain intact (Flege, 2008). Finally, similar sounds seem to be the most challenging since they display some phonetic characteristics of L1 sounds but the two are not identical. In this situation, due to the 'equivalence classification', the establishment of a new category may fail, and as a result, this will lead to a non-native like realisation of the L2 sound.

As discussed above, for Flege (1995), AOL may play a role on the learner's ability to perceptually discriminate between the phonetic characteristics of L1 and L2 sounds, in the sense that the younger the speaker, the greater the chance that the L1-L2 differences will be perceived. Even if SLM focuses mainly on the end state, its assumptions regarding the difficulty of reaching a native-like performance are strongly linked to initial and developmental stages of non-native sound perception (Flege, 2003).

In a more recent paper, Flege and Bohn (2021) present the revised Speech Learning Model (SLM-r) which focuses on the learning of L2 vowels and consonants across the lifespan. While the original SLM aimed to account for age-related limits on the ability to produce L2 sounds in a native-like fashion, the SLM-r aims to provide a better understanding of how phonetic systems reorganise over the lifespan in response to phonetic input received during naturalistic L2 learning. Although some aspects of the original SLM have been carried forward to the SLM-r without change, the SLM-r replaces the original 'age hypothesis' with a new hypothesis that may help explain age-related effects on L2 speech learning. Flege and Bohn (2021) argue that differences in learning outcomes between a learner's first language (L1) and second language (L2) are inevitable due to three factors. First, L1 sounds can 'substitute' L2 sounds because the learner automatically links L2 sounds to their L1 phonetic inventory. Secondly, preexisting L1 phonetic categories can interfere with or block the formation of new phonetic categories for L2 sounds. Lastly, the input for learning L2 sounds differs from that of monolingual native speakers of the target language. The authors say that SLM-r has yet to be tested empirically.

SLM is the result of a considerable amount of research on L2 speech conducted by Flege and his colleagues (e.g. Flege, 1987; Yeni-Komshian et al., 2000; Flege, 2003). The model prior to SLM-r has been employed as a reference framework in many studies on the acquisition of L2 speech (e.g. McAllister et al., 2002; Flege et al., 2003; Flege and McKay, 2004; Flege and Wayland, 2019; Combei et al., 2020). Nevertheless, as pointed out by some of these studies, SLM does not provide any standardised measure of perceived L1-L2 phonetic distance. In addition, the first versions of the model do not fully account for the process of phonetic discrimination. Another shortcoming of SLM regards its scope: the model only covers the lev-

el of phonemes and allophones, so no predictions are available for the acquisition of prosody or the acquisition of phonological processes. Nonetheless, Flege is aware of the weight of prosodic features and states that "nonsegmental dimensions are an important source of foreign accent" (Flege, 1995: 233). On a different note, SLM does not explicitly discuss the process by which a new phonetic category is formed, and that no prediction is made about whether this process may be improved or disrupted; also, SLM does not describe the effect that the formation of a new phonetic category has for the non-native phonological production (Gut, 2009). Finally, linguistic and non-linguistic factors, other than just Flege's (1995) AOL, have a role on the acquisition of L2 speech. In fact, later on, Flege (2018) reinterprets the impact of this variable, claiming that 'input' is actually a better predictor for the pronunciation accuracy.

1.2.2 Models based on universal principles

The influence of the native language is not the only factor taken into consideration for the design of theoretical models of L2 speech. Scholars have also built theories around the role of language universals in the acquisition of L2 phonology. Broadly speaking, universals of language are statements that are true for all languages (Bickel, 2010). But as shown below, this definition might prove to be too wide. The next section will briefly present some models that aim to explain the acquisition of L2 phonology, based on the concept of typological markedness and universal natural phonological processes.

In order to explain the concept of typological markedness, the definition of language universals provided above needs to be further developed. Based on the presence or absence of specific properties of any given language, universals are either absolute[6] or implicational. Absolute universals refer to universals that are always true of all languages (Whaley, 1997), while implicational universals have preconditions, meaning that two language properties can be placed in an 'if X then Y' relationship. This condition is unidirectional, in the sense that the presence of feature 'X' implies the presence of feature 'Y' but not vice versa. Also, the implicated feature is considered as less marked.

---

6. If just one language fails to support the accuracy of a supposedly absolute universal, then that generalisation is labelled as a nonabsolute universal, since it "admits exceptions […], it represents significant tendencies (e.g. most languages have adjectives), […], and it has a high degree of probability" (Whaley, 1997: 32).

Therefore, markedness universals concern occurrence relationships (Major, 2001), and the typological markedness may be defined as the "asymmetric relationship that is inferred to hold between language structures" (Gut, 2009: 25). For instance, the presence of final voiced obstruents in any given language implies their presence also in initial and medial position but not vice versa (Eckman 1977, 1984, 1985; Eckman and Iverson, 1994; Major, 2001). Furthermore, as Major (2001) points out, the markedness applies also to the order of L1 speech acquisition: for example, infants acquire front unrounded vowels prior to front rounded vowels.

Two decades later after Lado's (1957) CAH, Eckman's (1977) Markedness Differential Hypothesis (MDH) brings the concept of typological markedness[7] in the field of SLA. Eckman (1977: 315) suggests that CAH needs to be updated to include the concept of 'degree of difficulty' which is determined by typological markedness. He indicates that typological markedness is independent of any language and anything regarding L2 acquisition, so incorporating it into the CAH can help predict both the areas and the degree of difficulty for L2 learners. Additionally, Eckman argues that typological markedness is a natural and reasonable concept of difficulty based on certain assumptions about language and human learning.

While aiming to recover CAH, Eckman's (1977) MDH also attempts to explain and predict L2 phonological acquisition. One of the major tenets of the MDH is that unmarked phenomena are acquired before marked phenomena. According to Eckman (1977), there are three key claims regarding the areas of difficulty that a learner of L2 may encounter. Firstly, areas of the L2 that differ from the learner's L1 and are more marked than the L1 are likely to be difficult to acquire. Secondly, the relative degree of difficulty of these areas of difference in the L2, which are more marked than the L1, corresponds to the relative degree of markedness. Lastly, areas of the L2 that differ from the L1 but are not more marked than the L1 are not expected to present significant difficulties for the learner. Hence, by considering this notion of typological markedness, it is possible to predict the areas of difficulty for an L2 learner and the relative degree of difficulty. In addition, the conclusion that emerges from the MDH hypotheses is that not all L1-L2 differences will cause the same type of difficulty for the L2 learner.

Another model that employs the concept of typological markedness is Eckman's (1991) Structural Conformity Hypothesis (SCH). He reveals a series of issues with his MDH, namely that it has to refer to both the L1 and L2, not just universals, and it has to explain the L2 learner's difficulties in

---

7. Eckman generally uses the concept of markedness in its purely implicational hierarchical sense.

areas where L1 and L2 features do not differ. For these reasons, he proposes SCH. Rather than relying on universal generalizations to predict the order of acquisition, relative difficulty, or prevalence of errors in SLA, Eckman's (1991: 24) SCH seeks to test whether interlanguages (ILs) conform to universal generalizations. In other words, the focus is on examining the actual characteristics of interlanguages to verify whether they align with the predicted patterns.

Following Adjemian (1976), who claims that ILs are in fact languages, Eckman's (1991) SCH suggests that all universals that hold for primary languages should also hold for ILs. Conversely, phonological structures or processes that are not attested in the learners' L1 will not be found in their version of L2 speech. The claims of MDH and SCH have been tested on various empirical studies (Eckman, 1991; Carlisle, 1997, 1998, 1999) in the field of phonology, for instance, regarding the fricative-stop principle[8] and the resolvability principle[9]. The results of these studies suggest that, in most cases, the empirical data adhere to the afore-mentioned universal generalizations.

However, the scope and the explanatory and predictive power of MDH and SCH are limited. Considering that the two models focus on the extent to which some linguistic structures pose challenges to an L2 learner, one cannot make any predictions regarding the stages of the acquisition process. As already argued for other theories of L2 phonology, MDH and SCH do not account for prosodic phenomena, since, in this case, they cover only syllabic structures. Gut (2009) also criticises the fact that MDH and SCH consider neither the impact of perception, nor non-linguistic factors in L2 phonological acquisition.

Another theory that fits the category of models based on universal principles is Major's (2001) Ontogeny Phylogeny Model (OPM) that is a revised version of his first (1987) Ontogeny Model (OM). In short, OM assumes that "transfer processes decrease over time, while developmental processes increase and then decrease" (Major, 2001: 80). However, OM is described by Major (2001) himself as a 'model of performance' rather than 'competence'. In addition, the model only refers to L1 transfer and

---

8. Eckman's (1991: 24) assumptions based on Greenberg's (1978) fricative-stop principle: "If a language has at least one final consonant sequence consisting of stop + stop, it also has at least one final sequence consisting of fricative + stop" (e.g. 'picked' and 'rift': /pɪkt/ and /rɪft/ in English).

9. Eckman's (1991: 25) assumptions based on Greenberg's (1978) resolvability principle: "If a language has a consonant sequence of length m in either initial or final position, it also has at least one continuous sequence of length m - 1 in this same position" (e.g. 'fixed' and 'fix': /fɪkst/ and /fɪks/ in English).

developmental processes, but it is not explicit regarding the L2 nor the stages of development. Moreover, OM does not make any claims regarding markedness.

For all these reasons, the revised OPM was proposed. Major (2001: 81) explains that OPM considers two different perspectives. First, ontogenetically, the model focuses on the development of the speaker's IL, or their language development over time. Then, phylogenetically, OPM considers larger populations and factors such as language contact and change. Therefore, OPM covers both the individual and evolutionary aspects of language development. In a nutshell, OPM theorises an interrelationship between L1 transfer, L2, and universals, all governed by similarity and markedness. Major (2001) states that the model deals with language development in individuals and groups over a variable period of time (i.e. from a few seconds up to a lifetime), or even over generations, as well as with changes in languages and language families.

OPM makes four main claims with respect to chronology, style, similarity, and markedness. During the first stages, L1 transfer is greatest but afterwards it decreases gradually, while on the other hand, L2 structures are inexistent at the beginning, and they increase in later stages. Regarding the universals, their influence increases in initial stages but then it decreases. The stylistic corollary states that as style becomes more formal, L2 increases, while on the contrary, L1 decreases; additionally, the role of universals increases and then it decreases. Next, based on the similarity corollary of the OPM, the acquisition of structures that are similar in L1 and L2 firstly displays a small influence of L2 and universals, while on the other hand, the role of L1 transfer is more substantial. Finally, regarding the markedness, the acquisition of marked phenomena is theorised this way: L2 increases slowly, L1 decreases, and the influence of universals increases rapidly but then it decreases slowly.

The hypotheses of OPM are supported by various empirical studies (Major, 1994, 1996; Hancin-Bhatt and Bhatt, 1997) most of them concerning the acquisition of phonemes and syllable structures. Nevertheless, Major (2001) claims that the OPM's scope is even wider, including other phonological structures and non-phonological phenomena.

Criticism to OPM may be brought with respect to insufficient evidence regarding the role of linguistic and non-linguistic factors in the acquisition of L2 phonology. Additionally, the model does not clearly present the acquisition mechanisms that could affect the increase and decrease of the L2 components. Another weakness of OPM regards the unclear definitions of 'universals' and the ambiguous interrelation of the proposed components of IL grammar (Gut, 2009). Finally, since Major (2001, 2002) reveals that OPM is not shaped around any formal framework,

there are no proposals regarding sound representations and mappings of the acoustic signals, nor regarding the exact mechanisms of the phonological system (Escudero, 2005).

Many of the models presented in this chapter have been employed in empirical studies either by the authors that proposed them or by other scholars. As far as the Italian context is concerned, various applications of these frameworks are to be found in Costamagna and Giannini (2003) and in Costamagna and Marotta (2008). Most of the studies discussed in these volumes apply the models presented in this chapter especially on segmental features that are explored by means of acoustic analyses. For instance, Marotta and Barth's (2008) study on the production of Italian sonorants by German learners applies Major's (2001) OPM, Best's (1995) PAM, and Flege's (1995) SLM. As far as the usage of the afore-mentioned models in this book, I believe that theories accounting for the acquisition of L2 phonology as a function of L1 – especially Flege's (1995, 2003, 2007) SLM – are more adequate for explaining foreign-accented speech. That being said, this work will not provide a fine-grained acoustic analysis, so the postulations of SLM will be discussed at a very general level in relation to the data at hand.

2. Speaking with a foreign accent in Italy

The ever-growing Erasmus+ program – granting free mobility to European students – as well as the waves of migration towards Italy[1] in the last decades have increased more than ever the likelihood of hearing foreign-accented Italian on a daily basis. Foreign accents reflect our identities and origins, but at the same time they transmit information about our socio-cultural and socio-demographic backgrounds as well as other languages we had already spoken when we started to learn the target L2. In addition, our foreign accent may reflect how open we are to embrace the language and the culture of the foster country, or it could simply be a predictor of the quality and the quantity of the input we have received for the specific language we are learning (Flege, 2018).

Judging by the amount of scientific production on foreign-accented speech, it is evident that scholars in SLA have long been intrigued by this topic. In fact, a large body of theoretical and applied research has focused on the characteristics of non-native pronunciations, contrasting them to native speech. Apart from analysing the segmental and suprasegmental features of foreign accent, scholars have also tried to explore the learners' motivations of improving their accent and to observe their attitudes towards the country where the target language is spoken.

The aim of this chapter is to explore how the study of foreign accent has evolved over time and to present the state of art in SLA with respect to this topic. First, section §2.1 will provide some preliminary remarks on the concept of 'foreign accent' and its social facets, including its reflection in the media and cinema. Section §2.2 will present a survey of relevant studies on the production and the perception of non-native speech, exploring the

---

1. According to *Bilancio Demografico* (Demographic Report) published on the 20th of March 2023 by the Italian National Institute of Statistics (ISTAT), on the 31st of December 2021 there were 5,030,716 foreign citizens, from more than 200 nations, residing on the Italian territory. The report is available at: http://dati.istat.it/Index.aspx?DataSetCode=DCIS_POPSTRRES1 (accessed on the 20th of March 2023).

methodological challenges associated with the study of foreign accent and the role of several factors on the L2 pronunciation.

2.1 The perception and representation of foreign accent

In the globalized world nowadays, where many people are active speakers of at least one L2[2], foreign accent has evolved from initially being treated "largely as a theoretical puzzle" to a topic of "great public interest" (Moyer, 2013: 3). Especially in the last two decades, there has been a continuous interest in investigating features of non-native pronunciation, as well as in exploring "the social relevance and reception of L2 accent overall" (Levis and Moyer, 2014: 4).

As already noted in §1.2, scholars have formulated various theories on the acquisition of L2 speech, some concluding that the major difficulty lies in retuning one's perception and production mechanisms, so as to take into account totally new patterns or contrasts, but also similar – yet, not identical – L2 cues (Flege, 2003, 2007; Kuhl et al., 2008). Some L2 speakers never attain a native-like pronunciation, and as a matter of fact, they might sometimes find it challenging to establish themselves socially and professionally in contexts in which their foreign accent is associated with social stigma. Miller (2003) suggests that conforming to the language norms of the dominant majority can bring about social, personal, and academic benefits. Being acknowledged and accepted as a legitimate user of the target language by other legitimate users of that language is not always straightforward, especially if one speaks a different language or the target language with an accent, or in a way that is not considered native-like.

2.1.1 Foreign accent and the Italian society

Considering the scope of this monograph, the social dynamics linked to foreign-accented speech can only be briefly discussed within the following pages. The topic is largely covered in Gluszek and Dovidio (2010), Lev-Ari and Keysar (2010), Lippi-Green (2012), Levis and Moyer (2014), Calamai

---

2. According to United Nations International Migration Report 2017 available at https://www.un.org/en/development/desa/population/migration/publications/migrationreport /docs/MigrationReport2017.pdf (accessed on the 31st of January 2023), the number of international migrants worldwide has reached 258 million in 2017, showing an ever-growing trend: in 2000 there were 173 million people that lived outside the country they were born.

(2015), Timming (2017), Russo et al. (2017), Calamai et al. (2020), Nodari et al. (2021), Kang and Yaw (2021), and Bazzi et al. (2022).

The implications of speaking Italian with a non-native accent are frequently echoed in the media. As recent research has shown, some forms of foreign-accented Italian tend to be stigmatized or stereotyped (Bianchi and Calamai, 2012; Calamai, 2015; Orrù, 2017; Calamai et al., 2020). And as a matter of fact, all-encompassing labels for foreign accents have been reported in the media: *accento slavo*, *accento (dei paesi) dell'Est*, *accento dei paesi dell'Est* ('Slavic accent', 'accent of Eastern European countries', or 'Eastern European accent'), comprising presumably accents of speakers whose L1s are as typologically diverse as Romanian, Hungarian, Russian, Polish, Serbian, Albanian, and Bulgarian; in addition, *accento nordafricano* ('North African' accent) is also an ambiguous phrase to refer to foreign-accented speech.

Using the web as a corpus, both literally and with the ItWaC corpus (Baroni et al., 2009), as well as the subcorpus CORIS-STAMPA (Rossini Favretti et al., 2002), some examples[3] collected from various Italian newspapers and popular websites, as well as animation films and cartoon series will be presented here, in order to give a glimpse of how foreign-accented speech has been framed in the Italian media:

1. *L'accento sta al colpevole come il "presunto" sta al sospetto: sono entrambi modi, forse necessari, per dire senza dire. [...] [G]li accenti dei criminali, forniti come indicazione di origine: ammesso e non concesso che nel grande melting pot criminoso non esistano ormai banditi nordafricani con accento slavo, banditi slavi con accento nordafricano, banditi italiani che imitano entrambi gli accenti.* (Accent stands for the perpetrator as "alleged" stands for the suspect: both are ways, perhaps necessary, of saying without saying. [...] The accents of the criminals, provided as an indication of origin: assuming and not conceding that in the great criminal melting pot there are not now North African bandits with Slavic accents, Slavic bandits with North African accents, Italian bandits imitating both accents; Bartezzaghi, Accento slavo, *La Repubblica*, the 21st of September 2005)[4].

2. *[L]'Uomo con il violino immancabilmente borbotta qualcosa con il suo vocione dall'accento romeno, o forse bulgaro, o magari croato,*

---

3. Translations into English are provided between the brackets.
4. Available at: https://ricerca.repubblica.it/repubblica/archivio/repubblica/2005/09/21/accento-slavo.html (accessed on the 2nd of January 2023).

33

ma in ogni caso di chiara provenienza euro-orientale. (The man with the violin invariably mumbles something in his big voice with a Romanian accent, or perhaps Bulgarian, or perhaps Croatian, but anyway an Eastern European accent.; Culicchia, Gente di Torino, *La Stampa*, the 8th of August 2008)[5].

3. *Seduto di dietro, a sinistra, mi lascio sedurre dall'accento francese di Myriam.* (Sitting back, to the left, I let myself be seduced by Myriam's French accent.; Speciali Giro d'Italia, *Gazzetta dello Sport*, the 5th of May 2008)[6].

4. *In un ottimo italiano, dall'elegante accento inglese, la contessa Caroline, moglie di Leonardo Marco Emo Capodilista.* (In excellent Italian, with an elegant English accent, Countess Caroline, wife of Leonardo Marco Emo Capodilista.; Nell'archivio degli Emo c'è la storia del paese, *L'Arena*, the 26th of May 2009)[7].

5. *È un uomo con accento nord africano che fa la spola tra Cascina Gobba e piazzale Loreto.* (He is a man with a North African accent who shuttles between Cascina Gobba and Piazzale Loreto.; Vecchi, Marysthell, leader del "gruppo Olgettina". Pupilla di Silvio e donna di un narcotrafficante; *Il Fatto Quotidiano*, the 28th of January 2011)[8].

6. *Roma: i killer parlavano italiano con accento dell'Est.* (Rome: the killers were speaking Italian with an Eastern European accent.; Roma: i killer parlavano italiano con accento dell'Est, *SkyTG24*, the 7th of January 2012)[9].

5. Available at:
https://www.lastampa.it/2008/08/08/torinosette/gente-di-torino-PyZPppEDxH7o6K8jur6mpN/pagina.html (accessed on the 2nd of January 2023).

6. Available at:
https://www.gazzetta.it/Speciali/Giroditalia/2008/it/Primo_Piano/2008/05_Maggio/19/pop_presi.shtml (accessed on the 2nd of January 2023).

7. Available at: http://www.larena.it/home/provincia-in-primo-piano/nell-archivio-degli-emo-c-è-la-storia-del-paese-1.2677381 (accessed on the 2nd of January 2023).

8. Available at: https://www.ilfattoquotidiano.it/2011/01/28/marysthell-leader-del-gruppo-olgettina-pupilla-di-silvio-e-donna-di-un-narcotrafficante/89025/ (accessed on the 2nd of January 2023).

9. Available at:
https://tg24.sky.it/cronaca/2012/01/07/cinesi_rapina_tor_pignattara_roma_caccia_killer.html (accessed on the 2nd of January 2023).

7. *[I] tedeschi soddisfano tutti i cliché del cattivo, a cominciare dall'accento tedesco.* (Germans fulfill all the clichés of the villain, starting with the German accent.; Raimondo, Mamma li crucchi, *Il Fatto Quotidiano*, the 6th of August 2012)[10].

8. *[G]li investigatori hanno accertato che il presunto capo dell'organizzazione, [...], si faceva chiamare 'Mimmo', ma aveva un chiaro accento nord africano.* (The investigators have ascertained that the alleged head of the organisation, […], went by the name of 'Mimmo', but he had a marked North African accent.; Iezzi, Market dell'eroina in riviera, quattro arresti, *Il Resto del Carlino*, the 13th of October 2012)[11].

9. *Puntuale e cordiale, con quel tipico accento tedesco [...] che lo caratterizza, riesce a mettere in riga con una parola i suoi collaboratori.* (Punctual and friendly, with that typical German accent [...] that characterises him, he can get his staff in line with one word.; Tieni, Ernst Knam, *Vanity Fair*, the 26th of April 2013)[12].

10. *[L]'assistente di volo si esprime in perfetto italiano, anche se viziato da un accento rumeno.* (The flight attendant expresses herself in perfect Italian, although spoiled by a Romanian accent.; Abia, Carpatair: cosa ha sbagliato Alitalia? *Giornalettismo*, the 4th of February 2013)[13].

11. *[I]l 48enne attira il genere femminile con quel suo sensuale accento francese e quegli atteggiamenti da uomo di altri tempi.* (The 48-year-old [actor] attracts the female gender with his sensual French accent and old-fashioned manly attitudes.; Tradimento da Oscar: le donne dicono no ai Vip belloni, *IMG Press*, the 15th of April 2014)[14].

12. *Anche esotico, nonostante l'accento tedesco sia così poco armonioso. Julian Draxler verso l'Italia?* (Also exotic, despite the fact that

---

10. Available at: https://www.ilfattoquotidiano.it/2012/08/06/mamma-li-crucchi/318031/ (accessed on the 2nd of January 2023).

11. Available at: https://www.ilrestodelcarlino.it/ascoli/cronaca/2012/10/13/786114-market-eroina-in-riviera-quattro-arresti.shtml (accessed on the 2nd of January 2023).

12. Available at: https://www.vanityfair.it/vanityfood/foodstar/13/04/26/knam-re-del-cioccolato-intervista-ricetta (accessed on the 2nd of January 2023).

13. Available at: https://www.giornalettismo.com/archives/748927/cosa-ha-sbagliato-alitalia/2 (accessed on the 2nd of January 2023).

14. Available at: https://www.imgpress.it/caffetteria/78381/ (accessed on the 2nd of January 2023).

the German accent is so unharmonious. Julian Draxler to Italy?; Di Marzio: "La Juventus e la scommessa Draxler: il gioco vale la candela? Così si presenta il tedesco", *JN24*, the 19th of July 2015)[15].

13. *[Q]uali sono gli accenti che fanno innamorare gli italiani? [I]l 38,5% degli intervistati italiani trova che l'accento francese sia il più attraente. L'accento spagnolo segue a ruota (32,3%) e quello inglese si aggiudica un bel terzo posto.* (What accents make Italians fall in love? 38.5% of Italian respondents find the French accent the most attractive. The Spanish accent follows closely (32.3%), and the English accent takes a nice third place.; La lingua dell'amore? L'italiano batte tutti, *ANSA*, the 16th of June 2015)[16].

14. *Sono anni che studio l'economia italiana [...] – ci dice con un elegante accento inglese – e posso dire che non c'è affatto da sorridere.* (I have been studying the Italian economy for years [...] - he tells us in an elegant English accent - and I can tell you that there is nothing to smile about at all.; ESCLUSIVA MP – Milan e Inter, si rischia grosso. L'allarme dell'esperto: "O si cambia o saranno guai entro 5 anni", *Mondo sportivo*, the 1st of April 2015)[17].

15. *Bell'aspetto, maturo, con un simpatico accento spagnolo: uno di cui fidarsi, almeno fino a prova contraria.* (Good-looking, mature, with a pleasant Spanish accent: one to trust, at least until proven otherwise.; Panelli, Cliente educato raggira la commessa. Paga con un biglietto falso da 500, *La Nazione*, the 27th of March 2016)[18].

16. *Sarà l'intramontabile fascino dell'accento francese, [...], fatto sta che è lei, [...], la più amata dagli utenti Bancalavoro.* (It may be the timeless charm of the French accent, the fact is that she is the most loved by Bancalavoro users.; *ItWaC*, 2016).

15. Available at: http://www.jn24.it/di-marzio-la-juventus-e-la-scommessa-draxler-il-gioco-vale-la-candela-cosi-si-presenta-il-tedesco/ (accessed on the 2nd of January 2023).
16. Available at: http://www.ansa.it/lifestyle/notizie/societa/nuove_abitudini/2015/06/16/la-lingua-dellamore-litaliano-batte-tutti_dce612bf-7c4c-445f-97a5-d4fc1e80b928.html (accessed on the 2nd of January 2023).
17. Available at: http://www.mondosportivo.it/2015/04/01/milan-e-inter-si-rischia-grosso-lallarme-dellesperto-o-si-cambia-o-saranno-guai-entro-5-anni/ (accessed on the 2nd of January 2023).
18. Available at: https://www.lanazione.it/empoli/cronaca/cliente-educato-raggira-la-commessa-paga-con-un-biglietto-falso-da-500-1.2014075 (accessed on the 2nd of January 2023).

17. *[I]l ragazzo dal viso angelico e dal fascinoso accento spagnolo.* (The boy with the angelic face and charming Spanish accent.; *ItWaC*, 2016).

18. *[S]ono andati a scuola di dizione e hanno perso il loro irritante e anacronistico accento russo.* (They took pronunciation classes, and they lost their irritating and anachronistic Russian accent.; *ItWaC*, 2016).

19. *[C]ol suo musicalissimo accento spagnolo, ci parla di sé e delle sue opere.* (In his very musical Spanish accent, he tells us about himself and his works.; *ItWaC*, 2016).

20. *[Q]uel suo roco accento tedesco che la fa assomigliare ad una Marlene dell'età dell'oro.* (Her hoarse German accent reminds us of Marlene in her golden years.; *ItWaC*, 2016).

21. *Parla un italiano vivacizzato dall'accento spagnolo.* (He speaks Italian enlivened by Spanish accent.; *ItWaC*, 2016).

22. *Schama parla con educato accento inglese, servendo il tè con le paste.* (Schama speaks with polite English accent, serving tea with pastries.; *CORIS - STAMPA*, 2017).

23. *[I] cattivi parlano con accento russo?* (Do bad guys speak with a Russian accent?; *CORIS – STAMPA*, 2017).

24. *[S]arà stato il suo accento straniero a rendere Ibrahim non degno della stessa attenzione e dello stesso aiuto, che noi crediamo sia dovuto ad ogni essere umano.* (It may have been his foreign accent that made Ibrahim unworthy of the same attention and help, which we believe is due to every human being.; *Consiglio della Municipalità 2*, Solidarietà per Ibrahim Manneh e per tutte le vittime della malasanità, Comune di Napoli, the 15th of September 2017)[19].

25. *Francamente non saprei dire se si trattasse di un accento russo anziché che serbo, sicuramente un uomo dell'Est.* (Frankly, I couldn't

---

19. Available at:
http://www.comune.napoli.it/flex/files/7/a/6/D.f3d5f42c22d5cdbab347/o.d.g._15.09.2017 malasanit.pdf (accessed on the 2nd of January 2023).

tell if it was a Russian accent rather than a Serbian one, but he was surely from Eastern Europe.; Caccia all'uomo, *Libero Quotidiano*, the 12th of April 2017)[20].

26. *I tre malviventi, secondo quanto appreso, avevano il volto travisato e l'accento dell'est Europa* (The three thugs reportedly had faces in disguise and spoke with from Eastern European accents.; In casa di un imprenditore edile, *Firenze Post*, the 27th of January 2018)[21].

27. *Sempre con lo stesso, affabilissimo tecnico: tuta e parlantina rapi-da, accento dell'est Europa, spiegazioni su un pezzo da cambiare, richiesta immediata di contanti per andarlo a comprare. Prima di sparire.* (Always with the same, very affable technician: suit and quick talk, Eastern European accent, explanation of a part to be changed, immediate request for cash to go buy it. Before disappear-ing.; *La Repubblica*, the 31st of March 2020)[22].

28. *I tre uomini, al momento ricercati, secondo le informazioni fornite dalla Martini alla polizia, avrebbero parlato con un accento dell'Est Europa.* (The three men, who are currently wanted, reportedly spoke with an Eastern European accent, according to information provided to police by Martini.; Milano, "legata e rapinata in casa": città sconvolta, chi è questa ragazza, *Libero Quotidiano*, the 13th of Janu-ary 2023)[23].

Attitudes towards foreign accents are arguably subjective, as they are shaped by personal beliefs, cultural background, and experience. However, accent bias is a form of linguistic discrimination and a harmful practice be-cause it leads to unjust treatment of people based on their pronunciation skills (Foucart et al., 2020). Such biases can limit personal and professional options and erect obstacles in a variety of situations, such as education, pro-fession, and social relationships, perpetuating societal disparities.

---

20. Available at: https://www.liberoquotidiano.it/news/italia/12356892/igor-russo-killer-budrio-vittima-faccia-ascia.html (accessed on the 2nd of January 2023).

21. Available at: https://www.firenzepost.it/accento-dellest/ (accessed on the 2nd of Jan-uary 2023).

22. Available at: https://ricerca.repubblica.it/repubblica/archivio/repubblica/2020/03/31/falsi-tecnici-per-la-truffa-dei-ricambiMilano09.html?ref=search (accessed on the 15th of January 2023).

23. Available at: https://www.liberoquotidiano.it/news/milano/34526603/milano-legata-rapinata-casa-citta-sconvolta-chi-questa-ragazza.html (accessed on the 15th of January 2023).

When browsing the online Italian media archives above – keeping in mind how anecdotal this type of evidence may be – the general idea imbued is that some accents are more prestigious than others. Some accents seem to be a source of fascination and charm, adding a touch of uniqueness and personality to a person's speech. Attributes such as 'refined', 'cultured' or 'elegant' are attached to the English accent (examples 4, 14, and 22). Likewise, the French-accented Italian is described as 'seductive', 'sensual', or 'fascinating' (examples 3, 11, 13, and 16). This is consistent with previous literature on the topic: the French accent is generally perceived more positively than other accents, like Russian. This concurs frequently but not necessarily with attitudes toward the respective nationalities (Rakić and Steffens, 2013; Roessel et al., 2019). Similar positive attributes are chosen to depict the Spanish accent, which is also considered 'melodious' (examples 15, 17, 19, and 21).

The excerpts above suggest that foreign accents are sometimes used as a way to identify the origin of people and to support certain stereotypes. For example, in the context of crime reporting, accents are used to indicate the supposed nationality or ethnicity of the criminals (examples 25, 26, 27, and 28), even though such assumptions can be inaccurate. The German accent appears to be perceived as 'disharmonious' and 'hoarse', transmitting the idea that the German speaker is 'tough', 'serious', or 'rigid' when she/he speaks Italian (examples 7, 9, 12, and 20). On the same note, the Russian accent as well as the media-constructed 'Eastern-Europe accent' and 'Slavic accent' are described as 'irritating' and it is implied that they are spoken by 'the bad guys' and the criminals (examples 18, 23, 25, 26, 27, and 28), with the accent being used as the main identifier. In example 10, a Romanian flight assistant is described as speaking 'spoiled Italian, implying that speaking with a Romanian accent is perceived as a negative characteristic. In general, from the data analysed, it appears that the media might use foreign accents to support existing stereotypes and perpetuate prejudices, rather than talking about accents to celebrate diversity. This is problematic as it can lead to an artificial sense of distance between people of different L1s who might otherwise have much in common.

The examples discussed above are not surprising. Dobrow and Gidney (1998: 115) reported similar trends in English-speaking countries. They found, for instance, that 'Germanic' and 'Slavic' accents are often used to give voice to 'bad' characters in cartoons, such as Bluestone's Ghost in Scooby Doo or the gangsters in The Adventures of Tintin. There are various other instances of this kind of attitudes in the English-dubbed cartoons. For instance, Boris, Natasha, and, the Fearless Leader – three 'bad' characters in Ward and Scott's (1959) *The Adventures of Rocky and Bullwinkle*

and Friends – speak in Russian-accented English; there seems to be a similar rationale behind the choice of what most likely seems to be a Russian accent for Peg-Leg Pete, the main antagonist in Disney's (1941) *Timber*; while the Big Bad Wolf in Disney's (1936) *Three Little Wolves* speaks English with a German accent, considering the historical moment, probably an attempt to satirise Hitler; also Dr. Doofenshmirtz in Povenmire and Hughes' (2007) cartoon series *Phineas and Ferb* also speaks German-accented English. All these examples used by Dobrow and Gidey's (1998) to support the claim that foreign accents used to voice animated villains might echo Americans' animosity towards nations against they fought war against in the past (e.g. World War I and II, the Cold War, etc.).

It is also worth mentioning Dobrow and Gidey's (1998) observation that another accent, although not foreign, often used to voice villains is the British accent. In fact, among the cartoons surveyed for the present study, I found that Disney's Maleficent in *Sleeping Beauty* (1959), Cruella de Vil in *The Hundred and One Dalmatians* (1961), Madam Mim in *The Sword in the Stone* (1963), Horned King in *The Black Cauldron* (1985), Shere Khan in *The Jungle Book* (1967), Scar in *The Lion King* (1994), and Judge Claude Frollo in *The Hunchback of Notre Dame* (1996) sounded British. At this point it may be useful to mention that, Bob Kennedy and Steven Weinberger, two linguists quoted in a USA Today newspaper article[24] claim that villains may also speak with a Mid-Atlantic accent (also known as 'Transatlantic accent'), a form of speaking English that uses a mixture of British and American English (Fox and Combley, 2014), essentially an artificial accent common among the Hollywood actors in the first part of the 20th century. Differently from the General American English, the Transatlantic accent displays, among others, the following features: absence of /æ/ tensing; absence of /i/ tensing in word-final position, that instead is generally produced as [ɪ]; absence of mergers before /l/; absence of /t/ and /d/ flapping; absence of palatalization (Skinner, 1990; Fletcher, 2005). The USA Today newspaper article mentioned provides some examples of villains speaking with a Mid-Atlantic accent: Disney's Evil Queen in *Snow White and the Seven Dwarfs* (1937), Lady Tremaine in *Cinderella* (1950), and Ursula in *Little Mermaid* (1989). Some other examples I found are Disney's Jafar in *Aladdin* (1992), Governor Ratcliffe in *Pocahontas* (1995), and Yzma in *The Emperor's New Groove* (2000). An interesting finding is that the accents of Disney villains do not lead back to their origins. For instance, even if *Aladdin* is set in the fictional city of Agrabah, presumably

---

24. Available at: https://eu.usatoday.com/story/life/entertainthis/2014/05/30/disney-maleficent-villains/77278712/ (accessed on the 1st of June 2018).

somewhere in the Middle East, Jafar is voiced with the Transatlantic accent. Similarly, Disney's (1999) Gaston in *Beauty and the Beast* does not speak English with a French accent, he speaks American English.

Although Count Dracula, the main character of Murdocca's animation films *Hotel Transylvania 1* (2012), *Hotel Transylvania 2* (2015), and *Hotel Transylvania 3* (2018), is not a villain, it is worth mentioning Adam Sandler's attempts to imitate the Romanian-accented English. The result is realistic because the actor produces several cues of Romanian accent (e.g. devoicing of final obstruent consonants; dentalization; interdental fricatives realised as stops, sometimes also with dentalization; postalveolar approximants realised as dental trills; vowel shortening; vowel backing; etc.). However, some traits of the more common Russian accent are present as well (e.g. palatalization of consonants before front vowels). In the Italian version of these films, Claudio Bisio voices Count Dracula with a likewise believable Romanian accent, especially in the version aired in 2012. Various features of Romanian-accented Italian are present (e.g. consonant degemination; open-mid vowels realised as close-mid vowels; alveolo-palatal lateral realised as [l+j]; etc.), but the imitation of the Romanian accent is sometimes stereotypical or displaying features of the Russian accent, instead (e.g. lengthening of the open central unrounded vowel; insertion of voiced palatal approximant before front vowels; and palatalization). This may be due to the overall confusion Italians have regarding these accents, and as I have already mentioned, this confusion also led to the emergence of the concept of *accento (dei paesi) dell'Est*, or *accento slavo*.

Interestingly, for a considerable period, the Italian-dubbed classical cartoons rarely captured the foreign accents featured in the original American versions. Only a handful of characters in Italian adaptations had a foreign accent. For instance, Dr. Doofenshmirtz in Povenmire and Hughes' (2007) *Phineas and Ferb* speaks German-accented Italian; while some characters in Bluth's (1986) *American Tail* – although not being villains – keep their accents also in the Italian version (e.g. the pigeon Henri has a French accent in Italian and the mouse Papa Mousekewitz has a Russian accent).

In order to delve deeper into this issue, I analysed 33 classical Disney cartoons[25], specifically focusing on the accents of the villains. In most cas-

---

25. The Disney cartoons analysed here were the following: *Alice Comedies* (1923-1927), *Three Little Pigs* (1933), *Snow White and the Seven Dwarfs* (1937), *Fantasia* (1940), *Pinocchio* (1940), *Cinderella* (1950), *Alice in Wonderland* (1951), *Peter Pan* (1953), *Lady and the Tramp* (1955), *Sleeping Beauty* (1959), *One Hundred and One Dalmatians* (1961), *The Sword in the Stone* (1963), *The Jungle Book* (1967), *The Aristocats* (1970), *Robin Hood* (1973), *The Rescuers* (1977), *The Fox and the Hound* (1981), *The Black Cauldron* (1985), *Basil The Great Mouse Detective* (1986), *Oliver and Company* (1988), *The Little Mermaid*

es, standard and, very rarely, regional varieties of Italian are used to voice the villains (e.g. Duke Weaselton, the small-time least weasel crook in Disney's (2016) *Zootopia*, is dubbed with a Casertan accent by Francesco Matano). One possible explanation for this observation is that in the United States, culturally and historically, there has been a tendency to associate specific foreign accents, such as the Russian accent during the Cold War era, with villains. However, in the Italian context, these associations may not be relevant, as they do not reflect Italian beliefs. For this reason and to establish cultural relatability and resonance, it may seem advantageous to voice characters with regional accents in the dubbed version of American films. Moreover, the 1950's marked the starting point of a standardization process for the Italian language, hence, there has been an evident attempt to provide the target public – often children and young people – with established pronunciation models.

There are, however, several recent films, that are dubbed with foreign-accented Italian, such as Attal's (2009) *Le Concert*. Almost all voices in this film are Russian-accented. This choice generated mixed feelings among the public, suggesting that the Italian audience might not yet be used to listen to such long samples of foreign-accented speech:

- *L'unica cosa che mi ha disturbato è stato il doppiaggio con accento russo lungo tutto il film. Cosa che lo rende pesantissimo ed a volte poco intellegibile. Bisogna fare uno sforzo per capire certi passaggi, soprattutto dove la recitazione si fa più concitata. Non vorrei passare per il criticone di turno, ma dove si è mai visto che i film americani vengono doppiati con accento americano, o quelli francesi con accento francese?* (The only thing that bothered me was the Russian-accented dubbing throughout the film. Which makes it very heavy and sometimes unintelligible. One has to make an effort to understand certain passages, especially where the acting becomes more concise. I don't want to come across as the usual critic, but where on Earth have you seen American films dubbed with an American accent, or French films dubbed with a French accent?; *Mymovies*, the 30th of January 2011, online[26]).

(1989), *The Rescuers Down Under* (1990), *Beauty and the Beast* (1991), *Aladdin* (1992), *The Lion King* (1994), *Pocahontas* (1995), *Toy Story* (1995), *The Hunchback of Notre Dame* (1996), *Hercules* (1997), *A Bug's Life* (1998), *Mulan* (1998), *The Emperor's New Groove* (2000), *The Princess and the Frog* (2009), *Tangled* (2010), *Wreck-It Ralph* (2012).

26. Available at: https://www.mymovies.it/film/2009/ilconcerto/pubblico/ (accessed on the 30th of January 2023).

- *Fastidiosissimo il ridicolo accento russo che il doppiaggio italiano ritiene di affibbiare a tutti gli orchestrali e che rischia seriamente di trasformare il film in una farsa.* (Most annoying is the ridiculous Russian accent that the Italian dubbing deems to affix to all the members of the orchestra, which seriously risks turning the film into a farce.; *IBS – recensioni*, the 30th of August 2010, online[27]).

As mentioned above, some foreign accents have been reported to be perceived as having a lower status or credibility, and they may even be stigmatized. In fact, Bianchi and Calamai (2012) and Calamai (2015) found that, for instance, the Romanian accent was the most stigmatized accent in the verbal-guise experiments they conducted. The examples I extracted from the web and corpora also seem to confirm their findings. For example, in one instance the Romanian accent is described as an 'alterator' of the Italian language (example 10). Additionally, the Romanian accent is most of the times put into the hotchpotch of the concept of *accento (dei paesi) dell'Est*, or *accento slavo* – three common phrases in the Italian media[28] – even if it has very few or no features in common to other accents sharing category built by the media (examples 25, 26, 27, and 28 above). Similar attitudes are reported towards the so-called *accento nordafricano*, which, however, is a less frequent phrase in the corpora queried here[29].

However, formulating a theory able to account for the reasons why some foreign accents are stigmatized is challenging. More realistically, one could assume that these tendencies are country-dependent, and as a result, the unpopularity of some accents may be due to ethnic and socio-cultural animosity. In fact, Chakraborty et al. (2017) state that accent bias may be a consequence of ethnocentrism. One may recall, for instance, Victor Hugo's (1859: 93) words on Louis Bonaparte: "[he] read, with a foreign accent which was displeasing, a speech", or Thomas Babington Macaulay's (1974: 243) account of one of the protagonists of his historical essays: "his foreign accent and his foreign attachments were offensive to the national prejudic-

27. Available at: https://www.ibs.it/concerto-film-radu-mihaileanu/e/8032807032184 (accessed on the 30th of January 2023).
28. For *accento slavo* and *accento dell'Est* there are 25 concordances of in the 1,909,826,324 tokens ItWaC corpus (Baroni et al., 2009); 15 concordances in the 380,823,725 tokens Repubblica corpus (Baroni et al., 2004); 2 concordances in the 38,000,000 tokens CORIS-STAMPA subcorpus (Rossini Favretti et al., 2002); while an exact-match web search returns 16,809 hits for *accent* dell'Est*, 539 hits for *accento dei paesi dell'Est*, and 4,058 hits for *accent* slav**.
29. For *accento nordafricano* there are no concordances in ItWaC (Baroni et al., 2009) nor in the CORIS-STAMPA subcorpus (Rossini Favretti et al., 2002); while an exact-match web search returns 706 hits for this syntagma.

es". Therefore, the perception of foreign accents in a certain language is often sifted through the listener's preconceived stereotypes, many of which are likely to be shared with other native speakers of that language. Cultural prejudice or racism towards people of certain origins have been historically attested (Gabriel, 1994; Lipsitz, 2011). It is plausible to think that these attitudes might as well have affected the way in which the observer perceives the foreign accent of the people that she/he hears and thus discriminates. Consequently, one could infer a certain circularity in which foreign accents serve as (re-)activators of prejudice. In the examples provided above, most contexts of *accento slavo*, *accento (dei paesi) dell'Est*, or *accento nordafricano* regarded crime news. These frequent associations to which Italian people are exposed through media induce overgeneralizations that could easily shape the listeners' perception of these foreign accents. This might eventually lead to stigma, prejudice, and discrimination. People speaking L2 Italian might therefore be judged or treated unfairly based on their foreign accent rather than their merits and qualities. As a result, communication, socializing, and even job and school prospects may be hampered. Recognising accent biases and actively looking for different resources to overcome it is essential nowadays. Media makers and content providers have the duty to show objective portrayals of the various cultures and accents in Italy and to refrain from fostering damaging stereotypes.

A different type of attitude, although still related to accent bias is discussed in Lev-Ari and Keysar's (2010) study, where the authors show that the native speakers' difficulty to process foreign accent results in non-native speaker being perceived as less credible. They found that statements such as 'Ants don't sleep' were considered less true when uttered by non-native speakers of English. Even if more studies should focus on this under-explored topic, Lev-Ari and Keysar's (2010: 1096) conclusion is a thought-provoking one "[a lower] credibility may have an insidious impact on millions of people, who routinely communicate in a language which is not their native tongue". These tendencies, however, do not apply to all geographical contexts. In fact, Stocker (2016) claims that the study she conducted in Switzerland – a country with four official languages – reveals a different trend, namely that foreign accent does not affect credibility, even if raters had prejudice with respect to the credibility of some speech communities. Likewise, regarding Chinese-accented Italian, the study of De Meo et al. (2011) on Chinese advanced learners of Italian suggests that there is no effect of foreign accent on credibility. Still, they find that suprasegmental features are linked to credibility ratings, in the sense that a reduced tonal range and longer silent pauses lead to a significant increase in listener trust. Regarding the persuasiveness levels of Chinese-, French-, and

Russian-accented Italian, Vitale et al. (2012) test the reactions of native speakers of Italian during a perceptual experiment and the results show that native speakers are more persuasive than non-native speakers. On the same note, in their attempt to understand why some accents are stigmatized, Gluszek and Dovidio (2010) propose a framework that includes the listeners' perception (i.e. attitudes and stereotypes) and the speakers' perspective (i.e. beliefs and communication challenges), as well as the dynamics involved in communication (i.e. online processes and interaction outcomes). The authors conclude that communicative, social, and contextual factors interact altogether and that speaking with a foreign accent is associated with a sense of not belonging.

Engaging in discussions about accent bias holds significant societal value, as people who speak the L2 may feel that their accent is a heavy burden to bear due to the pervasive stigma and prejudice that exist in society. This can show up in a number of ways, such as a reluctance to use the L2 at all or feelings of insecurity or shame while speaking in public. Speakers may experience alienation and lack of confidence in their language abilities if accents are framed as indicators of inferiority, criminality, or more general, foreignness. This can perpetuate the wrong belief that intelligence or communication skills are somehow reflected in the accent. Additionally, as seen in the examples taken from media and corpora, foreign accent may be a stigma trigger and amplifier. Therefore, it is essential to acknowledge and fight the stigmatization of accents in our society and to encourage a more tolerant and inclusive approach to language diversity. Italian media has an important role in this respect because they can promote a more equal society that encourages L2 speakers to feel more confident when using their language abilities. Eventually this can contribute to creating an environment where L2 speakers feel more welcome in Italy.

2.1.2 Towards a definition of foreign accent

In the first months of life, able-bodied infants have the potential to perceive, discriminate, and learn the sounds and prosody of all languages (Eimas et al., 1971; Grieser and Kuhl, 1989). However, as children reach one year of age, they become aware of the speech contrasts that are phonologically relevant in their L1s and begin to ignore the rest (Werker et al., 1981). Some scholars have suggested that achieving native-like pronunciation in a foreign language becomes increasingly difficult with age. Rejecting the idea that a loss of neural plasticity and the neurofunctional reorganization have an effect on foreign accent, and following Moyer (2013: 12), I believe that this age-related disparity occurs essentially due to the fact that L2

learners, unlike L1 learners, start the acquisition process with a pre-existing foundation in their L1 phonology, which can serve as a basis for metalinguistic awareness and also cause potential interference for any L2 learned.

More recent research, such as Flege and Wayland (2019), criticise Lenneberg's (1967) claim that neural maturation could block the process of L2 speech learning, pointing out that instead, speakers cannot prevent the interaction between their L1 and L2 phonetic subsystems. At the same time, not all late learners manage to master the language at all levels, some prefer or are forced to remain grounded in culture of their home country and therefore will "continue to use other languages in their daily lives" (Gut, 2007: 75), an aspect that might slow down the process of achieving native-like pronunciation. Additionally, several studies have pointed out that with the passing of time, the late learners' sensitivity towards subtle phonetic cues of a second language is not entirely lost, but rather it is 'obscured' (Piske, 2008). Moreover, older learners seem to learn phonological patterns faster than young learners (Moyer, 2013). In the same strain, Flege (2018) claims that 'input', rather than age, is responsible for the way we sound when we speak a foreign language. That being said, speakers that do not sound native-like are said to have a 'foreign accent'.

According to most models presented in Chapter 1, at least in the first stages of learning a new language, learners employ phonological processes, pronunciation, and orthography norms of their L1 and those of other L2s they know, when speaking and reading the target L2. In fact, French learners of L2 Italian might find it hard to produce the alveolar trill [r] – as /r/ is not present in the phonemic inventory of their L1 – so they might, for instance, replace it with voiced uvular fricatives, [r] as [ʁ] in words such as *rosso*[30]. Native speakers of Italian can perceive this subtle segmental feature as deviating from the broadly accepted Italian pronunciation. More generally, native speakers of any given language understand if their language is spoken with a foreign accent. Accent-detection and accent-rating studies have shown that native speakers are able to recognise non-native pronunciations in a wide variety of speech samples, even in very short excerpts (Flege, 1984).

We all possess a personal comprehension of what a foreign accent entails and we can recognise non-native speech when we hear it. There have been several definitions and explanations for foreign accent and the exact circumstances that condition its presence or absence. The cues of foreign-accented speech have long been investigated both perceptually and instrumentally. In fact, various definitions of 'foreign accent' are built around the

---

30. Translation into English: 'red'.

concept of segmental and suprasegmental features that deviate from the pronunciation norms of the foreign language being acquired (Scovel, 1969; Gut, 2007).

One early definition of foreign accent is the one proposed by Jenner (1976: 167), namely "the complex of interlingual or idiosyncratic phonological, prosodic and paralinguistic systems which characterise a speaker of a foreign language as non-native". In the realm of non-native speech research, Flege (1981, 1984, 1988b) explores the concept of a foreign accent, defining it as an experiential perception by listeners. This perception stems from noticeable acoustic dissimilarities (and possibly visual cues) discerned in the pronunciation of sounds and other speech units, between individuals who are native speakers and those who are not. Similarly, Munro (1998: 139) refers to foreign accent in terms of a "non-pathological speech produced by second language learners that differs in partially systematic ways from the speech characteristics of native speakers of a given dialect".

Jilka (2000) shifts the perspective in the definition of 'foreign accent', by pointing out the fundamental relevance of the native speakers' perception:

[An] instance of foreign accent consists in a deviation from the generally accepted norm of pronunciation of a language that is reminiscent of another language, i.e. the speaker's native language. It has to be emphasised that such a deviation must be defined in terms of its perception by listeners who are native speakers of the respective language and not in terms of differences in articulation that may be instrumentally measurable. Only those deviations that are perceived as such can be considered instances of foreign accent (Jilka, 2000: 9).

In addition, following Markham's (1997) approach to non-native speech, Jilka (2000: 10) draws a distinction between two types of foreign accent. On the one hand, he claims that 'phonological foreign accent' only affects entire phonological categories and is attributed to cognitive limitations that result in incorrect or missing representation of phonemes in the second language. On the other hand, 'phonetic foreign accent' occurs when the learner has acquired the correct phonological representation of a category but has not yet developed the appropriate physical output routines. This results in incorrect productional representation of the phonetic realisation of a phonological category at the phonetic output level.

Jilka's (2000) and Munro's (1998) definitions are more inclusive since they take into account the native listeners and they contrast foreign accent to pathological speech and to regional accents. In fact, regarding Italian, one may argue that due to the multitude of established regional varieties, even native speakers fail to comply with the pronunciation norms of the 'standard' Italian. Defining 'standard' is, however, a challenging issue.

Building upon Crystal's (1987: 3) perspective, Moyer (2013: 86) argues that every individual has an accent; she challenges the notion of a neutral or standard accent that, however, remains deeply ingrained in the collective consciousness of speakers. She raises important questions about the origins of this idea and the reasons behind its enduring influence. One possible explanation is that a standard accent provides a benchmark for correctness, and it contributes to our understanding of social structure and our own position within it. On a similar note, Bertinetto and Loporcaro's (2005) study on the sound pattern of standard Italian points out that

from the [Florentine] dialect, Standard Italian inherited its phoneme inventory, but not all allophonic processes. […] Standard Italian is nowadays part of the active verbal repertoire of just a minority of educated people from Central Italy (especially Tuscany), besides being used by professional speakers or trained stage actors (the single idiolects spoken by these groups of people may, however, include sporadic features typical of Regional Italian) (Bertinetto and Loporcaro, 2005: 131-132)

Consequently, when defining non-native Italian speech, one should take all these aspects into consideration and diminish the emphasis on the notion of a 'standard' language variety or accent. Thus, a more encompassing definition of foreign-accented speech may be rendered in these terms:

A non-pathological manner of speaking a non-native language (L2) that deviates from the pronunciation that is typically regarded as native by the native (L1) speakers of that language.

2.2 Foreign accent in scholarly debate

This section is dedicated to a brief presentation of previous relevant research on foreign-accented speech. More detailed reviews can be found in Piske et al. (2001), Marotta and Boula de Mareüil (2010), Moyer (2013), and Combei and Marotta (2019). The focus is on the variables reported to have an impact on foreign accent, in an attempt to highlight the methodological complexity of recording and analysing non-native speech. The last part of this section will offer a state-of-art survey of studies on the perception of foreign-accented Italian.

2.2.1 Methodological considerations

Foreign accent has been extensively examined in theoretical literature, applied linguistics, and more recently, speech technology. Previous research has covered both production and perception of non-native speech, even if early studies concentrated mostly on the level of production (Selinker, 1972; Backman, 1979; Bernini, 1988, etc.). Due to the advent of technology, instrumental acoustic analysis is able to explore even fine cues of foreign accent (Cunningham, 2008; Mora et al., 2015, Zhi and Li, 2021).

In more recent times, a gamut of work has started to focus on the perception of foreign-accented speech, especially as far as L2 English is concerned (Flege et al., 1995; Munro and Derwing, 2001; MacKay et al., 2006; Levi et al., 2007; Piske, 2012; Atagi and Bent, 2016; Ahad et al., 2021, etc.). However, some studies have also explored other accents, such as foreign-accented German (Jilka, 1999), foreign-accented French (Vieru et al., 2011), foreign-accented Dutch (Pinget et al., 2014), and foreign-accented Italian (Marotta, 2008; Marotta and Boula de Mareuil, 2010; Pellegrino et al., 2015; Calamai et al., 2020; Nodari et al., 2021; etc.). Apart from the target language investigated, these and many other studies differ with respect to the speaker's L1. Research has been performed both on informants sharing the same L1 (Munro and Derwing, 1995; McKay et al., 2006) or on those having different linguistic backgrounds (Purcell and Suter, 1980; Flege, 2003; Calamai, 2015). In most cases, samples of non-native speech have also been compared with material produced by native speakers.

Studies on foreign accent have varied consistently with respect to the number of informants analysed, ranging from one participant (e.g. Hammarberg, 2001) – common for longitudinal or panel studies in which the researcher is generally interested in examining the speaker's progress – to hundreds of participants (Flege et al., 1995). These numbers are generally linked to the type of research conducted. Studies focusing on the global accent or on fluency tend to consider more speakers.

Regarding the elicitation techniques, various longitudinal studies (i.e. the material is collected from the same speakers in multiple instances over a period of time) and cross-sectional studies (i.e. the material is collected during a single recording session) have used read speech, either lists of (pseudo-)words (Elliott, 1995), or complex phrases or reading excerpts (Bongaerts et al., 1997; Frontera and Paone, 2018). In other cases, speech was elicited by means of the map task technique (Conklin et al., 2015), speakers were asked to simply talk about specific or free-choice topics (Byrne et al., 1998), or to describe images provided by the interviewer

(Thompson, 1991). Sometimes speakers were supposed to repeat pre-recorded speech (Neumeyer et al., 1998; Piske et al., 2001). This last technique, however, might not capture what Wolfson (1976: 189) calls 'natural speech', which is considered to be equivalent to 'appropriate' speech rather than 'unselfconscious' speech.

A lot of studies have, in fact, addressed the question of collecting spoken data from non-native speakers (Mayfield Tomokiyo and Burger, 1999), as it is generally considered challenging to record 'authentic' and 'natural' non-native speech. In fact, Mayfield Tomokiyo and Burger (1999: 7) say that the speech of "semi-fluent non-native speakers, whether they are at a real information desk or recording a contrived scenario […] will most likely be planned". Additionally, the type of elicitation technique may have an impact on the way listeners rate the degree of foreign accent in a speech sample. Oyama (1976) and Thompson (1991) report that read speech is perceived to be more strongly accented than spontaneous speech. Similar results are reported by Kolly and Dellwo (2013) on L2 German and L2 French. This fact is ascribed to the speaker' reading skills in an L2, since in most cases late learners (e.g. immigrants) do not receive education in the country they move. However, Patkowski (1990) claims that spontaneous speech may also pose challenges for two reasons: firstly, there is a higher likelihood that speakers produce lexical and morphosyntactic errors; secondly, speakers might tend to avoid using L2 words and sounds they are not familiar with.

As concerns the domains investigated, most research on foreign accent has addressed the production of segments. Additionally, in many cases, scholars are interested in investigating production issues linked to foreign-accented speech, while research on perception – L2 speaker's perceptual skills and the consequent effect on production, as well as perceived foreign accent – is far less common in the SLA literature.

One final methodological issue regards the tools and evaluation metrics used to investigate foreign accent. SLA research generally embraces one or a combination of two approaches when investigating foreign accent: the use of instrumental analyses on samples of non-native speech, in order to mark segmental and suprasegmental deviations with respect to native speech; and the elicitation of quantitative and qualitative judgments obtained from native speakers by means of accent perception experiments (e.g. recognising or rating samples of non-native speech; perceptual correlates of foreign accent; etc.). Considering the scope of this monograph, the methodological issues of acoustic-phonetic analysis will not be discussed; reviews and examples of good practice are presented in Granlund et al. (2011), Jiang (2018), Bartelds et al. (2020, 2021), and Koffi (2021).

Instead, greater attention is given here to some techniques and metrics borrowed from perceptual dialectology and folk linguistics to examine attitudes regarding foreign accents. Perceptual dialectology and folk linguistics are two related fields of research that focus on the beliefs and attitudes of non-linguists about language varieties, dialects, and accents. According to Preston (1999a, 199b), perceptual dialectology investigates what ordinary people believe about the distribution of language varieties in their own and surrounding speech communities and how they implement these beliefs. It is noteworthy that the beliefs and attitudes held by ordinary individuals towards language and accents can significantly diverge from the discourse among expert linguists. Preston (2021) takes a broad approach to folk linguistics, including not only the comments that non-linguists make about linguistic topics but also attitudes and reactions they have to language varieties and language use, including both overt and subconscious responses. Several studies surveyed in this book have used approaches that derive from perceptual dialectology and folk linguistics to investigate the ways in which L1 speakers use their knowledge to recognise non-native speech, to assess the overall degree of perceived foreign accent, to indicate deviant segmental and suprasegmental features, and to discriminate among multiple varieties of foreign-accented speech. Generally, in accent perception experiments listeners – expert or naïve native raters, as well as non-native raters (even if this group is not very common) – perform one or more of the aforementioned tasks.

Regarding the evaluation of 'accentedness', most studies employ rating scales that usually have at one end a value corresponding to 'no foreign accent', 'native speaker', or 'native-like accent', whereas at the other end a value labelled as 'heavy foreign accent' or 'strong foreign accent'. In fact, Southwood and Flege's (1999) experimental evidence shows that the degree of foreign accent can be measured on equal-appearing intervals, suggesting that a 9-point scale should be employed. However, the range of the values between these two ends varies greatly in studies where the degree of foreign accent is examined. In some cases, scales have very few values, for instance, a 3-point scale (Tahta et al., 1981), while in other cases, such as Flege et al. (1995), a 255-point sliding scale was used. According to Piske et al. (2001), the 5-point scale is the most used.

Other methods of investigating perceived foreign accent are accent identification experiments (Vieru and Boula de Mareüil, 2006); during this type of task, raters are generally asked to label L2 speech samples with the speaker's presumed L1; in some cases, a set of possible languages are provided in the experimental setting, while in other cases they are not. Additionally, listeners may also be asked to provide subjective evaluations of

foreignism feature saliency. In this case, authentic or synthetic speech samples (i.e. with altered prosody or segments) may be provided in order to trigger judgments regarding the features the experimenter is interested in (e.g. Marotta and Boula de Mareüil, 2010).

Finally, it is worth mentioning that various studies on perceived foreign accent have also used simulated speech, elicited with the 'matched-guise' technique, consisting in samples recorded by professional actors imitating specific accents (e.g. Johansson, 1978). In fact, Giles and Bourhis (1976) and Gluszek and Dovidio (2010) claim that this technique is used to avoid the effect of speaker-specific variables (e.g., pitch) on the results.

2.2.2 Factors influencing foreign accent

The results of instrumental and perceptual experiments are often interpreted as a function of the variables that are assumed to affect the L2 pronunciation. This section will present the factors that were most frequently reported in empirical-based research: age of L2 onset, input (comprising the quality and quantity of L2 exposure and training, and the amount of L2 use), gender, and sociopsychological factors. Various studies have focused on this topic (Purcell and Suter, 1980; Flege et al., 1995; Piske et al., 2001; Moyer, 2013) and may, therefore, offer additional insights.

Undoubtedly, among all, the age of onset – the starting point of the second language acquisition process – has been the most investigated variable in the literature. Depending on the type of acquisition – spontaneous or guided – this concept may be rendered in terms of 'age of arrival', 'age of first exposure', or simply 'age of learning'. Moyer (2013) dedicates a whole chapter to survey the SLA literature investigating the effects of age on L2 phonology; in her study she provides arguments for and against the age of onset theory. Chapter 1 introduced Lenneberg's (1967) CPH, which suggests that there is a rather limited period of time (i.e. from age 2 until puberty – presumably until 9-13 years old – a time-frame corresponding roughly to the lateralization process and the neural maturation) for the accurate acquisition of L2 phonology. However, as already discussed, a large body of empirical research does not support Lenneberg's (1967) CPH. There are some examples in which younger learners outperformed older ones in pronunciation and oral communication tasks (Fathman, 1975; Munro and Mann, 2005; Combei et al., 2020, etc.), but there is no model based on empirical evidence predicting that "before age X, a person is guaranteed to develop a native accent and, after age Y, a foreign accent is unavoidable" (Munro and Mann, 2005: 337). Similarly, as Piske et al. (2001) explain, if the predictions of CPH were true, there should be a significative variance

between the degree of perceived foreign accent of the speakers that started to learn the L2 before the end of CP and those that started after it. However, this is not the case, since various empirical studies (Bongaerts et al., 1997; Moyer, 1999; etc.) have shown that even speakers that started to learn the L2 after puberty managed to attain native-like pronunciation. At the same time, some studies claim that there are detectable cues of foreign accent in speakers whose age of learning is early childhood (Thompson, 1991; Flege et al., 1997; etc.). In the same strain, Piske et al. (2001: 197) say that factors beyond the age of acquisition also play a role in determining the level of foreign accent in a speaker's L2.

The impact of age of onset on L2 pronunciation is complex to analyse since it is often linked with other variables, such as the individual's age, duration of stay in an L2-speaking setting, and frequency of L1 and L2 usage, which may also affect their accent. In fact, Moyer (2013: 62) combines age to sociopsychological factors and claims that adolescents and adults are highly conscious of their ego boundaries, whereas children are uninhibited and enjoy playing with language, imitating new sounds freely. She suggests that this difference could be the reason why children are more easily capable of sounding like native speakers, while adults overcome their foreign accent more rarely.

All in all, I believe that even if an early age of onset might be desirable, it cannot be viewed as the only factor responsible for not speaking in a native-like manner. More realistically, and following Flege's work (1995, 2018), I consider that input affects L2 pronunciation the most. Various attempts of defining 'input' have been reported in the SLA community, starting with Krashen and Terrell's (1983) Input Hypothesis, which states that learners can acquire language only by means of 'comprehensible input'. Flege's (2008: 175) provides a definition of 'input' that is more specific for L2 speech, namely all "L2 vocal utterances the learner has heard and comprehended, including his own, regardless of whether these utterances have been produced correctly […] or incorrectly". Drawing upon the definition above, the concept of input, as it is used in this monograph, is the following:

A composite variable englobing the quality and quantity of L2 exposure and use, and potentially but not necessarily, length of stay in the L2-speaking country, as well as specific L2 (pronunciation) training.

Studies on the impact of input on foreign accent are not common in SLA literature (Moyer, 2008; Flege, 2008; Flege, 2018; Flege and Wayland, 2019; Combei et al., 2020). Moreover, quantifying the overall effect of the variables that define the concept of input poses methodological chal-

lenges. For this purpose, Flege (2008) proposes Csikszentmihalyi and Larson's (1987) Experimental Sampling Method to measure both the quality and the quantity of L2 input. The use of this technique implies a more detailed assessment of the speaker's L2 use. The factors that input incorporates have been reported individually in literature as predictors of foreign accent. The length of stay in the country where the L2 is the prevalent language, the L2 use in general and the ratio of L1-L2 use in particular were examined in various studies. In some cases, the length of stay is associated with L2 experience, even if this is not necessarily true (e.g. poorly integrated immigrants – people that might have a social network consisting mainly of other immigrants with the same origin – that do not use L2 even if they have a long residency in the foster country). This variable was reported to have an impact on the overall degree of foreign accentedness in Flege et al. (1995). Moreover, Purcell and Suter (1980) found that a composite variable consisting in two predictors – length of stay and sharing house with L1 speakers – is useful in explaining the variability of L2 speakers' pronunciation accuracy. Similarly, Piske et al. (2001) claim that L1 use had a significant effect on the degree of perceived foreign accent; in fact, speakers that used their L1 during their daily lives received ratings that suggested they had a stronger foreign accent than those who used L2 more often than L1.

Finally, the role of L2 (pronunciation) training on accentedness has been investigated in a limited number of studies. However, according to Piske et al.'s (2001) review of the studies that assessed the factors affecting foreign accent, it appears that formal L2 instruction does not to have a significant effect on pronunciation skills. This might be due to the fact that language teachers do not give enough consideration to pronunciation skills during language classes. In addition, I believe that the role of pronunciation instruction on the degree of foreign accent was difficult to explore in previous studies because the effect of this variable may interact with other factors: inappropriate training, either because the instructor's approach is not tailored on the learner's needs or simply because the quality and the quantity of the training is inadequate; methodological issues, either due to corpus sampling (i.e. generally, in a classroom setting, there is plenty of individual variation among learners), or due to the metrics used to assess pronunciation.

Regarding gender, there is contrasting evidence in the literature. Some studies (Suter, 1976; Thompson, 1991; Polat and Mahalingappa, 2010) report a significant role of this factor, in the sense that, generally, women have a better pronunciation accuracy. However, a study by Conceição Silva and Almeida Barbosa (2017) claims that in an accent perception experiment, Brazilian male speakers were judged as having less marked accents

in L2 Spanish than female speakers. However, in many other studies, gender is not reported as a predictor of the degree of foreign-accented speech (Elliott, 1995). Additionally, as Moyer (2013) suggests, even if some studies report gender differences, they may be confounded by other factors, such as age or immersion style. In fact, studies on adolescents (Davies 2004; Heinzmann, 2009; etc.) report different performance between male and female speakers, but these differences may be influenced by psychological factors, such as anxiety, a common condition among adolescents. Finally, in case the L2 immersion style is immigration, for instance in adult couples, male speakers might display a better performance, since they are more integrated than their spouses and therefore have a better input (Moyer, 2013).

Besides the age of L2 onset and input, other factors affecting L2 pronunciation were mentioned in empirical studies. Some are the speaker's motivation; her/his attitude towards the L2, the country where the L2 is the prevalent language, and its inhabitants; and personality. However, the role of motivation and attitude on the degree of perceived foreign accent has only been systematically investigated in the last two decades (Levis and Moyer, 2014). In fact, in SLA research there is still no standard measurement meant to quantify attitude and motivation, or the speaker's reliability when it comes to self-assessing these two factors. Anecdotal evidence suggests that instrumental motivation (i.e. motivation for well-paid jobs or higher grades), as well as integrative motivation (i.e. a better integration in the foster country and the desire to connect with its community) could make L2 speakers become more aware of their pronunciation. For instance, Purcell and Suter (1980) say that integrative motivation was a significant predictor for pronunciation accuracy, while Bongaerts et al. (1997) report a similar result for instrumental motivation. However, evaluating the contribution of the L2 speaker's attitudes – an umbrella concept incorporating to speaker's mental outlook on L2 learning and L2 in general, as well as her/his sentiment for the foster country and for its inhabitants, and the desire to be and sound like them – is a complex task. In fact, this may be the reason why very few studies have addressed this issue. For example, Elliot (1995) and Moyer (2007) report a positive correlation between favourable accent ratings and the desire to sound native. But other attitudes have not been consistently explored. One example of such study is Thompson (1991), where no significant effect on accent rating scores ia reported for L2 speaker's attitudes. And even if Moyer's studies (2004, 2007) suggest that attitudes towards linguistic and cultural assimilation in the L2, as well as the intention to reside for a long time in the foster country correlate with accent, in one of her subsequent studies she says that "[t]here is little doubt that atti-

tudes play some role in phonological attainment, however, the directness of that relationship is uncertain" (Moyer, 2013: 70). Similarly, very few studies have explored the role that the speaker's identity has on the degree of perceived foreign accent. Based on anecdotal evidence (Marx, 2002; Miller, 2003), it is generally considered that a strong sense of belonging to the native country would lead to poor pronunciation skills in L2, but systematic studies are needed. The role of other sociopsychological factors (e.g. personality, social cognition, self-concept; etc.) has not been not frequently tested in studies on foreign-accented speech. However, in one such study, Zárate-Sández (2017) claims that personality explais a considerable portion of the variance in the degree of perceived foreign accent, more specifically, extraversion and neuroticism are significant predictors of foreign accent.

2.2.3 The perception of foreign-accented Italian

By the end of the 1990's most empirical research on the perceived foreign accent had been conducted on English. In more recent time, however, various studies on non-native Italian speech have been published. In this subsection I am presenting an up-to-date review of over 20 empirical studies that have treated directly or at least to some extent the perception of foreign-accented Italian[31].

A limited number of studies investigated how segmental and suprasegmental features of L2 speech affect the perception of foreign-accented Italian. One of the first is Boula de Mareüil et al.'s (2004) work on Spanish, French, and Italian accents. After having processed and investigated the samples by means of the prosody transplantation technique and accent perception experiments delivered to native speakers of French, Italian and Spanish, the authors report that prosody plays a greater role than the segments in the identification of the Spanish and Italian language and accent. Starting from this work, Marotta and her colleagues published other studies on the perceived foreign-accented Italian. While stressing the important role of prosody on the perception of foreign accent and reviewing studies on non-native speech in general, and Italian in particular, Marotta (2008) also explains in detail the methodology and results in Boula de Mareüil et al. (2004). On a similar note, while referring to studies on L2 speech, Marotta (2009), argues that perception has not received much attention and

---

31. Considering the scope of this monograph, only studies that examine the Italian listeners' perception of foreign accent will be presented. Theoretical research as well as purely acoustical studies on non-native Italian speech will not be discussed.

that the impact of prosody in both learning a second language and perceiving a non-native accent has been largely overlooked. In the same paper, Marotta (2009) presents the results of a pilot study consisting in an accent identification experiment. The speech samples were produced by L2 speakers of Italian whose mother tongues were either French, Spanish, German, or English. On the one hand, the author reports that, in most cases, listeners (i.e. native speakers of Italian) were able to perceive foreign accent even in samples produced by experienced L2 speakers, while, on the other hand identifying the correct accent was a difficult task. Only the samples of English-accented speech were consistently labelled as such. Moreover, it appeared that in the samples of read speech, the accent was more easily recognisable than in those of spontaneous speech.

Starting from these pilot studies, Marotta and Boula de Mareüil (2010) perform a more complex perceptual experiment. They record read and spontaneous speech from 10 female speakers, two for each of the following L1s: Italian, French, Spanish, English, and German. In an accent perception experiment they ask 127 listeners (native speakers of Italian) to identify the speakers' L1 and to rate their degree of foreign accent. The authors report that, overall, native speakers of Italian, as well as English, and German speakers are identified more accurately than French and Spanish speakers, suggesting that this might be due to the vicinity of Italian with the two Romance languages. Regarding the degree of perceived foreign accent, German speakers are labelled as having a stronger accent than all the other informants, both in samples of read and spontaneous speech. The quantitative analyses suggest that there are statistically significant differences between the correct accent identification in read and spontaneous speech, meaning that in the former the foreign accent is identified more accurately than in the latter. Finally, the authors test the effect of the listeners' familiarity with the foreign accents investigated and the accuracy of accent identification, but they reported a weak correlation.

Devis (2005) assesses the perception of Spanish and Catalan-accented Italian in an experiment that consists not only in an identification task, but also in signalling segmental features of foreign-accented speech. She records read and spontaneous speech from four experienced speakers of L2 Italian: two monolingual L1 Spanish speakers, and two bilingual Spanish-Catalan speakers. The material is rated by 26 native speakers of Italian from three different regions (Emilia-Romagna, Calabria, and Tuscany) that are instructed to identify segmental deviations. The author reports that Spanish-accented Italian, due to its cues (e.g. vowel lowering, absence of consonant gemination; rolled alveolar trills; etc.), is more recognisable than Catalan-accented Italian.

Pellegrino (2012a) also tests the role of suprasegmental and segmental features on the perception of foreign-accented Italian. The degree of foreign accent of 10 speakers (2 native speakers of Italian, and 8 Chinese learners of L2 Italian) is assessed by 56 Italian listeners. At the same time, the author performs a spectrum-acoustic analysis on the speech samples, by calculating the following features: vowels and syllables duration, the length of stressed open and closed syllables; articulation and speech rate, fluency, tonal range, percentage of silence and mean duration of empty pauses. In addition, she examined the syllable composition and pronunciation errors. Based on the results of the accent perception experiment and the data from instrumental analyses, Pellegrino (2012a) concludes that both segmental and suprasegmental cues influence the degree of perceived Chinese-accented Italian.

In another study on 16 Chinese learners and a control group of 2 native speakers of Italian, Pellegrino (2012b) examines the effect of suprasegmental features on the degree of foreign accent judged by 56 Italian listeners, and the reaction times of other 20 Italian listeners during an accent identification task. The results of instrumental and perceptual analyses show that the features that contributed most to the degree of perceived of foreign accent were the percentage of speech, the number of silent pauses, fluency, and tonal range. An increase in tonal range, fluency, and speech rate correlates to a milder foreign accent. Regarding the reaction times, the author reports an average of 5 seconds for the identification the native speakers and over 10 seconds for non-native speakers. Finally, tonal range, articulation rate, and speech rate seem to have an effect on accent identification as well.

Similarly, Vitale et al. (2014, 2017) assess the prosody in yes/no questions and statements produced by 12 Chinese learners and 4 native speakers of Italian (females, aged 24 on average). Instrumental analyses are performed on 224 utterances (questions and statements divided evenly), by extracting mean values for pitch, the number and duration of inter-pause speech intervals, the number of syllables per inter-pause speech interval, the duration of silent pauses and disfluencies, the percentage of disfluencies, articulation rate, phonation rate, tonal range, fundamental frequency (f0) at the midpoint of each of the last three syllables in each sentence. Authors claim that prosody may account for differences between elementary, intermediate, and advanced Chinese learners of Italian. Similarly, the perceptual experiment – using prosody transplantation – conducted on 40 Italian listeners reveals that non-native segments with a native Italian prosody are judged as less accented than native Italian segments with a non-native

prosody. In addition, the importance of prosody is confirmed in discriminating between questions and statements.

Pellegrino et al. (2015) also test the Italian listener' discrimination abilities. In this study, the task consists in identifying foreign-accented and pathological speech in samples of read material produced by Italian and Ukrainian deaf speakers, as well as by hearing-abled Japanese and Ukrainian L2 speakers. The 17 native Italian listeners judge the samples in terms of 'accentedness', 'comprehensibility' and 'speech type'. It seems that listeners have difficulties in distinguishing between native and non-native deaf speech. In addition, the authors claim that due to the segmental and suprasegmental cues characterising the Japanese informant's speech, listeners thought she was a deaf speaker. Still, when examining instrumentally her vowel space, as well as the consonantal deviations her speech is different from that of the deaf speakers.

Pellegrino and Dellwo (2015) examine the role of rhythmic cues, from the amplitude envelope (ENV), on the perception of non-native Italian. The authors ask 10 native speakers of Italian to judge the native-likeness in authentic utterances produced by a German speaker of L2 Italian and by an Italian L1 speaker, and in synthetically manipulated utterances with either the ENV of a German speaker of L2 Italian or that of a native Italian speaker. The results show that rhythmic features in the speech ENV impact the listeners' perception of foreignness.

For their contrastive studies on L1 Italian and L2 Italian of Chinese, Albanian, Romanian, and Polish speakers, Romito and Tarasi (2012) as well as Romito et al. (2012) conduct a perceptual experiment aimed at exploring which features are perceived as more salient by 58 Italian listeners. The results reveal that the features that were perceived as most deviant from L1 Italian were those at the segmental level, especially the absence of consonant gemination, reported for all L2 speakers, while the prosodic features mentioned as more salient were (in this order) intonation, stress, and rhythm.

Frontera and Paone (2018) investigate the relationship between the motivation of learning L2 Italian and the prosodic accuracy in this language. Their corpus is made up of 12 speakers: 4 Romanians, 4 Arabs, and 4 Italians. The L2 speakers have either moderate or high motivational indexes, half of them are inexperienced while the other half are experienced speakers – indexed by length of residence (LOR). The speakers' prosodic accuracy is assessed on read declarative sentences, through acoustic measurements and perception tests. The authors claim that there might be a possible correlation between prosodic accuracy, motivation indexes and LOR levels. However, based on the native speakers' perceptual judgments, no signifi-

cant correlation is reported between high motivation rates and more native-like performances; even if, motivation seems to determine a weaker perceived foreign accent when L1 and LOR are similar. The values for Romanian prosodic cues (i.e. articulation rate and tonal pitch range) are comparable to those reported for Italian speakers, independently of the motivational index or the LOR. Regarding the Arabs, a high motivational level generally means a better prosodic accuracy. However, the LOR does not explain divergences in non-native productions.

De Meo et al. (2015) also investigated the acoustic cues of non-native speech. They ask Chinese, 'Slavic', and Italian listeners to discriminate between native and non-native speech and to signal the most salient cues of foreign accent. The samples are produced by six female speakers whose mother tongues were either Arabic, Chinese, French, Japanese, Russian and Vietnamese. A control group consisting in 3 Italian speakers is included as well. The authors report that the intonation is the most relevant feature for discriminating between native and non-native speech, followed by the articulatory quality. Finally, it appears that the 'native accent' label depends on the listeners' and speakers' L1: the Arabic accent is easily recognised by native speakers of Italian; Chinese listeners tend to assign 'native accent' label more easily than other listeners; while on the contrary, Slavic listeners assign it the least frequently.

Mori and Barkat-Defradas (2005) verify the role of VOT on Moroccan speakers of L2 Italian. They perform acoustic analyses but no perceptual experiment, claiming that a long lag VOT for /k/ and /t/ may be considered an indicator of the Moroccan accent. Segmental traits (both vowels and consonants) of the L2 speech produced by Moroccan informants are analysed also in Mori (2007). The author reports that non-native consonants are more salient perceptually and they contribute to the perception of foreign accent.

In a different type of study, Vitale et al. (2012) test the reactions of native speakers of Italian in a perceptual experiment aimed primarily at evaluating the levels of persuasiveness of non-native speech. While listening to 8 radio spots recorded by native speakers of Italian, Chinese, French, and Russian, 164 listeners indicate whether the speakers are native nor not, they assess the degree of foreign accent and persuasiveness of each voice, and they report the features (i.e. articulatory accuracy, intonation, speech rate, pauses, or other) that influence the rating. The results show that five out of six L2 speakers are rated as foreign-accented by almost all listeners. The accents of most speakers are judged as 'strong'. As far as the persuasiveness is concerned, only the French female speaker receives positive ratings. In addition, persuasiveness in neutral radio spots seems to correlate to a flat

and less variable pitch contour. The authors claim that the other features examined are not significantly relevant.

As mentioned at the beginning of this chapter, Calamai and her collaborators have conducted several studies on the attitudinal judgements of Italian listeners with respect to non-native speech. Bianchi and Calamai's (2012) work focuses on how prejudice and stereotypes are associated to foreign-accented Italian. Their sociolinguistic study employs the verbal guise technique, and it investigates Italian listeners' reactions to samples of non-native speech, produced by male speakers of L1 Albanian, Romanian, and American English. The authors also record two native speakers of Italian: one speaking Florentine Italian and the other standard Italian. The judgments are provided by 1st year university students attending B.A. programmes in Humanities. Regarding the identification task, the Florentine speaker is labelled either as 'Tuscan', or 'Florentine'. Next, the Romanian accent is recognised by 44% of the respondents, followed by American English (38% labelled as 'English', and 2% as 'American') and Albanese (33%). The standard Italian voice is judged as the most pleasant by 68% of the listeners, followed by the Florentine (64%), the Albanese (40%), the American (27%), and the Romanian (16%). Judgments regarding the speakers' socio-cultural backgrounds and personality follow the same trend, revealing that the Romanian accent is the most stigmatized.

In a similar experiment, by means of the verbal guise technique, Calamai (2015) examines how Albanian-, Romanian-, and General American-accented speech is perceived by 97 young native speakers of Italian (i.e. high-school students). In the perceptual experiment, listeners have to identify the speakers' nationalities and to provide profiles (i.e. accent pleasantness, level of educational qualification, personality traits, degree, profession, cultural background) for them by answering various questions (e.g. 'Do you think that he usually reads newspapers?', 'What kind of job do you think he has?', etc.). The Italian accent was recognised by all listeners, followed by Romanian (82%), Albanian (60%) and American (26%). Another finding of this study is that the listener's attitude towards a certain nationality may have an effect, "both on the perception of the linguistic information and on the social profile attributed to the speaker" (Calamai, 2015: 303). In fact, as one may expect, the Italian speaker is the most positively judged, while on the other hand, confirming the finding in Bianchi and Calamai (2012), the Romanian accent seems to be the most stigmatized.

More recently, in Calamai and Ardolino (2020), two separate experiments are performed to examine biased attitudes towards non-native Italian speech, specifically Chinese-accented Italian, among teachers and students in high schools in Tuscany. In particular, the study employs overt question-

naires and the Implicit Association Test protocol to explore the existence of implicit negative attitudes towards Chinese-accented Italian. The results reveal a significant disparity between implicit and explicit attitudes towards foreign accent. The research also highlights the sociolinguistic implications of these findings for education and future research. The authors suggest that combining implicit and explicit attitudes in the same experimental design may provide a more comprehensive analysis of listeners' perceptions and reactions to non-native speech.

Another study by Calamai et al. (2021) focuses on discrimination in school settings, describing the results of field investigations conducted in Tuscany from 2018 to 2020. The ethnic and linguistic stereotypes held by teachers towards their students are examined. The study reports the results of two sections of the questionnaire: one on expectations relating to academic success and the other on ethnic stereotypes. The findings indicate that teacher prejudice towards students varies based on social status, gender, and migration background, and students with a migration background from certain ethnic groups receive more positive assessments than others. The study also highlights the importance of comparing field evidence to reference literature and comments from teachers to ensure the validity of research questionnaires.

A similar study by Nodari et al. (2020) investigates the attitudes of Italian school teachers towards students' foreign-accented speech in three different cities in Tuscany, Italy. The authors collected 144 questionnaires from both lower and upper secondary schools, asking about teachers' ideology towards foreign accents and their opinions towards multiculturalism and multilingual education at school. The results suggest that teachers hold negative biases towards foreign-accented Italian and that there is a correlation between language attitudes and the evaluation of multilingual and multicultural educational practices. These attitudes appear to be consistent across different cities, suggesting that exposure to linguistic diversity in Italy is not a significant factor in shaping attitudes towards foreign-accented speech.

Finally, Pellegrino et al. (2015) conduct perceptual analyses with the purpose of investigating transfer phenomena from the Neapolitan dialect on L2 Italian speech. Their corpus consists in 15 non-native speakers aged 18-50, from Sri Lanka, Bangladesh, Pakistan, Ukraine, Burkina Faso, Kosovo, Dominican Republic, Russia, Senegal, and Venezuela, and residing in Naples. Based on the qualitative perceptual analysis, the realisation of the voiceless postalveolar affricate /ʧ/ as a voiceless postalveolar fricative /ʃ/ is the most frequent phenomenon, followed by final vowel variation (e.g. [o] as [ə]). Subsequently, the authors perform quantitative analyses on the [ʧ]

as [ʃ] phenomenon and they report that its occurrence is correlated to dialect exposure rather than to the length of stay in Naples.

2.2.4 Specific pronunciation training

On the 3rd of January 2018 edition of the Public Broadcasting Service's *In My Humble Opinion*, entitled *We stigmatize accents, but language belongs to everyone*[32], Judy Woodruff interviewed the Pulitzer Prize nominee and Columbia University Professor Hernan Diaz. After a brief consideration on accent-reduction courses, Diaz tackled topics relevant for this monograph such as accent profiling, accent discrimination, and stigma:

Walking around campus the other day, I was perplexed to see flyers advertising accent reduction or even accent elimination. Having been born in Argentina, grown up in Sweden, and spent most of my life in the United States, I have, to some degree, a foreign accent in every language I speak. […] Does my accent need correcting? I don't think so. To sound like who, exactly? A native speaker? What would that even mean? […] Even though everybody has an accent, there certainly is such a thing as accent discrimination. Most of us have either suffered or witnessed it at some point. I can easily tell when I'm not being understood or when someone is underscoring a difference in pronunciation just to show me my place, because accent discrimination is, in the end, all about place, who belongs and who doesn't (Diaz, *We stigmatize accent, but language belongs to everyone*, *In My Humble Opinion*, Public Broadcasting Service, the 3rd of July 2018, television programme).

In the excerpt above Diaz questions the usefulness of pronunciation training and accent elimination classes, pointing out that sounding like a native speaker, especially when the target language is American English, may be, to a great extent, a utopia. His considerations are reasonable and generally embraceable, but his criticism to pronunciation training – he implies accent reduction and elimination courses might have emerged a result of accent discrimination and stigma – does not take into account the speaker's personal, instrumental, and integrative motivations to sound native-like. Bearing these aspects in mind in the next pages I will engage in exploring the relation between pronunciation instruction and foreign accent. The intent is not to provide techniques and methods for teaching pronunciation to learners of L2 (for such purposes, several resources exist, such as Ballerini and Rosati, 1990; Costamagna, 1996, 2000; Calabrò, 2010, 2015, 2016; Sis-

---

32. Available at: https://www.pbs.org/newshour/show/we-stigmatize-accents-but-language-belongs-to-everyone (accessed on the 23rd of January 2023).

inni, 2016; Calabrò and Mairano, 2017; etc.). Instead, I will discuss the findings of the studies that tested the role of specific pronunciation training on foreign accent. Some general considerations will also be drawn up.

In the preceding chapters, the term 'pronunciation' was used on multiple occasions, but it is now important to provide a generally accepted definition of 'pronunciation' within the SLA framework:

Pronunciation in language learning, […], is the practice and meaningful use of the target language phonological features in speaking, supported by practice in interpreting those phonological features in a target language discourse that one hears (Burgess and Spencer, 2000: 191-192).

As other SLA studies and the previous sections of this work have shown, the degree of importance placed on pronunciation may vary depending on the learner's goals and the level of fluency required in their L2 usage. However, speaking with a foreign accent may have repercussions on how the learner's performance in L2 is perceived by the native speakers of that language (Lippi-Green, 2012). This may lead to implications at personal, social, communicative, and professional levels: preoccupation of not mastering the language and therefore reluctance to interact, incomplete integration to the community of native speakers, a more challenging road to better job opportunities and university degree programs, etc. For this purpose, some anecdotal examples regarding the feelings and concerns of speakers of L2 Italian with respect to their accents will be provided below. These testimonials[33] are extracted from online forums, magazines, and newspapers:

1. *Dunque, sono un ragazzo dell'Est Europa (bielorusso) e vorrei farvi una domanda: ma è vero che tutte le ragazze in Italia hanno paura dell'accento straniero?* (So, I am a guy from Eastern Europe (Belarusian) and I would like to ask you a question: but is it true that all girls in Italy are afraid of foreign accent?; Dron, *Stranieri in Italia – Forum Stranieri*, the 22nd of September 2005, online[34]).
2. *[C]ome ha sentito il mio accento straniero, […], ha iniziato a strillare. Davanti a tutti mi ha insultato e diceva di voler parlare con un'italiana.* (As soon as she heard my foreign accent, […] she started to yell. She insulted me in front of everyone, saying that she wanted

---

33. Translations into English are provided between the brackets.
34. Available at: http://www.stranieriinitalia.it/forum-stranieri/viewtopic.php?f=6andt=5832 (accessed on the 18th of January 2023).

to speak to an Italian.; ElMo, Siete mai state vittime di un episodio di razzismo?, *Yahoo! Answers*, 2008, online[35]).

3. *È dura, anche perché il rapporto con le altre ragazze non è sempre facile. […] Beh, all'inizio mi prendevano un po' in giro per il mio accento straniero, mi chiedevano da che paese vengo, dove ho imparato l'italiano. Anzi, temevo che questo potesse ostacolarmi per l'ingresso in finale.* (It's hard, also because the relationship with the other girls is not always easy. [...] Well, at first they teased me a little bit about my foreign accent, they asked me what country I was from, where I learned Italian. In fact, I was afraid that this might hinder me from entering the finals.; Bellezza brasiliana con cuore romagnolo, *Il Resto del Carlino*, the 14th of September 2011, online[36]).

4. *I romani […] appena sentono il mio accento straniero mi identificano come badante.* (Romans [...] as soon as they hear my foreign accent they identify me as a caregiver; Colonnelli, Corriere della sera – Cinquantamila, Occhi brasiliani guardano Roma. *Claudileia Lemes Dias, ritratto spietato dell'arroganza*, the 15th of September 2012, online[37]).

5. *Ero restia a cantare in italiano perché si sente il mio accento straniero.* (I was reluctant to sing in Italian because you can hear my foreign accent.; Bizzini, Kay McCarthy L'irlandese italiana, *Distorsioni*, the 18th of July 2014, online[38]).

6. *Appena ha sentito il mio accento straniero ha iniziato ad insultarmi con frasi ingiuriose anche pesanti, ridendomi in faccia.* (As soon as she heard my foreign accent, she started to insult me, using very offensive insults and laughing in my face.; Io e il mio cucciolo aggredi-

35. Available at: https://beta.answers.yahoo.com/question/index?qid=20101104094243AAYijm0andsort=N (accessed on the 18th of January 2023).

36. Available at: https://www.ilrestodelcarlino.it/ravenna/cronaca/2011/09/15/581235-bellezza_brasiliana.shtml (accessed on the 18th of January 2023).

37. Available at: http://www.cinquantamila.it/storyTellerArticolo.php?storyId=0000002214927 (accessed on the 18th of January 2023).

38. Available at: http://www.distorsioni.net/canali/interviste/lirlandese-italiana (accessed on the 18th of January 2023).

ti da un cane sciolto, *Cronache maceratesi*, the 20th of February 2016, online[39])

7. *[D]all'altra parte del telefono, quando hanno sentito il mio accento straniero, mi è stato chiesto di che colore fossi.* (From the other side of the phone, when they heard my foreign accent, they asked me about the colour of my skin.; Assandri, Niente stanza in affitto perché sono di colore, *La Stampa*, the 19th of July 2016, online[40]).

8. *Un'altra cosa che ho notato è che appena mi sentono parlare con il mio accento straniero, si rivolgono a me con il "tu", e non con il "lei" come si converrebbe ad una persona che non si conosce.* (I also noticed that as soon as people hear my foreign accent they address informally, instead of using the more formal and appropriate Italian pronoun, as one would expect to hear.; Una russa expat in Italia racconta…, *Donne che emigrano all'estero*, the 30th of November 2016, online[41]).

The examples provided above show that having a foreign accent may can indeed pose challenges to non-native speakers of Italian. Besides facing clear episodes of accent bias and discrimination (examples 2, 6, and 7), or stigma (examples 1, 4, and 8), the L2 pronunciation skills seem to pose integration and interaction challenges (examples 1 and 3), leading to reluctance to speak, perform, and sing in this language (examples 3 and 5). These testimonials and Hernan Diaz's recount on the leaflets advertising accent reduction classes, suggest that L2 speakers feel the need to reduce their foreign accent. In fact, a quick search on the web returns an impressive number of online courses, specialised on accent reduction or elimination, that promise to improve the learners' English pronunciation (e.g. Accent Advisor's Accent reduction training online course[42], Accent Eraser's Accent reduction training online course[43], Pronunciation Pro's English pronunciation course[44], etc.). As far as Italian is concerned there are less pro-

---

39. Available at: https://www.cronachemaceratesi.it/2016/02/20/io-e-il-mio-cucciolo-aggrediti-da-un-cane-sciolto/772314/ (accessed on the 18th of January 2023).

40. Available at: http://www.lastampa.it/2016/07/19/cronaca/niente-stanza-in-affitto-perch-sono-di-colore-g5zZxeIF5aZiInCCcT75VL/pagina.html (accessed on the 18th of January 2023).

41. Available at: http://donnecheemigranoallestero.com/russa-expat-italia-racconta/ (accessed on the 18th of January 2023).

42. Available at: https://accentadvisor.com (accessed on the 19th of July 2019).

43. Available at: http://accenteraser.com (accessed on the 19th of July 2019).

44. Available at: https://www.pronunciationpro.com (accessed on the 19th of July 2019).

nunciation courses advertising the reduction of foreign accents. Still, there are courses available, both in paper- and web-based formats, specifically designed for individuals who aspire to speak Italian like a native speaker or enhance their pronunciation skills; they are meant to be used either individually or in classroom settings. At the date this monograph was written I managed to find various onsite and online pronunciation courses, as well as several pronunciation coursebooks[45], such as Corso di pronuncia italiana by Studiamo.COM[46], Corso di pronuncia italiana per stranieri by Language Academy[47], Corso di italiano by Loecsen[48], Italian Pronunciation Course in Florence by Parola Italian Language School[49], Imparare l'italiano in Italia – Corso di dizione per cantanti lirici, Centro Culturale Conero[50], and Lezioni di pronuncia, Italiano a Venezia[51]. In some of these cases, L2 speakers are promised to reduce or even cancel their foreign accent rapidly.

A review of the websites of the afore-mentioned courses reveals that when it comes to challenging aspects of Italian pronunciation, they report mainly issues at the segmental level and propose exercises aimed at reducing articulation errors, completely overlooking prosody. In most cases, the pronunciation lessons proposed consist in perception and production exercises, as well as in contrastive analyses between the target language and the learner's production.

In this regard, Moyer (2013: 147) discusses about the type of pronunciation skills that teachers tend to prioritise, saying that phonological instruction focuses primarily on individual sounds in isolation. She argues that this approach has received significant criticism, as there is a growing consensus that isolated drills are insufficient, and a standardised approach is not suitable. Such an approach contradicts the pedagogical principle of learner-centeredness and ignores the increasing recognition of individual differences in SLA.

---

45. I refer to all sorts of courses and coursebooks that specifically aim to improve the L2 learners' pronunciation skills in Italian. For the sake of brevity, I do not mention here general-purpose pronunciation coursebooks for Italian (e.g. Carboni and Sorianello, 2011).

46. Available at: https://www.studiamo.com/ita/italiano-online-pronuncia-italiana.php (accessed on the 3rd of January 2023)

47. Available at: http://www.languageacademy-italian-for-foreigners.com/corso-pronuncia-italiano-per-stranieri-milano (accessed on the 3rd of August 2019).

48. Available at: https://www.loecsen.com/it/corso-italiano (accessed on the 3rd of August 2019).

49. Available at: https://www.parola.it/italian_pronunciation_course.html (accessed on the 3rd of August 2019).

50. Available at: https://www.linguaitaliana.com/it/dizione-per-cantanti-lirici.html (accessed on the 3rd of January 2023).

51. Available at: http://it.italianoavenezia.com/venezia-Lezioni+di+pronuncia-5-1.asp (accessed on the 3rd of January 2023).

Besides the classroom- and web-based pronunciation courses presented above, it is worth mentioning three practical coursebooks for non-native speakers that address Italian pronunciation, namely Costamagna's (1996) collection of *Pronunciare l'italiano. Manuale di pronuncia per stranieri*; Dall'Armellina et al. (2005) *Giocare con la fonetica*; and Calabrò's (2010) *E tu... come pronunci? Eserciziario di fonetica italiana per italiani e stranieri*. The three textbooks – all containing a CD-ROM – provide non-native students with a more exhaustive guide to Italian pronunciation. In particular, Costamagna's (1996) course covers all aspects of Italian phonology, and its 2002 version is also supported by a pronunciation training software. Dall'Armellina et al. (2005) as well as Calabrò's (2010) coursebooks propose a series of exercises aimed at teaching specific pronunciation skills to L2 learners of Italian.

However, hardly ever such specific pronunciation books developed as a result of SLA research are actually used in classroom; in fact, it has been often stressed that when an L2 is taught, instructors tend to focus on improving their students' syntax skills and enriching their L2 lexicons, disregarding pronunciation (Derwing et al., 1998; Harmer, 2001; Derwing and Munro, 2005; Darcy et al., 2012; Calabrò, 2015; Derwing and Munro, 2015). Eventually, this would interfere with the learners' wishes and aims to sound native-like. Even if most teachers acknowledge the importance of pronunciation instruction, not all of them have enough preparation to teach it properly (Kelly, 2000).

Most pronunciation exercises opt for a behaviourist bottom-up approach (Reimann, 2018), oftentimes in the form of the audio-lingual method (Celce-Murcia et al., 1996). This means that learners listen to audio samples, previously recorded by native speakers of Italian, that consist in vowels and consonants in various environments and combinations, as well as in words, phrases, or longer utterances, and eventually they may be asked to repeat them. Teachers and/or, more recently, computer-based interfaces provide corrective feedback. Some other exercises based on the audio-lingual method may include reading and memorizing tongue twisters, recognising minimal pairs, listening to songs and learning them – a type of exercise often used with young learners – (e.g. Costamagna et al.'s (2010) *L'italiano con le canzoni*); reading aloud in classroom specific text excerpts, etc. The rationale behind this approach is that examples and repetition help learners perceive and produce difficult structures of L2 speech.

The effectiveness of pronunciation training has been discussed in the SLA community. It is however necessary to highlight that the interaction between L2 teaching and SLA research has not always been entirely symmetric and commensurable. In fact, it has been speculated that the sophisti-

cation of modern research has created a gap between the two fields (Ellis, 1997). As SLA research findings are not always actualised in classroom, outdated approaches on L2 teaching might hamper with the learner's interest and efforts to attain native-like pronunciation. In addition, pronunciation is often overlooked in L2 teaching. This may be due to the fact that language teachers themselves do not receive enough training on how to teach pronunciation, or they lack knowledge of phonetics and phonology (Wang and Munro, 2004), even if, most of the times they acknowledge the importance of pronunciation instruction. On a similar note, Moyer (2013: 147) highlights that SLA does not provide L2 teachers with practical guidance or useful insights, so they must determine for themselves the significance of pronunciation skills for their learners and how to address them formally. Furthermore, although newer L2 coursebooks cover some of the most relevant aspects of phonology by means of targeted exercises it appears difficult to quantify the amount and the quality of the pronunciation training learners really get during language courses.

In this regard, testing the effectiveness of pronunciation training is a particularly challenging task. Up to a decade ago, only few empirical studies had aimed to assess the relationship between specific pronunciation instruction and the degree of foreign accent. In more recent years, some studies tested the role of pronunciation training on the learner's foreign accent (e.g. Saito, 2011; Kissling, 2013; Algethami, 2017; Combei et al., 2020; etc.). Before presenting the review on the role of pronunciation training on foreign-accented speech, it is important to note that there are many methodological differences between the studies that addressed this topic in terms of the target language, the number and type of learners (i.e. L1, age, L2 level, etc.), the amount and type of training (i.e. most of the times the focus is only on segments, less frequently on prosody, or on both levels), the metrics used to measure pronunciation skills (i.e. overall degree of foreign accent assessed by human raters, generally native speakers; instrumental analysis of acoustic properties of speech; etc.). For this reason, the results of the studies to be presented below are hardly comparable.

One of the first evidence of the benefits of pronunciation instruction comes from Elliott's (1997) experimental study on L2 Spanish learners – their L1 was English – that improved their perception and production of some specific Spanish segments (i.e. liquids and stops) after one semester of specific training. However, no improvement is reported regarding fricatives and vowels. Kissling (2013) tests the effect of pronunciation and phonetic instruction on L1 English speakers who were learning Spanish. The author claims that the learners' realisation of some Spanish sounds improved after the training.

Also, Moyer (1999) reports that the accents of some English-speaking learners of L2 German who had received special pronunciation feedback were rated as very similar to those of L1 speakers. Similar results are obtained by Missaglia (1999) on a group of Italian learners of German, and by Derwing et al., (1998) that investigate the performance of L2 English learners of various L1s; in both studies, the speakers' pronunciation skills improved, even if in Derwing et al. (1998) significant differences are obtained only in reading productions.

A totally different result is reported by Derwing and Rossiter (2003), who say that specific pronunciation instruction does not show any significant effect on the overall accent of the L2 English learners they examine. Next, Derwing et al. (2014) present an experiment that tested the role of a short-term pronunciation course on L2 speakers of English who had lived in Canada for a long time. The authors report that the specific training has no effect on the learner's foreign accent. Similarly, Algethami (2017) assesses the effect of instruction of English segments on the degree of perceived foreign accent, by analysing L1 Arab learners that had previously attended 11 hours of explicit pronunciation lessons. The results of this study do not reveal a significant improvement, as far as the learners' foreign accent is concerned. Neither Saito (2011) finds any improvement in the pronunciation skills of L2 English learners whose L1 was Japanese, after they had received specific training on English segments. However, Saito and Lyster (2012) report that due to targeted instruction of the English /r/, followed by corrective feedback, the L1 Japanese learners manage to produce the sound more accurately.

Next, Schmid and Pedrazzini's (2016) study tests the effect of detailed phonetic training on the production of German plosives. In a class-room experiment in Switzerland, 10 L2 learners of German, whose L1 is Italian, received a detailed phonetic training regarding the differences between German and Italian plosives. The learners' reading task were recorded twice, before and after the training. A control group of 10 learners that did not attend explicit pronunciation training was also recorded. The effect of pronunciation instruction is tested instrumentally, especially for the Voice Onset Time (VOT) and %Voice (i.e. the percentage of duration by which the signal of 'voiced plosives' is periodic). The authors report no statistically significant effect of the specific pronunciation training on the production of plosives; in fact, only 4 students revealed higher VOT values of the voiceless German plosives after the training.

As far as L2 Italian is concerned, there are relatively few studies that explore the relationship between targeted pronunciation training and the improvement of pronunciation skills, or more specifically perceived foreign

accent. Here five of the most representative studies are summarised. De Meo et al. (2013) explore the effects of imitation and self-imitation techniques on the overall degree of perceived foreign accent, the intelligibility, and the communicative effectiveness of 26 Chinese speakers learning Italian. The authors employ prosodic transplantation (i.e. transferring suprasegmental features from speech samples produced by L1 speakers to the samples produced by L2 speakers), while the training was performed in a computer-assisted learning setting. Four types of speech acts were considered: grants, orders, requests, and threats. The learners were divided into two groups, and they practiced imitation and self-imitation. The effect of self-imitation is tested by comparing pre- and post-training performances of both groups. Based on the ANOVA results, it appears that both instruction techniques help learners improve their pronunciation "providing a spin-off for prosody learning, communication effectiveness and intelligibility improvement, and foreign accent reduction" (De Meo et al., 2013: 97-98), but the self-imitation strategy leads to a more native-like prosodic performance.

In a similar study, De Meo et al. (2016) assess the role of imitation and self-imitation techniques on the acquisition of prosody. Their experiment was conducted on Chinese learners of L2 Italian, but a control group of native speakers of Italian was employed as well, for imitation purposes. During the experiment, 19 Chinese learners attended individual sessions, aimed at training intonational patterns for assertions, orders, and requests in Italian. For the imitation phase, the learners listened to sentences produced by native speakers, they exercised on their own, and eventually their imitation of the native speakers' model was recorded. Then, for the self-imitation phase, learners trained on speech samples obtained by means of prosodic transplantation (i.e. the suprasegmental features were transferred from the native speakers of Italian). After this training phase, the learners were recorded once again. The effect of both techniques is evaluated by performing comparisons between the learners' performance before and after the training. The perceptual measurements were given by 46 native Italian listeners who assessed the accentedness and the communicative purpose of the audio material they listened to. The authors claim that both methods of prosody training revealed effective at least to some extent: there was an improvement for the prosody of short sentences. However, comparing these results with those in De Meo et al.'s (2013) study, De Meo et al. (2016) claim that unlike unexperienced learners, the more proficient speakers are rated as less-accented and more communicatively efficient after the training.

Pellegrino and Vigliano (2015) assess the effectiveness of the self-imitation technique in the development of native-like prosody. They analyse 7 intermediate Japanese learners and 2 native speakers of Italian in a

reading task on 3 different pragmatic functions: granting, orders, and requests. Non-native speakers performed the task twice, before and after the self-imitation prosodic training. The training was performed on speech samples obtained by transferring the suprasegmental features of the native speakers on the speech uttered by the learners. To assess the effects of this technique, 17 Italian listeners rated pragmatic function and accentedness on the pre- and post-training speech samples. Based on the outcomes of this perception experiment, the authors claim that the self-imitation improves the learners' performances in terms of communicative effectiveness. However, the accentedness ratings do not vary significantly before and after the training.

By means of a survey-based study, Calabrò and Mairano (2017) test the effectiveness of their *Minimal Pair Finder* (MPF)[52] application on 12 Chinese learners of L2 Italian. This tool was built using word lists from the *CoLFIS* corpus[53] and from the dictionary *Grande Dizionario Italiano Garzanti*[54]. The learners managed to find the target pair investigated in this study – [m] vs. [m:] – easily, both perceptually and graphically. Also, based on the conversation in classroom and on the results of the surveys provided at the end of the activity, it seems that this type of training is evaluated positively by the Chinese learners. The authors conclude that this type of exercise may help learners develop their phonetic and phonological awareness.

Finally, Nicora et al.'s (2018) study investigates the effect of a perception-production training on the production of Italian yes/no questions (i.e. lexical stress and intonation pattern) by Hiberno-English learners. For this experiment 5 female speakers (age: 16-27) having an A2[55] level in Italian were enrolled, and they were divided as follows: the first group (3 speakers) attended an explicit perception-production training, while the control group (2 speakers) attended conversation classes. More specifically, the training focused on improving the speakers' skills regarding their phonological/pragmatic awareness as well as the phonetic implementation. The authors claim that only learners that had participated to the explicit pronun-

---

52. Available at: http://phonetictools.altervista.org/minimalpairfinder (accessed on the 16th of January 2023).

52. Available at: http://linguistica.sns.it/CoLFIS/Home.htm (accessed on the 1st of March 2023).

54. The word list of *Grande Dizionario Italiano Garzanti* is available at: http://www.garzantilinguistica.it/lemmario-italiano/ (accessed on the 16th of January 2023).

55. A description of *CEFR* (Common European Framework of Reference for Language Skills) levels is available at: https://europass.cedefop.europa.eu/it/resources/european-language-levels-cefr (accessed on the 16th of January 2023).

ciation training were able to improve their production of both the lexical stress and the intonation pattern in the yes/no questions in Italian.

The findings of most studies reported above might suggest that teaching pronunciation is beneficial in enhancing the learners' awareness of segmental and suprasegmental features of L2 speech, but due to the multitude of methodological differences it is still particularly difficult to draw universally generalizable conclusions.

Pronunciation teaching has also been explored through a technological lens. The advent of innovative technologies has greatly enhanced the field of L2 teaching and learning. For instance, in recent years, three-dimensional (3-D) talking heads have become more and more common in teaching. Some examples are Graesser et al.'s (2004) AutoTutor, Massaro's (2005) and Massaro et al.'s (2005, 2006) Baldi, Cosi et al.'s (2002) Baldini: Baldi Speaks Italian, Cosi et al.'s (2003) and Leone et al.'s (2012) LUCIA. Computer-based language courses have also started to employ talking heads in order to teach L2 learners how specific sounds, words, or discourse structures should be uttered. Almost two decades ago, Massaro (2005) presented his work regarding the possible applications of talking heads in language learning. Massaro's (2005) and Massaro et al.'s (2006) Baldi is a 3-D animated talking endowed with synthesized and natural speech. It was employed in language-learning programs for various types of learners, including impaired ones, and it contains a variety of exercises, among which those that focus on improving speech articulation and on developing phonological awareness. Learners were able to train their perception and production of speech segments in isolated contexts and in various constructions. Besides the possibility to see the talking head from various angles, another particularly helpful feature of this tool is that it may be set to display midsagittal perspectives, or even to remove the skin in order to reveal the internal articulators. This could help learners increase their phonetic and phonological awareness. Massaro (2005) and Massaro et al. (2006) reported that the learners' pronunciation skills improved regardless of the training method employed on Baldi.

Nowadays, Mobile or Computer-Assisted-Language-Learning platforms (MALL and CALL) are endowed with state-of-art speech synthesizers, and they may realistically substitute human tutoring when the latter is not available (e.g. the learner cannot access language courses taught by native speakers; impaired learners; etc.). Learners can benefit from these tools in various way: potentially unlimited training of the perception of sound contrasts (e.g. discrimination of minimal pairs); learning, imitating, and consolidating L2 specific segmental and suprasegmental features (e.g. the perception and production of consonant gemination, lexical stress, intonation,

etc.); and monitoring their progress autonomously. Some of these applications also provide real time pronunciation feedback automatically, by assessing the goodness of fit of the L2 speaker's production to the L1 models stored (e.g. Duolingo[56]). Nevertheless, one of the drawbacks of this type of feedback is that the tool may sometimes be under- or over-fitted: in the first case, even samples of incorrect pronunciation may be rated as 'correct'; while, on the other hand, in case of over-fitted models, the tool could rate as 'incorrect' samples of otherwise acceptable pronunciations. For these reasons, other CALL platforms allow learners to store their productions and eventually human feedback is provided (e.g. Living Language Platinium[57]).

An even more specific type of MALL/CALL tool is the computer-assisted pronunciation training (CAPT). Although systematic research has been conducted in this field, both in industry and in academia, most studies consist in descriptions of CAPT technologies and applications (e.g. Eskenazi, 1999; Setter, 2008; Silke, 2012; Walker, 2014; Fouz-González, 2015; Chen and Li, 2016; etc.). There is still more to be done regarding the effects and benefits of teaching and learning pronunciation by means of these tools. According to Chen and Li (2016), CAPT applications may be divided into those specialised: in pronunciation assessment, generally employed for testing purposes in language centres; and in pronunciation teaching and learning, whose direct users are language teachers and learners. Most applications of the latter group are dedicated to the English language (e.g. Pronunciation Power[58]; Accent Master[59]; etc.). As far as L2 Italian is concerned, Costamagna's (2002) *Pronunciare l'italiano* software on a CD-ROM offers various lessons on all aspects of Italian phonetics and phonology: 12 on vowels; 36 on consonants; and 19 on words stress, syllables, intonation, and other general activities aimed at improving the learners' perception and production skills.

While reviewing CAPT, Hardison (2004) and Levis (2007) stress that these tools have some advantages over the conventional pronunciation instruction delivered in a traditional classroom environment: CAPT may be used without interruption, as it is 'tireless' (e.g. it maximises retention and learning processes since stimuli are seen more often and they consist in various input typologies); its activities are specifically tailored on the learner and the topic to be learned; it allows the learner to train on her/his own; it has consistency in terms of interface and features (e.g. learners receive feedback always and in most cases it is immediate); it offers visual support

---

56. Available at: https://www.duolingo.com (accessed on the 16th of January 2023).
57. Available at: https://www.livinglanguage.com (accessed on the 16th of January 2023).
58. Available at: https://www.englishlearning.com (accessed on the 23rd of January 2023).
59. Available at: http://www.accentmaster.com (accessed on the 24th of January 2023).

(e.g. for training articulation; for enhancing attention and working memory; etc.); and it may be easily used as a complementary tool during traditional language lessons.

Very few empirical studies have assessed the effects of CAPT on the learners' overall pronunciation skills. One of the first studies that aimed to do so is Tanner and Landon's (2009) work on L2 English. The authors claim that the learners that joined 11-weeks self-directed CAPT programme managed to improve their prosody in English, especially regarding their perception and production of pauses, lexical stress, and intonational curves. Next, Luo (2016) tests a pronunciation training technique, for Taiwanese learners of English, that combined CAPT and in-class instruction. Parallelly, two control groups were exposed only to teacher-directed in-class pronunciation training. The author claims that, based on the post-training results, integrating the CAPT technique proved more efficient in reducing the learners' pronunciation issues compared to only conventional instruction.

Mehrpour et al. (2016) investigate whether the use of an accent reduction software during conversation classes would bring any benefits for Iranian learners of English enrolled in a foreign languages program at the university. The Farsi version of the Accent Master software was employed for this purpose. The application includes various lessons on segmental and suprasegmental elements, and it is endowed with a series of audio-visual features (e.g. animated IPA transcription; waveforms; pitch trace; intonation contours; side and front views of articulatory organs; etc.), displaying comparisons between the learners' production and those of the model speakers. The experimental group was made up of 16 learners who trained on the Accent Master software. Parallelly, the learners in the control group were not exposed to this CAPT application, but they attended traditional pronunciation lessons. The authors report that the learners improve their pronunciation skills after the CAPT exposure. The goodness of pronunciation is assessed by the authors themselves and by a native speaker of English, using a pronunciation battery of tests that focused on various segmental and suprasegmental markers. The scores obtained by the learners in the experimental group are significantly higher than those of the control group. However, learners in both groups have an overall good communicative ability in English. After the CAPT training, the learners in the experimental group improve their pronunciation skills.

Finally, Del Monte's (2011) study that explores the effect of speech technologies in language learning concludes that using speech technology alone can be highly beneficial in language learning environments, but it is not sufficient. Additional tools are required to address the pedagogical challenges present in all learning scenarios. When it comes to prosodic learn-

ing, only a small number of problems related to prosody can be identified and adequately addressed through this type of feedback. Therefore, there is a considerable way to go before CALL systems can be used for L2 teaching. The incorporation of animated tutors in a complete system for literacy tutoring or pronunciation instruction presents the most demanding scenario. While animated tutors are already a reality, considerable work is still needed to tackle pedagogical issues in both the visual and speech domains.

The results of these studies suggest that CAPT indeed may be a solution in those situations in which exposure to the target language is not possible (e.g. unavailability of a native-speaker language teacher), since they are easily-accessible sources of authentic input. However, more studies on more target languages, Italian included, are still needed to systematically explore whether these tools have indeed any role on diminishing the degree of foreign accent. This work aims to bring a contribution to this debate, by testing, among other factor, the role of pronunciation training on the perceived foreign-accented Italian.

3. Foreign accent and technology

Over the years, the way people communicate and engage with the world has undergone significant transformations. More recently, natural language understanding (NLU) and natural language generation (NLG) have been commonly used in Artificial Intelligence (AI) technologies for a better interaction between humans and machines. The constant progress in this field has led to the creation of highly performant man-machine interfaces able to assist users in complex tasks, such as driving or working.

Speech technology and its subfields have become some of the most active and interdisciplinary areas of research. Automatic speech recognition (ASR) alone is a large field that it is based on intensive research in many different fields: acoustics, digital signal and speech processing, phonetics and phonology, computer science and engineering, etc. Besides the large number of academic contributions, a lot of commercial applications in this field have been designed and developed. In recent years, systems based on ASR, NLU, and NLG have become more common than ever. For many daily-life actions people use applications that employ ASR, at least to some extent, for instance to:

- dictate texts or e-mails (e.g. IBM's Watson Speech-to-Text[1]; etc.);
- perform hands-free voice searches online on their mobile phones (e.g. Google[2] search engine released its voice search in 2012);
- ask for directions to the GPS navigator (e.g. Google Maps[3]; Here WeGo[4]; etc.);

1. Available at: https://www.ibm.com/watson/services/speech-to-text/(accessed on the 27th of January 2023).
2. Google search engine is available both as a mobile phone application and a web application. Available at: https://www.google.com (accessed on the 27th of January 2023).
3. Google Maps is a free service both as a mobile phone application and as a web application. Available at: https://www.google.com/maps/ (accessed on the 27th of January 2023).
4. Here WeGo is a free service both as a mobile application and as a web application. Available at: https://wego.here.com (accessed on the 27th of January 2023).

- communicate with the virtual assistants on mobile phones and computers (e.g. Siri[5]; Cortana[6]; Bixby[7]; Google Assistant[8]; Alexa[9], etc.);
- use automatic voice translation and interpretation (e.g. Tywi[10] offers an automatic voice translation service for 78 languages; Google Translate[11] and iTranslate VOICE[12] manage to perform automatic voice translation directly on the mobile phone).

Additionally, in case of visually- or hand-impaired users, the voice technology applications listed above as well as the tools designed specifically for these impairments have aimed to reduce the communication difficulties that these groups have been facing (Liu et al., 2015; Mansoor et al., 2015; etc.). For instance, screen-reading and dictation software (e.g. Nuance TALKS[13]) have become available for most languages. Moreover, various applications built for able-bodied users have been adapted with accessibility features aimed to improve the everyday lives of impaired users as well (e.g. Nuance Dragon[14] speech recognition software).

Among all the tools presented above, it is worthwhile to note that the wide availability of voice technology has especially led to an impressive rise of voice searches. In fact, the importance of voice search and its ever-growing daily use are also the most relevant topics of the Internet Trends Report[15] by Kleiner Perkins Caufield Byers that stresses how the queries associated to voice-related commands have increased 35 times since 2008.

5. Available at: https://www.apple.com/it/ios/siri/ (accessed on the 27th of January 2023).

6. Available at: https://www.microsoft.com/it-it/windows/cortana (accessed on the 27th of January 2023).

7. Available at: https://www.samsung.com/it/smartphones/galaxy-s8/intelligence/ (accessed on the 27th of January 2023).

8. Available at: https://assistant.google.com/intl/it_it/ (accessed on the 27th of January 2023).

9. Available at: https://alexa.amazon.com (accessed on the 27th of January 2023).

10. Available at: http://www.translateyourworld.com/en/about/ (accessed on the 27th of January 2023).

11. Google Translate is a free service both as a mobile phone application and a web application, the latter is available at: https://translate.google.it (accessed on the 27th of January 2023).

12. Available at: http://itranslatevoice.com/index.html (accessed on the 27th of January 2023).

13. Available at: https://www.nuance.com/mobile/mobile-solutions/talks-zooms.html (accessed on the 27th of January 2023).

14. Available at: https://www.nuance.com/dragon/accessibility-solutions.html (accessed on the 27th of January 2023).

15. Available at: https://www.kleinerperkins.com/perspectives/2016-internet-trends-report (accessed on the 27th of January 2023).

According to the afore-mentioned report, in 2015, Siri handled more than 1 billion weekly requests through speech; likewise, in 2016, 25% of searches on Windows 10 taskbar were made through speech, while similarly, 20% of the searches on the Android systems in USA were again voice searches; three of the most frequent queries are: "Navigate home", "Call mom", and "Call dad". The report also quotes Andrew Ng – a Baidu[16] Chief Scientist – who claims that by the end of 2020 half of all searches were either through speech or images.

There is a simple reason behind the success of these applications: human beings can speak faster than they can type text on their keyboards or on their display devices, almost 3 times faster, according to Ruan et al.'s (2018) study. Also, using voice services, rather than typing, is easier, more comfortable, and safer in various circumstances, such as driving, walking, working, house-holding, or in general, when performing multiple tasks simultaneously, since an intensive and continuous use of hands and eyes is not required. According to the Internet Trends Report, the main environments where US users perform voice searches or requests are, in this order: at home (43%), in the car (36%), on the go (19%), at work (3%). As far as Italy is concerned, according to the FIND[17] report on mobile search, in 2016, 53% of the Italian smartphone users had never used the voice search, while 46% had performed a search through speech at least once in their lives (16% had used it often, 12% only sometimes, and 18% rarely or only once). However, there is a clear increasing trend of using voice search in Italy, considering that in 2013, only 38% of the users had tried to use this service.

For all the reasons above, the market demands that voice-based applications should be able to manage all kinds of real-life vocal input, including voices characterised by noise, age, pathological and accented speech (both regional and foreign accents), etc. Ideally, the performance of these tools should be good enough to decipher the message and fulfil the speaker's re-

---

16. Baidu Inc. is a Chinese multinational company that provides various technology and internet-related services, among which the most popular search engine in China, called Baidu, available at http://www.baidu.com (accessed on the 27th of January 2023).

17. To the best of my knowledge, the 2016 report by FIND called *Mobile Search in Italy - Gli italiani e la ricerca da smartphone* report was the only source available regarding the Italian users' behaviour on search engines. As far as voice searches are concerned, this report only takes into account the searches performed on smartphones. Available at: http://www.findsdm.it/report/FIND-Mobile-Search-in-Italy-2016_download.pdf (accessed on the 27th of January 2023). According to Digital Consumer Trends Survey 2022 by Deloitte, in 2022, 27% of Italians claimed to have a speaker with voice assistant (in 2017 it was 4%). Available at https://www2.deloitte.com/it/it/pages/technology-media-and-telecommunications/articles/digital-consumer-trends-2022.html (accessed on the 27th of January 2023)

quests in all circumstances. As far as foreign accent is concerned, different techniques, such as acoustic model adaptation and pronunciation adaptation have been used to overcome this issue, but to build these models, corpora of non-native speech are required.

3.1 Corpora of non-native speech

The advent of innovative speech-based applications has brought a lot of advantages to users worldwide, but at the same time it has posed various methodological challenges for developers and scholars working in the field. In this scenario, it has been evident that in order to develop efficient speech technologies, one had to develop good speech corpora[18]. Corpus development has, indeed, received a lot of attention, as speech corpora are indispensable resources for developing these technologies. The performance of speech technologies relies heavily on the corpora adequacy.

Depending on the purpose of use – language or pronunciation teaching and learning, sociolinguistic research, acoustic analysis, ASR, speaker verification, accent identification, etc. – the corpus should consider a large array of variables relevant for that specific application (Sturim et al., 2016). Below, Gibbon et al.'s (1997) definition of speech corpora is reported:

[A] spoken language corpus is any collection of speech recordings which is accessible in computer readable form and which comes with annotation and documentation sufficient to allow re-use of the data in-house, or by scientists in other organisations (Gibbon et al., 1997: 79).

Currently, a significant number of speech corpora are available. As one may expect, most of these corpora are built on mainstream varieties of languages (i.e. L1 rather than L2; 'standard' rather than regionally-marked; etc.). Only some speech corpora can be queried online. Here I am listing examples of corpora built for speech technologies[19]:

---

18. Depending on their purposes and forms, 'spoken language corpus', 'phonological corpus', or 'speech database' are terms used to refer to speech corpus (Durand et al., 2014).

19. A more complete list of speech corpora is available at the University of Pennsylvania's Linguistic Data Consortium (LDC) https://www.ldc.upenn.edu and on CLARIN platform https://www.clarin.eu/resource-families/spoken-corpora (accessed on the 28th of January 2023).

- Air Traffic Control corpus – ATC[20];
- CSLU 22 Language corpus[21];
- English Broadcast News Speech corpus – HUB4[22];
- Fisher Spanish Speech corpus[23];
- Fisher Levantine Arabic Conversational Telephone Speech corpus[24];
- ICSI Meeting Speech corpus[25];
- Robust Automatic Transcription of Speech corpus– RATS[26];
- Robust Speaker Recognition 2015 corpus – RSR2015[27];
- TIMIT Acoustic-Phonetic Continuous Speech corpus[28];
- YOHO Speaker Verification corpus[29].

All these corpora contain speech productions of native speakers of the target languages, and in most cases, standard varieties were collected. State-of-art ASR systems have become more accessible and performant due to increasing sizes of speech corpora of native speakers – especially true for English – and to more complex acoustic and language models (Graves et al., 2013; Kitashov et al, 2018). However, scientists have also understood that to improve the performance of these technologies, they had to take into account speech variability, such as age, gender, speaking style, emotions, accents, etc. Therefore, on the quest to provide more effective speech-based applications, in the ever-growing multiethnic and multilingual contexts, a better robustness to non-native speech is desirable. To overcome the challenges that emerge from speech and speaker

20. Available at: https://catalog.ldc.upenn.edu/LDC94S14A (accessed on the 28th of January 2023).

21. Available at: https://catalog.ldc.upenn.edu/LDC2005S26 (accessed on the 28th of January 2023).

22. Available at: https://catalog.ldc.upenn.edu/LDC98S71 (accessed on the 28th of January 2023).

23. Available at: https://catalog.ldc.upenn.edu/LDC2010S01 (accessed on the 28th of January 2023).

24. Available at: https://catalog.ldc.upenn.edu/LDC2007S02 (accessed on the 28th of January 2023).

25. Available at: https://catalog.ldc.upenn.edu/LDC2004S02 (accessed on the 28th of January 2023).

26. Available at: https://www.darpa.mil/program/robust-automatic-transcription-of-speech (accessed on the 28th of January 2023).

27. Available at: http://archive.signalprocessingsociety.org/technical-committees/list/sl-tc/spl-nl/2012-05/the-rss2015-speech-corpus/ (accessed on the 28th of January 2023).

28. Available at: https://catalog.ldc.upenn.edu/ldc93s1 (accessed on the 28th of January 2023).

29. Available at: https://catalog.ldc.upenn.edu/LDC94S16 (accessed on the 28th of January 2023).

variability, in the last couple of years Appen[30] – a company specialised in creating solutions and providing human-annotated training data for machine learning and artificial intelligence – has recruited, via crowdsourcing, a wide variety of speakers all over the world, including Italy.

Similarly, the Mozilla Internet browser started a project, called Common Voice[31], aimed at improving the performance of the ASR services they provide. Users worldwide can, in fact, contribute in two ways: either by reading and recording their own speech sample or by validating the speech samples recorded by other users. One could not, by all means, retrieve the number of samples that were produced by non-native speakers of Italian, since the recordings on Mozilla's Common Voice platform are anonymous and freely accessible to everyone on a voluntary basis (i.e. no requirement is set). Also, no information is provided about the number of samples, probably since it is constantly growing. However, for this work, I performed an ad hoc experiment that consisted in validating 1000 utterances on the platform on 5 different trials in 5 different days. In one of the trials, I was able to find one voice that was clearly foreign-accented, presumably an example of German-accented Italian produced by a male speaker. As far as the English version of the platform is concerned, there were clearly more examples of foreign-accented speech; following the sample protocol, I found a total of 38 samples that were clearly spoken by non-native speakers of English. Without claiming any conclusions out of these trials, I do believe that the developers acknowledge the important role played by foreign accent on the overall performance of voice applications. However, in order for this tool to incorporate accent-based variability non-native speakers should voluntarily record themselves on the platform, or they should be prompted to do so.

And even if investigating the differences between native and foreign-accented speech is a necessary step in non-native speech recognition, at the moment, the number of learner speech corpora is still relatively low if compared to corpora of native speech. In addition, until recently, for commercial reasons, the research focused on English. Some of the largest resources for non-native speech are available for this language; here are some of them[32]:

---

30. Available at: https://appen.com (accessed on the 29th of January 2023).
31. Available at: https://voice.mozilla.org/en (accessed on the 29th of January 2023).
32. A large list of non-native speech databases is available on the University of Louvain's section dedicated to learner corpora is available at: https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html (accessed on the 28th of January 2023). Also, Gruhn et al. (2011) survey the non-native speech databases available at the date their study was published. The lists I will provide in this section contain various corpora that are not present in any of the two surveys.

- ATR-Gruhn[33];
- CSLU: Foreign Accented English Release 1.2[34];
- Interactive Spoken Language Education – ISLE[35];
- L2-ARCTIC[36];
- N4 Nato Native and Non-Native Speech database[37];
- Speech Accent Archive[38];
- The HIWIRE database, a noisy and non-native English speech corpus for cockpit communication[39];
- The Wildcat Corpus of native- and foreign-accented English[40];
- Translanguage English Database – TED[41].

Large speech corpora of foreign-accented English are owned by Beijing Haitian Ruisheng Science Technology Ltd – Speechocean, and they were specifically built for commercial purposes, both for training and testing speech recognisers. Aside from speech corpora of standard American-, Australian-, British-, and Canadian-English, they store corpora of Chinese-, Taiwanese-, and Indian-accented[42] English.

Until recently, ASR systems privileged English, and in fact, the interest for other languages has started to develop only in the last decade, probably as a consequence of the fact that commercial voice-based applications have aimed to reach larger customer targets. As the world has become more globalized, non-native speakers are not a minority group of customers an-

33. Available at: http://universal.elra.info/product_info.php?cPath=37_39andproducts_id=1924 (accessed on the 28th of January 2023).

34. Available at: https://catalog.ldc.upenn.edu/LDC2007S08 (accessed on the 28th of January 2023).

35. Available at: http://catalog.elra.info/en-us/repository/browse/ELRA-S0083/ (accessed on the 28th of January 2023).

36. Available at: https://psi.engr.tamu.edu/l2-arctic-corpus/ (accessed on the 28th of January 2023).

37. Available at: https://catalog.ldc.upenn.edu/LDC2006S13 (accessed on the 28th of January 2023).

38. Available at: http://accent.gmu.edu (accessed on the 28th of January 2023).

39. Available at: http://catalog.elra.info/en-us/repository/browse/ELRA-S0293/ (accessed on the 28th of January 2023).

40. Available at: http://groups.linguistics.northwestern.edu/speech_comm_group/wildcat/ (accessed on the 28th of January 2023).

41. Available at: https://catalog.ldc.upenn.edu/LDC2002S04 (accessed on the 28th of January 2023).

42. Available at: http://kingline.speechocean.com/category.php?id=273 (accessed on the 28th of January 2023).

ymore. Without claiming to be exhaustive, I will mention some of the largest non-native speech databases for languages other than English, favouring those available online:

- Bavarian Archive for Speech Signals Strange Corpus 1 – SC1 – 'Accents'[43] for German;
- Bavarian Archive for Speech Signals Strange Corpus 10 – SC1 – 'Accents II'[44] for German;
- French Learner Language Oral Corpora – FLLOC[45] for French;
- Individualized Feedback for Computer-Assisted Spoken Language Learning – IFCASL corpus[46] for French and German;
- Japanese Speech Database Read by Foreign Students – UME-JRF[47] for Japanese;
- SINOD[48] for Slovenian;
- Spanish Learner Language Oral Corpora – SPLLOC[49] for Spanish;
- The University of Toronto Romance Phonetics Database[50] – RPD for speech produced by native and non-native speakers of French, Italian, Portuguese, Romanian, and Spanish (the corpus will be further described below).
- West Point Arabic Speech corpus[51] for Arabic;
- West Point Heroico Spanish Speech corpus[52] for Spanish;
- West Point Russian Speech corpus[53] for Russian.

---

43. Available at: https://www.phonetik.uni-muenchen.de/Bas/BasSC1eng.html (accessed on the 29th of January 2023).

44. Available at: https://www.phonetik.uni-muenchen.de/Bas/BasSC10eng.html (accessed on the 29th of January 2023).

45. Available at: http://www.flloc.soton.ac.uk/search.php (accessed on the 29th of January 2023).

46. Available at: http://www.ifcasl.org/corpus.html (accessed on the 29th of January 2023).

47. Available at: http://research.nii.ac.jp/src/en/UME-JRF.html (accessed on the 29th of January 2023).

48. Available at: http://universal.elra.info/product_info.php?cPath=42_43&products_id=1417 (accessed on the 29th of January 2023).

49. Available at: http://www.splloc.soton.ac.uk (accessed on the 29th of January 2023).

50. Available at: http://rpd.chass.utoronto.ca (accessed on the 29th of January 2023).

51. Available at: https://catalog.ldc.upenn.edu/LDC2002S02 (accessed on the 29th of January 2023).

52. Available at: https://catalog.ldc.upenn.edu/LDC2006S37 (accessed on the 29th of January 2023).

53. Available at: https://catalog.ldc.upenn.edu/LDC2003S05 (accessed on the 29th of January 2023).

Since many non-native speech corpora are built for commercial purposes within private research centres, it is quite difficult to map all the resources of this type ever built. Most of the resources listed above are available online, but only few of them can be accessed freely, and in all the other cases users may need to a pay subscription.

Currently, the ASR systems for Italian which are integrated into the virtual assistant software commercially available (e.g. Google Assistant, Siri, Cortana, Alexa, etc.) perform well on native speech (Tamburini, 2022, for an overview of recent ASR architectures for the Italian language). Despite recent advances in this field, non-native accents still represent a challenge for the freely available applications listed above. This may be due to fact that there is significantly less training data available for speech recognition on non-native pronunciations.

However, considering that Italy is a multicultural country, with over 5 million foreign citizens, representing 8.5% of the entire population residing on its territory[54], it would be desirable to provide generally available ASR services and applications to a wider variety of users who speak Italian with non-native accents. As already mentioned, a good starting point would be creating adequate learner speech corpora for Italian. Besides being used like training sets for speech recognition or text-to-speech systems, these resources might be beneficial in the field CALL and MALL, as well as in speaker profiling applications. Additionally, SLA experts and language teachers working on Italian might also benefit from the presence of these corpora.

When this monograph was written, there were only two speech corpora specifically built for foreign-accented Italian freely available online, namely Dialoghi in Italiano Lingua Straniera – DILS[55] and The University of Toronto Romance Phonetics Database – RPD, the latter has been introduced above.

DILS consists of semi-spontaneous audio material obtained by means of the task-oriented dialogue elicitation technique. It contains 9 audio samples (for a total duration of 100 minutes) uttered by 18 speakers: 12 Dutch female speakers from Belgium, 3 Spanish females and 3 Spanish males. The cities where the material was recorded (Ghent and Almeria) and the learners' Italian language skills were disclosed, but other sociolinguistic and socio-cultural details are not provided. This speech database does not contain transcriptions and annotations.

---

54. Further information is available at *Bilancio Demografico* (Demographic Report) published on the 20th of March 2023 by the Italian National Institute of Statistics (ISTAT) at: http://dati.istat.it/Index.aspx?DataSetCode=DCIS_POPSTRRES1 (accessed on the 20th of March 2023).

55. Available at: http://www.parlaritaliano.it/index.php/it/dati/794-corpus-dils-dialoghi-in-italiano-lingua-straniera (accessed on the 29th of January 2023).

The RPD corpus is divided into four sections: the Romance Language Survey – RLS consisting in material produced by native and non-native speakers of the five major Romance languages: French, Italian, Portuguese, Romanian, Spanish; the Dialect Atlas of Argentina; the Experimental phonology database for French and Spanish; then, second and third language acquisition database for French and Spanish. Regarding the Italian set of RLS, the data were collected in Toronto from 15 native speakers and 15 second/third language learners having various L1s (i.e. English, Spanish, Albanian, and Amharic). Speakers read a set of words in a carrier sentence, they read the Italian version of the story *The Northwind and the Sun*, they retold the story of *The Red Ridinghood* by looking at a set of pictures; and finally, they talked about their favourite meal. The corpus is segmented and annotated and therefore it is also possible to perform queries on target phonemes, for instance. Various sociolinguistic data (e.g. gender, age, education level, etc.) are provided together with the audio files and the corpus interface allows the user to refine their searches according to the variables they are interested in.

In addition to the above-mentioned corpora, there exists a database of written and spoken non-native Italian, entitled *Archivio Digitale di Italiano L2* – ADIL2 (Palermo, 2009), which is purchasable in the form of a DVD. Based on the description in Palermo (2009) the corpus is endowed with a sophisticated search tool, it is accurately transcribed, and it contains an admirable amount of data collected.

Finally, it is also worthwhile to mention that there are several other learner corpora for Italian: *Varietà Apprendimento Lingua Italiana Corpus Online* – VALICO[56], which is a collection of non-native written Italian; *Corpus di apprendenti di Italiano L2* – CAIL2[57], also consisting in samples of Italian written by non-native speakers; Longitudinal Corpus of Chinese Learners of Italian – LOCCLI[58] , as the name suggests it contains material written by Chinese learners of Italian; Corpus of Chinese Learners of Italian – COLI[59], this resource, unlike the previous one contains also transcriptions of speech produced by Chinese learners of Italian; *Lessico dell'italiano parlato da stranieri* – LIPS[60]; *Corpus Parlato di Italiano L2*[61]. The last two corpora consist of transcriptions of audio samples produced by

56. Available at: http://www.bmanuel.org/projects/br-HOME.html (accessed on the 29th of January 2023).

57. Available at: https://www.unistrapg.it/cqpweb/cail2/ (accessed on the 29th of January 2023).

58. Available at: https://www.unistrapg.it/cqpweb/cina/ (accessed on the 29th of January 2023).

59. Available at: https://www.unistrapg.it/cqpweb/coli/ (accessed on the 29th of January 2023).

60. Available at: http://www.parlaritaliano.it/index.php/en/data/653-corpus-lips (accessed on the 29th of January 2023).

61. Available at: http://elearning.unistrapg.it/osservatorio/Interrogazione.html (accessed on the 29th of January 2023).

non-native speakers, but a multimodal version of the latter, containing annotated videoclips, is being built.

3.2 Dealing with foreign accent in speech technology

Following the findings in Tatman's (2017) Ph.D. dissertation, on the 15th of February 2018, The Economist published an online article entitled *In the world of voice-recognition, not all accents are equal*[62] that stresses how applications based on ASR systems (e.g. Siri, Google Assistant, etc.) are still unprepared to face accented-speech, showing a clear bias towards standard pronunciation:

[T]he speech-recognisers are largely trained on just one [accent] per country: "General American" and Britain's "Received Pronunciation". Speakers with other accents can throw them off. Some might consider that an unlucky but avoidable consequence of "having an accent". But everyone has an accent, even if some are more common or respected. The rise of voice-activated technologies threatens to split the world further into accents with privileges—in this case, the ability to command the Echo, Apple's Siri, Google Assistant and other such gadgets—and their poor relations. (In the world of voice-recognition, not all accents are equal; *The Economist*, the 15th of February 2018)

This newspaper article suggests a sort of 'electronic imperialism', which might lead to a form of exclusion; for the people that have a 'standard' pronunciation, speech technologies work well, while for all the others they might not. On the same note, recently, a lot of anecdotal evidence regarding tests on commercial voice assistants has become available. For instance, Wired[63], the monthly American magazine, tested the limits of the English-speaking versions of Siri, Echo[64], Google Home[65] in various circumstances[66]. In one experiment children of all ages were asked to perform some

---

62. Available at: https://www.economist.com/books-and-arts/2018/02/15/in-the-world-of-voice-recognition-not-all-accents-are-equal (accessed on the 29th of January 2023). It appeared also in the print edition under the headline 'Alexa's biscuits'.
63. Available at: https://www.wired.com (accessed on the 30th of January 2023).
64. Available at: https://www.amazon.com/all-new-amazon-echo-speaker-with-wifi-alexa-dark-charcoal/dp/B06XCM9LJ4 (accessed on the 30th of January 2023).
65. Available at: https://store.google.com/it/product/google_home (accessed on the 30th of January 2023).
66. To the best of my knowledge, at the time this book was written, no such experiments were available for Italian.

speaking tasks on the afore-mentioned everyday AI tools[67]; apparently, Google Home outperformed other tools. In another experiment they tested AI against a variety of accents (e.g. Australian, British, German, Italian, Japanese, etc.) to determine which tool was the best at understanding most people[68]. Even if the video claims that some tools outperformed others, it is clear that AI still struggles with foreign-accented speech.

To deal with these shortcomings, commercial applications would need additional training on non-native accents. As stressed in the previous section, this may be achieved during the design and developmental phases. However, in most cases, these tools are designed so that users themselves can train them on their speech variety after the purchase. A circular causality issue may arise from this type of approach: if, at the beginning, users realise that the tool does not support their accents and it does not understand them they will be discouraged and therefore they might even give up training the tool.

It has been shown that speech recognition is still unstable to speaker variability (Soky et al., 2021). Since state-of-art ASR systems are under-trained for non-native speech, especially for lesser-known accents, they perform poorly when they encounter these varieties, independently of the approach chosen to solve the task. One of the main issues of the inter-speaker variation is accent (Radzikowski et al., 2021). ASR systems trained on standard or on mainstream pronunciation models may not recognise non-native speech when evaluated on a mismatched test condition. They recognise words as a sequence of elements defined in a pronunciation dictionary, and consequently, sometimes it is not possible to get a full match with non-native speaker utterances, precisely because these productions deviate from the standard pronunciation. However, the variations caused by non-native speech are not casual, they are closely linked to the universal or L1 specific phenomena discussed in the previous chapters. Therefore, recognising foreign accents prior to the ASR would allow the system to improve its performance by adapting its acoustic model and by selecting alternative pronunciations.

A lot of studies have focused on compensation techniques for foreign-accented speech. Generally, ASR systems employ some acoustic models adapted with data taken from non-native speech corpora (Behravan, 2016; Weninger et al., 2019). Some of the most recent employ deep neural network (DNN) acoustic modelling approaches. Huang et al. (2014), for in-

67. Available at: https://www.youtube.com/watch?v=GZnUibN6m4Aandfrags=pl%2Cwn (accessed on the 30th of January 2023).

68. Available at: https://www.youtube.com/watch?v=gNx0huL9qsQandfrags=pl%2Cwn (accessed on the 30th of January 2023).

stance, propose a multi-accent DNN acoustic model with an accent-specific top layer – used to model the accent-specific patterns – and shared bottom hidden layers – to share knowledge between the native and the non-native models; a Kullback-Leibler-divergence-regularised adaptation model is used to train the accent-specific top layer. The authors claim that this approach reduces error rate in ASR for British and Indian accents. A different foreign accent adjustment approach was adopted by Kitashov et al. (2018). Their method uses the difference in pronunciation between specific accents – taken from the Speech Accent Archive – and General American English, at the word level, and it eventually creates new accented samples. Additionally, the model can learn all generalizations that previously were manually set by phonologists. This statistical method is used to generate a million phonological variations of words from the Carnegie Mellon University (CMU) Pronouncing Dictionary, finally a sequence-to-sequence Recurrent Neural Network (RNN) is trained to recognise accented words. The system reaches 59% accuracy in recognising these words.

An equally interesting approach is offered by Fukuda et al. (2018) that use data augmentation to improve the recognition of foreign-accented English. More specifically, modified copies of two accents, Latin American and Asian are created with voice transformation (i.e. modifications of glottal source and vocal tract parameters), noise addition, and speed modification. The authors claim that all augmentations improve the accuracy, especially speed modification. Additionally, the benefits of training accent-specific models with the augmented data are significant. On the other hand, supervised and unsupervised adaptations with the augmented data do not yield substantial improvements. The authors report that this approach reduces ASR Word Error Rate (WER) up to 30% over a baseline approach trained only with accented data.

There are several know strategies to overcome the issues linked to foreign accents in ASR, focusing especially on efficient techniques and features for the identification and classification of non-native speech[69]. Typically, one can distinguish between two types of approaches to automatic recognition of foreign accents: phonotactic and acoustic. On the one hand, the phonotactic approach implies that accents differ in terms of their distribution of phone sequences (Jalalvand et al., 2012); therefore, using probabilistic methods, the accent recognition task consists in estimating the probability of having a specific sequence of phones given the target accent (e.g. the method of phone recognition followed by language modelling). On

---

69. Research on this topic has been productive only for some languages, especially English. Here I will provide a selection of some of the most representative studies. To the best of my knowledge, no such research has addressed foreign-accented Italian.

the other hand, the spectral approach supposes that accents differ with respect to their spectral features (e.g. Mel frequency cepstral coefficients – MFCC [70]); therefore, speech samples are represented as a set vectors of acoustic features and the identification is determined by maximum likelihood estimations.

One of the first approaches to foreign accent classification is Hansen and Arslan's (1995) study on German-, Turkish-, and Chinese-accented American English. They propose a method that uses a source generator framework based on prosodic features. The Hidden Markov Model (HMM) algorithm determines effective prosodic and acoustic features for accent classification. The authors claim that although speakers vary in terms of pitch structure, some global aspects could predict accent; additionally, energy, duration, and spectral information, as well as phonemic substitution seem to perform well on accent detection tasks. Regarding the system performance, for unknown open speech structures a classification rate of 81.5% was achieved; the rate increases to 88.9% in case of a three-word set.

Vieru et al. (2007) have worked on the perception and automatic identification of non-native French accents. Their 2007 study attempts to automatically discriminate between various foreign accents in French (Arabic, English, German, Italian, Spanish and Portuguese). The approach proposed here consists in an automatic alignment into phonemes of non-native French recordings that allowed the authors to compute values for vowel formants, consonant duration and voicing, prosodic cues, as well as pronunciation variants combining French and foreign acoustic units (e.g. a rolled /r/). A total of 62 features were used to train the foreign accents classifiers. The authors report a 50% correct identification rate, by using a cross-validation method with some unseen data, for Logistic Regression and Support Vector Machines (SVM) algorithms.

Similarly, Boula de Mareüil et al. (2008), focused on the identification and description of foreign and regional accents in French. They performed perceptual experiments on native listeners, and then measured the cues that would be most informative in discriminating among the accents. According to the results reported, the most robust features for automatic accent identification are the following: the devoicing of voiced stop consonants, the shift of /e/ toward [i], confusion between /b/ and /v/ and /s/ and /z/, the so

---

70. The concept of Mel frequency cepstral coefficients (MFCC) is generally attributed to Bridle and Brown (1974) and Marmelstein (1976). MFCCs are coefficients of the Mel-frequency cepstrum (MFC), namely a representation of the power spectrum of a sound, computed on the Mel-bands (the Mel scale is a non-linear scale of frequency, scaled to the human ear), instead of the Fourier spectrum, by means of a linear cosine transform (Formiga and Alias, 2009). These features are commonly used in ASR.

called "rolled r", the schwa fronting or raising as far as the German, English, Arabic, Spanish, Italian and Portuguese accents are concerned, the /O/ fronting for the Northern France accents, the production of schwa and the dentalization of nasal vowels for the Southern France accents.

Similarly, Vieru et al.'s (2011) study presents an approach of characterisation and identification of non-native French. They conduct perceptual and automatic classification experiments on Arabic-, English-, German-, Italian-, Portuguese-, and Spanish-accented French, in order to identify the most robust acoustic cues for discriminating among these accents and assess how native speakers of French perceive foreign accents. The authors measured consonant duration and voicing, the first and the second formant for vowels, prosodic features regarding word-final schwa, and the percentages of confusions obtained using automatic alignment. In order to select the most robust features for the accent classification task and then classify the speakers, WEKA machine learning techniques were used. It seems that the features that mostly characterise non-native speech are the devoicing of stop consonants (e.g. /s/ vs. /z/), the rolled /r/, and schwa fronting and raising. The 50% accuracy rate obtained in automatic classification of the six foreign accents compares favourably to perceptual data.

Behravan and his colleagues have published various studies on foreign accent recognition that will be summarised below in a chronological order. First of all, in Behravan et al. (2014), a hybrid acoustic and phonotactic approach is proposed for recognising samples of Russian-, Albanian-, Arabic-, English-, Estonian-, Kurdish-, Spanish-, and Turkish-accented Finnish taken from the Finnish National Foreign Language Certificate (FSD) corpus. For this experiment a common set of speech attributes shared across different languages are defined. In order to extract the manner attributes from speech samples, shallow neural networks were used. Also, contextual information in the forms of delta and double delta features was computed from the attribute features and then appended; eventually they were modelled with *i-Vectors*. This technique outperforms by 16% the SDC-MFCC i-vector baseline system. In addition, contextual information also brings benefits for the recognition rate.

Next, Behravan et al. (2015) present a spectral feature based *i-Vector* recognition system, that is tested on non-native speech extracted from the FSD corpus. The authors evaluated three aspects of the recognition system: its parameters; data used for estimating its hyper-parameters; and 'language aspects', such as confusion patterns among the accents. The results show that training hyper-parameters from the application-specific dataset is better than training them from mismatched dataset results. The highest accent recognition accuracy was achieved with this hyper-parameter setting: UBM

with 512 Gaussians, i-vector dimensionality of 1000 and a heteroscedastic linear discriminant analysis (HLDA) dimensionality of 180. The authors claim that L1 traits are more common among older speakers and that the highest accuracy is achieved with speakers that have a lower level in L2.

Behravan et al. (2016) investigate the effects of HLDA and linear discriminant analysis (LDA), of the duration of training and test samples, of the place of articulation, and of feature level fusion on the accuracy of foreign accent recognition. The experiments are conducted on foreign-accented English and on data taken from the FSD corpus. The results of this study reveal that the best recognition accuracy is obtained by using manner attribute features with *i-Vectors*. Additionally, the attribute-based system performs better than the spectral-based system, irrespectively of the amount of training data and test utterance length. At the same time, incorporating contextual information, and using a manner-and-place-of-articulation-based system improves the overall accuracy recognition rate. This system performs better than a spectral *i-Vector* system on 7 out of 8 accents with a statistical significance level of 5%.

For his Ph.D. dissertation, Behravan (2016) proposes a system based on a universal acoustic characterisation of speech utterances. Speech attributes (i.e. manner and place of articulation) are extracted and modelled using an *i-Vector* representation paradigm, in order to characterise the non-native speech. The principal component analysis (PCA), is employed to capture the temporal context of attribute feature streams. In order to improve the *i-Vector* modelling, a HLDA is compared and contrasted with an LDA. Finally, an out-of-set (OOS) data selection approach is proposed. The system is tested on the FSD and the US National Institute of Standard and Technology (NIST) corpora. The author claims that this approach brings an improvement of accuracy in foreign accent recognition over other approaches, such as the Gaussian mixture model-universal-background model (GMM-UBM) spectral technique, or the *i-Vectors* systems based on shifted delta cepstrum (SDC) features.

Siddhant et al. (2017) propose a different approach to accent identification, consisting in a training phase on speakers' L1 speech together with the accented speech. The authors built a deep Siamese network model that learns the association between accented speech and the L1 speech. The Siamese networks were trained with *i-Vector* features extracted using either an unsupervised GMM model or a supervised DNN model. The accent identification tests were performed on the CSLU Foreign Accented English (FAE) corpus. The deep Siamese networks achieve a significant performance improvement of 15.4% on a 10-class accent identification task (Portuguese-, Hindi-, Farsi-, German-, Hungarian-, Italian-, Manda-

rin-, Russian-, Spanish-, Tamil-accented English), over a baseline DNN-based classification system that uses GMM *i-Vectors*.

Ahamad et al. (2020) address the challenge of identifying speech spoken by non-native speakers using ASR technology. The authors outline the requirements for a corpus of well-curated speech samples in non-native accents to train and test robust ASR systems. They introduce AccentDB, a database containing samples from various Indian-English accents as well as native-English and metropolitan Indian-English accents. The authors analyse the separability of the collected accent data and test several accent classification models that are evaluated against human-labelled accent classes. They also propose the task of accent neutralization using autoencoder models to transform non-native accents to native accents, aiming to enhance ASR systems at different stages of development.

Although the two studies summarised below investigated regionally or geographically proximate accents, I believe that they are relevant, since the methodologies may be employed for foreign accents, as well. First of all, Mannepalli et al.'s (2015) study presents a method of identifying Andhra-, Telangana-, and Rayalaseema- accented Telegu, an Indian language in the Southern part of the country. The system employed in this experiment is based on MFCC features and a Gaussian mixture model (GMM) for the classification task. The authors report an overall classification rate of 91%.

Other approaches to the identification of accented speech are presented in Brown's (2016) study. The author presents five different systems (i.e. GMM-UBM, GMM-SVM, Phon-GMM-SVM, ACCIDST[71] based Correlation, and ACCDIST-based SVM) for identifying geographically-proximate English accents (Berwick-upon-Tweed, Eyemouth, Carlisle and Gretna). The approach that better discriminates among these accents is the ACCDIST-based SVM with an overall accuracy of 87.5%.

The studies presented above propose highly engineered solutions either for reducing the effect of foreign accents in ASR, or for classifying them. However, very few linguistically meaningful conclusions may be drawn from some of the studies deploying these approaches, since generally, even the features used to model these systems are highly abstract. Another aspect that immediately emerges is that regarding the pro-

---

71. ACCDIST was defined as 'a metric of the similarity between speaker's accents that is largely uninfluenced by the individual characteristics of the speakers' voices' (Huckvale, 2007).

cessing of foreign-accented Italian speech, research is still at the beginning, and more studies are needed to reduce error rate in ASR. It closely relates to the fact that there are very few learner corpora of non-native Italian speech. Not only would these resources be beneficial as training sets for ASR systems or for developing linguistic profiling applications (e.g. automatic verification or identification of foreign accents), but they might as well be employed for conventional language classes of L2 Italian or for courses on CALL and MALL planforms.

4. A corpus of non-native speech for Italian

In the previous chapters, I stressed the importance of the availability of speech corpora for non-native speech. Not only would these resources be useful as training sets for ASR systems that continue to have a poorer performance on non-native speech, but they may also be employed for MALL and CALL tools, especially for improving the decisions of the pronunciation assessment exercises for L2 Italian (see Chapter 2). Speech corpora for non-native speech are used also for the training purposes of speaker profiling tools. These profiling techniques are useful to analyse speech patterns and provide insights into the speaker's age, L1, and other language background and proficiency. They are also used in various applications such as security, forensic investigations, and personalised advertising. As far as foreign-accented English and foreign-accented Hindi are concerned, such tools for accent classification or accent verification were proposed to assess the speaker's real origin, for instance in border entry settings (Kulshreshtha and Mathur, 2012). However, it should be noted that these tools can also be problematic as they can perpetuate biases and stereotypes based on the speaker's demographics and socio-cultural characteristics, leading to incorrect conclusions. The use of profiling tools should, therefore, be approached with caution and with awareness of their limitations.

Besides the applications in the field of speech technology mentioned above, L2 speech corpora may also be used in the context of teaching Italian as a foreign language (e.g. examples of authentic non-native speech are valuable resources for building targeted exercises on specific pronunciation errors). Moreover, empirical studies in various fields of linguistics (e.g. sociolinguistics; experimental phonetics; etc.) could benefit from resources of spoken L2 Italian. A lot of studies that have analysed non-native speakers of Italian are hardly comparable because they use ad hoc datasets, sometimes even insufficiently large to draw statistically significant conclusions. An adequately designed speech corpus of L2 Italian, both in terms of quantity and quality of the data collected, would allow scholars to perform comparable perceptual and instrumental-acoustic analyses.

Due to the limited availability of such resources for L2 Italian speech and considering all the aspects listed above, for this study I collected a non-native speech database that was used for the perceptual and classification experiments to be discussed in Chapters 5 and 6. In §4.1 I will present the phases of data collection, while in §4.2 I will describe the database.

4.1 Data collection[1]

The corpus used for the experiments presented in this work was designed, collected, and developed from January 2016 through July 2017. The aim was to provide a uniform collection of audio material produced by young adult non-native speakers of Italian residing in Italy[2].

4.1.1 Recruitment process

At the beginning of this research, the intent was to collect data for 11 different accents, corresponding to speakers whose L1s were: Maghrebi Arabic, Urdu, Mandarin Chinese, Albanian, Russian, English, German, French, Romanian, Spanish, and Italian (as a control group). The first 10 groups would have represented the L1s spoken by some of the major groups of foreign students residing in Bologna[3], either those engaged in exchange programmes or those regularly enrolled. However, recruiting an adequate number of speakers for these 10 groups proved to be particularly difficult. This issue might have arisen because participation was completely voluntary, with no monetary reward being provided to informants as an incentive. Since it was not possible to recruit enough speakers of Maghrebi Arabic, Urdu, Mandarin Chinese and Albanian, these four groups were dropped. Generally, some specific criteria of quality, quantity, and diversity were taken into consideration, for each L1, when the participants were selected (this will be further explained in §4.2). The following variables were considered:

1. Part of this section was published in Combei (2017), as a paper in conference proceedings.
2. To facilitate the data collection process and to limit unnecessary variability, I chose to recruit only people that resided in Bologna.
3. The data referring to the students enrolled at the University of Bologna are available at: https://www.unibo.it/it/ateneo/chi-siamo/luniversita-oggi-tra-numeri-e-innovazione (accessed on the 3rd of February 2023). Other foreign students attend the Academy of Fine Arts, Johns Hopkins University, Music Conservatory Giovan Battista Martini, Bologna Business School, Alma Graduate School, and Dickinson College.

- L1 (English, German, Russian, French, Romanian, Spanish, and English);
- Age (young adults);
- Gender (female, male, other);
- Occupation (regular or exchange students enrolled in Bachelor's or Master's degree, Ph.D. or specialization programs in Bologna).
- Age of onset (the first exposure to the Italian language: birth, infancy, adolescence, adulthood);
- Self-assessed proficiency level in Italian (ranging from A2 to C2 levels, based on *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment-CEFR*[4]);
- Length of stay in Italy (less than 12 months, less than 24 months, more than 24 months);
- Predominant language learning method (naturalistic, or guided and naturalistic);
- Presence or absence of specific pronunciation training in the language learning process.

Sociopsychological factors were not taken into account, since I was expecting the speakers to display similar patterns in terms of their attitudes and motivations.

When speakers were recruited, they were enrolled as regular or exchange students in Bachelor's, Master's, Ph.D., and specialization programs in Bologna, and they were contacted on their personal e-mail address. The e-mail message was sent to 600 people, and it contained a short description of the research project. Participants were informed about the tasks they would have performed, without detailing any of them. Nearly one fourth of the people contacted replied positively to the call.

4.1.2 Experimental protocol and recording

Upon the informants' acceptance to voluntarily participate in the project, they were invited one by one for the recording experiment. Each recording session lasted around 60 minutes and they were individual based. I guided and monitored the entire experiment. To guarantee uniformity, the same experimental protocol was employed for all subjects. The speakers

---

4. A description of *CEFR* (Common European Framework of Reference for Language Skills) is available at: https://europass.cedefop.europa.eu/it/resources/european-language-levels-cefr (accessed on the 3rd of February 2023).

understood that they would have been recorded and they gave their informed consent in writing to the use of their speech samples and sociolinguistic data for research purposes (see Appendix: A). They were also given an information sheet describing the project and the task they would have been asked to perform (see Appendix: B). Moreover, before each recording session, speakers were asked to fill in a detailed form regarding their sociolinguistic background (see Appendix: C). This allowed to gather information that would have been further coded in the corpus and used in the experiments conducted for this monograph. All these forms were written in Italian and in some cases they were translated orally into English to informants that did not comprehend them fully.

Since I wanted to avoid fallouts caused by stress and anxiety, speakers were explained the aims of the experiment pointing out that it was not a test, so they were expected to speak and behave as naturally as possible during the recording. Additionally, a pleasant environment was created. At the beginning of the session, speakers were made feel comfortable in a cozy setting; they were provided biscuits and room-temperature water; finally, a couple of minutes before the recording, they were invited to drink a glass of water to prevent throat dryness. The digital recordings were performed in a quiet environment with a *MacBook Pro 13'' Retina* (*i5 Core* 2.7GHz processor, 8 GB Ram) and a USB-powered plug-and-play *Samson METEOR MIC* cardioid pickup microphone, having the following characteristics: a condenser diaphragms of 25 mm; the frequency response of 20 Hz-20 kHz; a resolution of 16-bit, 44.1/48 kHz. The microphone was placed on a desk that had a flat surface, at 75 cm above the ground, in front of the speaker who was seated on a chair. The software used for all recording sessions was *Praat*[5]. The sampling parameters were set as follows: mono channel, 16-bit, 44.1 kHz, linearly encoded, *.wav* format.

4.1.3 Types of speech[6]

In order to have a more functional corpus, both read and spontaneous speech was collected. The speakers performed two tasks. First, they had to talk about themselves freely and describe spontaneously how they spent their most recent holiday. I opted to start with this task because it would have allowed speakers to get used to the microphone and the software progressively and more easily, without major consequences on the audio material produced. For anonymization purposes, the samples that contain names

5. Available at: http://www.fon.hum.uva.nl/praat/ (accessed on the 10th of February 2023).
6. This corresponds to the style variable examined in Chapter 5.

and explicit reference points to the speakers and other people were not stored in the corpus.

In the second part of the recording, the speakers were asked to read a 249-words article excerpt published on the Italian newspaper *Corriere della Sera*[7] (see Appendix: D). Its suitability was validated with other expert phoneticians at the University of Pisa. That specific reading fragment was chosen because it presented various levels of complexity: the text varied both in terms of word length and complexity; from a structural point of view, the sentences were sufficiently different and had various degrees of length; finally, the excerpt contained all Italian phonemes (Tables 4.1 and 4.2 for a general overview).

Table 4.1: The distribution of consonants in the reading task

Consonant	Nasals	Stops	Affricates	Fricatives	Approximants	Laterals	Trills
Number of tokens	112	221	18	102	20	74	75

Table 4.2: The distribution of vowels in the reading task

Vowel	Close front	Close-mid front	Open-mid front	Open central	Close back rounded	Close-mid back rounded	Open-mid back rounded
Number of tokens	102	106	26	131	27	111	12

A reading task is highly necessary for triggering difficulties that could emerge as a result of conflicting orthographic norms and conventions between the speakers' mother tongues and Italian (Wottawa and Adda-Decker, 2016). At the same time, it could allow speakers comparisons and analyses on the same type of material, fundamental both for the perceptual and the classification experiments conducted (see Chapters 5 and 6). The participants had two reading attempts and they were asked to read and speak as naturally as they could. In almost all cases, both reading attempts were stored.

---

7. Available at: http://cinquantamila.corriere.it/storyTellerArticolo.php?storyId=0000002228555 (accessed on the 3rd of February 2023).

4.2 Corpus description

This section will outline the main characteristics of the database. The first part will provide information regarding the speakers, while the second and the third will describe the audio material and the web application built to store and access the database interactively.

4.2.1 Speakers

The database is made up of the material produced by105 non-native speakers, and by 17 native speakers of Italian; the latter were recorded as a control group. When this research began, it was planned to recruit at least 15 speakers for each L1. The threshold set was reached for French, and it was exceeded for the other six groups (Table 4.3 for the exact number of speakers for each L1). However, as it will be pointed out in the next chapters, some of the speech samples collected were not used in the experiments, either because they were inadequate or because I wanted to balance some speaker-dependent variables.

Table 4.3: The number of speakers recorded

L1	Russian	English	German	French	Romanian	Spanish	Italian	**TOTAL**
Number of speakers	17	16	17	15	20	20	17	**122**

Since the aim was to collect speech produced by young adult speakers enrolled in some sort of degree or specialization programme in Bologna, the age range is 19-40 years, but most speakers are older than 20 and younger than 30 years (age mean: 25.07, standard deviation: 4.45). Despite the efforts to have a balanced gender representation – 600 people were contacted – the database is not gender-balanced, especially for the French and the Russian groups. After recruitment and recording, it appeared obvious that it was easier to engage female participants. In fact, 81 females (66%) and 41 males (34%) were recorded for after all for this corpus. The literature indicates that gender has not been reported as a major source of pronunciation issues, so it was assumed that gender imbalance had minor effects on the perceptual and the classification experiments conducted (Gruhn et al., 2011). As mentioned, the occupation variable was controlled, with all speakers being enrolled as regular or exchange students in degree pro-

grammes delivered by some of the universities based in Bologna. By the time the speakers were recorded, 26% of them had completed the high-school and they were enrolled in a Bachelor's degree, 45% of them had a Bachelor's degree and they were enrolled in a Master's degree, while 29% had a Master's degree and therefore they were attending Ph.D. and specialization courses.

At the corpus level, the age of Italian language onset (i.e. the age corresponding to the first consistent exposure to Italian, either at school or naturalistically) was distributed as follows: birth – corresponding to L1 speakers of Italian (14%); infancy (15%); adolescence (28%); and adulthood (43%). The data regarding the age of onset are predictable and they reflect how speakers learnt Italian. Considering that I recorded people that were living in Italy, either temporarily or for longer periods, I believe that none of them qualifies as learners that learnt Italian only by means of purely guided methods. Therefore, for this monograph, I decided to attach two levels to the variable regarding the method of Italian L2 learning, namely, 'naturalistic' – for those that had never attended conventional, private, or online Italian language classes – and 'guided and naturalistic' – for those that attended any type of Italian courses, and at the same time complemented their language skills, due to their naturalistic exposure to Italian because they were living in Italy. Leaving aside the Italian informants, the variable divided L2 speakers as follows: 56% of them learnt Italian naturalistically while 44% of them using both guided and naturalistic approaches. At the same time, at the corpus level, 25% of the learners were given specific pronunciation training, while 75% were no. As one may expect, only speakers that attended some form of guided training in Italian were taught pronunciation.

Regarding the time spent in Italy, since most informants were exchange students, 58% of them had spent 6-12 months in Italy by the time they were recorded. The remaining part had lived in Italy for 12-24 months (16%), or for more than 24 months (26%). Most L2 speakers had only lived in Bologna or in its suburbs. Therefore, not surprisingly, the great majority claimed that they had been exposed only to the variety of Italian spoken in Bologna. However, I acknowledge that the reliability of this statement is arguable. That is because especially unexperienced speakers were not aware of the differences between the regional varieties of Italian, even if, presumably they were exposed to them at university, or during their daily lives. As concerns the native speakers of Italian, they indeed had lived in Bologna, but for representativeness reasons, I opted for evenly engaging informants that were from the South, Centre, and North of Italy.

Considering that it was almost impossible to predict the speakers' proficiency level in Italian before meeting them, the balancedness is not guaranteed for all accent groups (e.g. no Romanian speaker had A2 waystage/elementary level in Italian). For brevity purposes, at the corpus level, this variable is represented as follows: waystage/elementary level - A2 (8%), threshold/intermediate level - B1 (30%), vantage/upper-intermediate level - B2 (17%), advanced level – C1 (16%), and proficiency level – C2 (15%). As mentioned, these proficiency levels are self-assessed by the speakers themselves. Finally, native speakers represent 14% of the corpus.

4.2.2 Material

The dataset includes material produced by105 non-native and 17 native speakers of Italian. In total there are 8 hours of speech, consisting of in roughly 70,000 words. An average of 4 minutes and 25 seconds of raw audio material, both read and spontaneous speech was recorded for each speaker. Some of the speakers had to end the registration session earlier than planned, so in those cases it was possible to record and store only their first reading attempt. Moreover, some first reading attempts were not included in the corpus, since the speakers interrupted the recording due to external factors (e.g. asking for clarifications, feedback, skipping a line, etc.). Regardless of that, the spontaneous speech collected (22% of the material, corresponding roughly to 1 hour and 40 minutes) is, however, inferior in size to the reading speech material (78%, corresponding roughly to 6 hours and 20 minutes).

Considering that the material was produced by a wide variety of speakers, including unexperienced learners, I observed several production issues. Some are typical features of L2 speech, and they may be explained as 'transfer' phenomena (e.g. segmental and suprasegmental errors; etc.), while others are due to stress and anxiety (i.e. hyper- and hypoarticulation, artificial segmentation, filled and empty pauses; etc.).

All raw samples were segmented manually into utterances corresponding to grammatical sentences for the reading material, and to phonological sentences for the spontaneous speech. The material was not qualitatively altered, so hesitation phenomena and disfluencies (i.e. false starts, filled and silent pauses, phoneme lengthening, mispronounced words), mouth clicks, and external noise were left as they were.

4.2.3 Web application

Since there is scarcity of corpora of spoken L2 Italian (see §4.1), one contribution of this work is that of compiling the database of non-native Italian described above. Following the model of similar tools (e.g. The University of Toronto Romance Phonetics Database – RPD, Speech Accent Archive, etc.), a repository to host the database was created. The web application (webapp) is capable of extrapolating and classifying the audio files from the dataset, according to specific criteria, that will be discussed below. For the creation of the webapp, the web framework *Django*[8], as well as several *Python* libraries (*MySQL-python*[9], *django-treebeard*[10], *django-filer*[11], *html5lib*[12], *sorl*[13], *wsgi*[14], *polymorphic*[15], *classy-tags*[16], *audiofield*[17], *appconf*[18], etc.) were employed. These choices allowed the use of a powerful *Object-relational mapping* (ORM)[19] system, equipped with a web interface for storing multiple data types into my *MySQL* database.

Furthermore, the *Django* web framework *ORM* favoured the construction of data collection models: a model is the only final data

8. A description is available at: https://www.djangoproject.com (accessed on the 5th of February 2023).

9. A description is available at: https://dev.mysql.com/doc/connector-python/en/connector-python-example-connecting.html (accessed on the 5th of February 2023).

10. A description is available at: https://django-treebeard.readthedocs.io/en/latest/ (accessed on the 5th of February 2023).

11. A description is available at: https://django-filer.readthedocs.io/en/latest/ (accessed on the 5th of February 2023).

12. A description is available at: https://html5lib.readthedocs.io/en/latest/ (accessed on the 5th of February 2023).

13. A description is available at: https://github.com/solariumphp/solarium (accessed on the 5th of February 2023).

14. A description is available at: https://wsgi.readthedocs.io/en/latest/ (accessed on the 5th of February 2023).

15. A description is available at: https://github.com/django-polymorphic/django-polymorphic (accessed on the 5th of February 2023).

16. A description is available at: https://django-classy-tags.readthedocs.io/en/latest/usage.html (accessed on the 5th of February 2023).

17. A description is available at: https://github.com/areski/django-audiofield (accessed on the 5th of February 2023).

18. A description is available at: https://django-appconf.readthedocs.io/en/latest/ (accessed on the 5th of February 2023).

19. The Object-relational mapping is a technique for converting data between incompatible type systems by means of a virtual object database. A description is available at: https://media.readthedocs.org/pdf/django-orm-cookbook/latest/django-orm-cookbook.pdf (accessed on the 5th of February 2023).

source containing the fields and the essential behaviours of the dataset and of the reference objects. Thus, each model is mapped to a single database table and each attribute represents a database field. The queries are performed by means of ad hoc *Application programming interfaces* (*API*s) for each model. The project is hosted on a server with a *CentOS 7*[20] operating system, and it is already configured for various types of *SQL*[21] and *NoSQL*[22] databases (*PostgreSQL*, *MongoDB*, *Cassandra*, etc.). Moreover, it supports the execution of some cloud computing platforms, such as *Amazon Web Services* (*AWS*)[23], which could improve its performance in case of an exponential growth of the computational complexity.

The database is stored as a webapp repository called *Corpus Audio di Italiano L2* (CorAIt)[24]; in the paragraphs bellow I will describe its front-end interface. The web database was conceived as a preview of the original material. Therefore, to overcome storing issues, and to observe the design model of Weinberger's (2015) *Speech Accent Archive* (Weinberger, 2015), the format of the audio files available on the CorAIt repository is *.mp3*. Samples coded in other formats (e.g. *.wav*, *.flac*, etc.) are available offline. In order to enable advanced queries, various layers of metadata were added to each audio file: the speaker's L1, gender, age of Italian language onset, age at the time the sample was recorded, level of Italian proficiency, Italian learning method, length of stay in Italy, proficiency in other foreign languages. Also, information regarding the type of sample (i.e. either read or spontaneous speech) and its quality was included (Figure 4.1).

20. A description is available at: https://www.centos.org (accessed on the 5th of February 2023).

21. A description available at: https://www.microsoft.com/it-it/learning/sql-training.aspx (accessed on the 5th of February 2023).

22. A description is available at: http://nosql-database.org (accessed on the 5th of February 2023).

23. A description is available at: https://aws.amazon.com/it/ (accessed on the 5th of February 2023).

24. CorAIt website is under construction, but the audio material is available on request and accessible from the dedicated section of the online repository, prior to registration and approval.

Figure 4.1: Interactive search tool and query results

For the purposes of this work, namely describing foreign-accented Italian by means of perceptual and computerised classification experiments, a complete transcription and annotation of the corpus was not necessary. Nevertheless, the phonetic transcription is on-going, and this feature will be released in the future. Moreover, following the example of the Speech Accent Archive, the grammatical sentences corresponding to the reading excerpts is already inserted under the audio samples of the reading task. Apart from the embedded audio player – which allows users to listen and download the audio sample – the window where the single result is displayed provides socio-demographic and socio-cultural information with respect to the speaker that uttered a certain speech sample, as well as qualitative information regarding the audio file (Figure 4.2).

Copyright © 2023 by FrancoAngeli s.r.l., Milano, Italy. ISBN 9788835154716

Figure 4.2: Results window

Recognising the existing imbalances in the database, particularly as concerns the gender variable, expanding the data collection efforts to enhance the representativeness of the corpus stands as one of the future objectives of this project. Nonetheless, the current version of the CorAIt database serves as a preliminary foundation; future iterations of the project can encompass other varieties of non-native speech to ensure a more comprehensive representation of the subject matter. Additionally, the database could benefit from more comprehensive orthographic and phonetic transcriptions, as well as annotations of disfluencies such as false starts, filled and silent pauses, phoneme lengthening, mispronounced words, as well as mouth clicks, and external noise. Although the non-native speech database was specifically designed to meet the requirements of this study on foreign-accented Italian, I believe it might be used for other purposes, too. Currently, CorAIt may only be accessed on request for academic research purposes. The website will be made available in the future, which could prove beneficial for teachers of L2 Italian, scholars in the field of SLA, and anyone interested in accessing a collection of speech samples from a diverse range of Italian speakers, including both native and non-native speakers.

5. Perceiving foreign-accented Italian

Chapters 1, 2, and 3 examined the treatment of foreign-accented speech in media, SLA, and speech technology research, with an emphasis on non-native Italian. Moving on, I will propose a series of experiments aimed at providing a representation of Russian-, English-, German-, French-, Romanian-, and Spanish-accented Italian, as compared to the native varieties of this language. More specifically, this chapter will present a three-task perception experiment aimed at measuring foreign accent through the ears of native speakers of Italian. First, section §5.1, will describe the experiment, focusing on its design, the choice of stimuli, tasks, and listeners. Then, sections §5.2 and §5.3 will present the results of the first two tasks of the experiment, modelling the responses, by taking into consideration listener-, speaker-, and style-related factors.

All measurements, descriptive analyses, and statistical models employed in this chapter are built on the R programming language and environment[1], using functions from various R packages[2]: RMongo, pastecs, doBy, coin, pwr, scales, ggthemes, ggfortify, ggvis, car, multcomp, vcd, devtools, testthat, roxygen2, RColorBrewer, RMarkdown, swirl, rcc, shiny, DAAG, MASS, leaps, relaimpo, gvlma, broom, sjplot, and the core tidyverse (ggplot2, scales, ggthemes, tidyr, dplyr, purrr, stringr, forcats).

5.1 The design of an accent-perception experiment

The perception of foreign accents has been at the centre of a lot of research that has addressed non-native Italian (see Chapter 2), but to the best of my knowledge no other study has investigated this topic on such a large-scale. With a focus on six foreign accents, the aim of this chapter is to es-

---

1. Available at: https://www.r-project.org (accessed on the 12th of February 2023)
2. The descriptions of the R packages used are available at: https://cran.r-project.org/web/packages/available_packages_by_name.html (accessed on the 12th of February 2023).

tablish which listener-, speaker-, and style-related factors influence the perceptual identification of non-native Italian speech. In this respect the degree of accent variation will be examined in detail, also by exploring how listeners assess the speakers' accentedness.

In order to trigger judgements from native speakers of Italian with respect to non-native speech, a three-task perception experiment was designed. This type of approach is common in perceptual dialectology (see Chapter 2). The experiment was computer-based, and it was conducted over the Internet, on a WordPress[3] platform. This allowed to reach a large, unbiased, and varied pool of listeners from all over Italy. The experiment was password-protected and there was also an IP verification for each informant in order to block multiple submissions. To ensure a solid and varied sampling, the experiment was delivered from May 2017 through January 2018. At the beginning of each session, an introductory page provided a short description of the experiment, indicating that it would last approximately 45 minutes and that more specific instructions for each task would be given progressively. Listeners were instructed in advance that the optimal environment for conducting the experiment would be a silent room, and that they would have needed a computer and headphones to complete the trial. To make the informants familiar with the audio samples, an example was provided on one of the introductory pages. Before going further, informants were asked to accept the conditions of the experiment and the informed consent for the use of the data collected; this represented a compulsory step, so in case an informant had declined these conditions the experiment would have been interrupted immediately.

In the first section of the experiment, listeners were asked to answer questions regarding their socio-demographic and socio-cultural background. This part was compulsory and qualifiable for further rounds. The variables and the levels that resulted after the sampling were the following:

- Gender (female, male);
- Age range (18-30, 31-45, 46-65);
- The region where the listener spent most of her/his life (Abruzzo, Apulia, Basilicata, Calabria, Campania, Emilia-Romagna, Friuli-Venezia Giulia, Lazio, Liguria, Lombardy, Marche, Molise, Piedmont, Sardinia, Sicily, Tuscany, Trentino-Alto Adige, Umbria, Valle D'Aosta, Veneto);

---

3. WordPress is a content management system based on a Hypertext Preprocessor and a MySQL. A detailed description of this platform is available at: https://wordpress.com (accessed on the 12th of February 2023).

- The highest level of education achieved (high-school, Bachelor's degree, Master's degree, PhD);
- The occupation (student, employee, unemployed);
- The proficiency level in foreign languages;
- The degree of familiarity with foreign accents, focusing on those investigated;
- The experience in the field of linguistics (having attended Bachelor's and Master's courses in linguistics).

In addition, the listeners had to indicate their L1 or L1s, because only monolingual[4] Italian speakers were admitted for further rounds. Finally, listeners were asked whether they had ever experienced hearing issues; those that answered yes were excluded from the experiment.

5.1.1 Listeners

Keeping in mind how a wide range of factors could affect the outcomes of this study, the aim was to reach a stratified sampling of listeners, without explicitly controlling any variable regarding the listeners. This also helped reducing intentional associations. Thus, the experiment was delivered on the internet, and it was disseminated by means of mailing lists to over 50 interest groups all over Italy. The e-mail provided a password that would have allowed participants to access the online platform. A sufficient number of informants was obtained for each level of the variables identified during sampling, even though a balanced distribution was not achieved, particularly concerning the age and gender variables. This limitation is acknowledged.

Out of the initial pool of over 500 participants who embarked on the experiment, a total of 288 completed it. The average time spent on the platform was 53 minutes. It is apparent that the length of the experiment played a significant role in deterring certain participants from completing it, with a considerable number abandoning the experiment at the onset of the third task. Nevertheless, the sample size of 288 respondents who successfully reached the end of the experiment remains a reasonable and adequate number to test the hypotheses initially outlined in this monograph, to draw conclusions, and advance cautious generalizations.

In terms of gender distribution, the listeners in this study were predominantly women, accounting for 82.99% of the participants, whereas men

---

4. The term 'monolingual' is used here in contrast it with 'early' or 'late bilinguals'.

comprised 17.01% of the sample[5]. Regarding the age groups, the majority of informants, specifically 82.02%, fell within the first age group (18-30 years). The second age group (31-45 years) accounted for 13.20% of participants, while the third age group (46-65 years) constituted a smaller portion at 2.78%. This distribution is unsurprising, considering that the experimental recruitment took place online, where younger respondents displayed a higher level of responsiveness compared to their older counterparts. The majority of the informants were Bachelor's students and their highest level of education was a high-school diploma (61.80%). There followed Master's students that had a Bachelor's degree (13.19%), and workers that had a Master's degree (11.11%). Another variable is the Italian region where listeners had spent most of their lives. A reasonably well-stratified sample was obtained, encompassing respondents from all Italian regions except Molise. Generally, the distribution of respondents corresponds somewhat to the population size of each region. In fact, most respondents were from Lombardy (17.70%), Piedmont (11.11%), Sardinia (10.07%), Sicily (9.72%), and Apulia (7.29%). In contrast, there were relatively fewer informants from Valle D'Aosta, Basilicata, and Marche.

The respondents' background in linguistics was also considered in this study. To ensure clarity and prevent misinterpretations, the question was formulated as follows: 'Have you ever attended Bachelor's and/or Master's courses in linguistics? N.B.: This question exclusively refers to specific courses in linguistics and should not be confused with foreign language courses.' (in Italian: *Ha seguito corsi di linguistica all'università, durante la Laurea Triennale o la Laurea Magistrale? N.B.: La domanda si riferisce esclusivamente a esami di linguistica, da non confondersi con esami di lingua straniera o di idoneità linguistica.*). The distribution of respondents within the 'linguistics' variable demonstrates a well-balanced representation. Participants who had attended Bachelor's and/or Master's courses in linguistics accounted for 59.37% of the entire sample, while those who had not represented 40.63%.

Another aspect of interest in this study was the respondents' proficiency levels in foreign languages. They were asked to self-assess their knowledge of English, German, French, Spanish, Romanian, and Russian, using a range from '0' (i.e. no proficiency) to '5' (i.e. excellent proficiency). As expected, nearly everyone (99.65%) had some degree of proficiency in English. Similarly, almost 80% of the respondents had some knowledge of French, German, or Spanish. On the contrary, a small percentage of respondents spoke or understood Russian (20.14%) or Romanian (4.86%).

---

5. No other level was reported for the gender variable.

Based on the self-assessed level of foreign languages, it appears that younger listeners possess a higher proficiency compared to older generations, particularly in English, Spanish, and French, as also indicated in Table 5.1. This disparity may be attributed to the fact that younger respondents, unlike their older counterparts, had greater opportunities for study abroad programs or language learning experiences.

Table 5.1: The listeners' language proficiency as a function of the age groups

Language	Age group: 18-30 Level (mean, sd)	Age group: 31-45 Level (mean, sd)	Age group: 46-65 Level (mean, sd)
English	3.85 (±0.95)	3.29 (±1.14)	3.25 (±1.17)
German	1.10 (±1.48)	0.78 (±0.32)	1.37 (±1.68)
French	1.90 (±1.44)	1.76 (±1.34)	1.50 (1.07)
Spanish	1.99 (±1.56)	1.97 (±1.62)	0.75 (±1.04)
Romanian	0.08 (±0.39)	0.19 (±0.69)	0.00 (±0.00)
Russian	0.45 (±0.98)	0.11 (±0.51)	0.13 (±0.35)

Lastly, participants were requested to provide information regarding their familiarity with foreign-accented Italian, with a particular focus on the six accents explored in this study. They were asked to rate their familiarity using a scale ranging from '0' (i.e. no familiarity with that accent) to '5' (i.e. a complete familiarity with that accent). Overall, 98.96% of the listeners had at least some familiarity with the English-accented Italian. Following closely was the French accent, which was familiar to varying extents for 92.36% of the respondents. There followed the Spanish accent (88.54%) and the German accent (78.12%). On the other end of the spectrum, the Russian accent (64.24%) and the Romanian accent (53.47%) received lower familiarity ratings. These findings may appear somehow unexpected considering that Romanian- and Russian-speaking communities represent significant populations in Italy. Nevertheless, this distribution of accent familiarity can be attributed to the composition of the sample of respondents, predominantly comprised of young people who may have had limited exposure to native speakers of Romanian and/or Russian. This trend is further reflected in Table 5.2, which highlights the average familiarity ratings across different age groups. For both Romanian and Russian accents, the

familiarity level remains around 1.00 across all age groups. Conversely, all age groups expressed a notable level of familiarity with Italian spoken with English, French, and Spanish accents.

Table 5.2: The listeners' level of familiarity with foreign accents as a function of the age groups

Accent	Age group: 18-30 Familiarity (mean, sd)	Age group: 31-45 Familiarity (mean, sd)	Age group: 46-65 Familiarity (mean, sd)
English	3.86 (±1.15)	3.32 (±1.28)	3.62 (±0.92)
German	2.30 (±1.77)	1.87 (±1.66)	3.25 (±1.16)
French	2.76 (±1.53)	2.79 (±1.47)	2.88 (±1.25)
Spanish	2.81 (±1.67)	2.97 (±1.60)	2.38 (±1.77)
Romanian	1.08 (±1.36)	1.45 (±1.52)	1.00 (±0.76)
Russian	1.54 (±1.56)	1.24 (±1.38)	1.25 (±1.04)

5.1.2 Tasks and stimuli

The tasks were delivered in this order: the identification of the speakers' L1, the evaluation of the degree of accentedness of the sample, and the subjective assessment of foreignism feature saliency. The informants that qualified for these rounds of the experiment were reminded to activate the audio on their computers and to use headphones for listening to the audio samples. Also, a detailed description of each of the three tasks of the experiment was provided.

Some elements of the experimental design (e.g., the duration of the experiment, the nature and duration of the stimuli, the sequencing of questions, the response scale, etc.) may contribute to priming and a response bias in experiments such as the one delivered for this study. Regarding the duration of the experiment, it was essential to have a substantial number of stimuli for each accent due to the necessity of testing multiple hypotheses. However, to minimise priming, biases, and fatigue, the online platform employed for collecting responses was designed with modern, colourful, and interactive features. Respondents were required to manually navigate from one question to another, ensuring their engagement with the new stimuli. Furthermore, the experiment was structured across six distinct pages, each presenting unique colours and layouts. In each task and in each experimental session, the stimuli were randomised to avoid or at least to reduce

responses and associations that could have emerged due to priming and confirmation bias. The experiment underwent a pilot phase involving five expert informants, namely one full professor in linguistics from the University of Pisa and four Ph.D. students specialising in phonetics and phonology at the same university. These expert informants possessed a strong background in the topic at hand, ensuring the selection of a well-balanced dataset. This was particularly crucial for the second task of the experiment, which involved rating the degree of accentedness in voices. The inter-rater reliability was calculated for both the expert listeners' and the informants' responses, specifically for this purpose. Further elaboration on this will be provided in section §5.3.

To prevent repetition priming, a large set of stimuli was employed for the experiment, comprising 112 different samples (16 for each L1, as detailed in Appendix E) from a total of 98 different speakers (14 for each L1). In both the first and second task, the 91 voices used (13 for each L1) were distinct from one another. This ensured that 49 different speakers were used for the first task, and 42 different speakers for the second task, without any overlap between the two. In the third task, although quantitative assessments were not sought from the listeners, 21 different voices were used, with 7 of them being entirely new, while the remaining 14 were previously presented in the first and second tasks. Nevertheless, the samples of the utterances were all different and they were not presented in other tasks.

The accent identification task involved presenting listeners with stimuli consisting of 10 to 15 seconds of speech and requesting them to identify the L1 of the speaker they heard. Each sample could be played twice, and the options were displayed on the screen in the following sequence: French, Italian, Romanian, Spanish, English, German, and Russian. In order not to influence the listener's choice, the transcription of the audio was not provided. Similarly, to prevent bias, no feedback on the accuracy of the identification was provided. For this specific part of the experiment, 49 samples from 49 distinct speakers were used, with seven samples per accent category, including an equal number of native Italian speakers. Since one objective of this study is to compare human performance in identifying foreign accents with the results obtained with machine learning classification systems, it was necessary to use only samples of read speech. The speakers' second reading attempts were used.

The stimuli were produced by a varied range of speakers and great attention was given to ensure balance across all the variables discussed in this study. As far as the gender distribution is concerned, it mirrored the distribution of the entire corpus, with 63.27% of the stimuli belonging to female voices and 36.73% to male voices. Next, in terms of the age of Italian on-

set, 48.97% of the speakers were firstly exposed to Italian during adulthood, 22.45% during adolescence, and only 14.29% during infancy; of course, Italian speakers, who were exposed to Italian from birth, represent 14.29% of the voices played in the identification experiment. As previously explained, the recording circumstances ensured that none of the speakers qualified as individuals who solely learned Italian through guided methods. Consequently, the variable 'learning method' consists of two levels: 'naturalistic' – referring to those who had never attended conventional, private, or online Italian language classes - and 'guided and naturalistic' – representing individuals who participated in any form of Italian language courses while also being exposed to Italian naturally, as they resided in Italy during the recording period. Excluding the seven samples from native Italian speakers who acquired Italian as their first language, the distribution of samples produced by non-native speakers is as follows: 52.38% learned Italian naturalistically, while 47.62% employed a combination of guided and naturalistic approaches. At the dataset level, 35.71% of the learners received specific pronunciation training, limited, therefore, to some of those who underwent guided training in Italian, while 64.29% did not receive such training. Regarding the duration of time spent in Italy, focusing solely on the group of non-native speakers, the distribution is as follows: 42.85% had resided in Italy for less than a year at the time of recording, 26.19% had lived in Italy between one and two years, and 30.95% had spent more than two years but generally less than five years in Italy. Moving on to the proficiency level in Italian for the accent identification task, only non-native speakers who declared at least a B1 level were selected from the corpus. Within the dataset, the distribution of proficiency levels among non-native speakers is as follows: threshold/intermediate level - B1 (40.48%), vantage/upper-intermediate level - B2 (26.19%), advanced level - C1 (11.90%), and proficiency level - C2 (21.43%).

In the second task of the experiment, the listeners were asked to assess the degree of foreign accent of the voices they heard using a six-gradients Likert rating scale (Likert, 1932). The top of the rating scale ('0') was labelled with 'no foreign accent-native speaker' (in Italian: *nessun accento – parlante madrelingua italiano*), whereas the bottom ('5') was labelled with 'very strong accent' (in Italian: *accento molto forte*). The in-between values were the following: '1 – mild accent' (in Italian: *accento lieve*); '2 – moderate accent' (in Italian: *accento moderato*); '3 – marked accent' (in Italian: *accento marcato*); '4 – strong accent' (in Italian: *accento forte*). The stimuli comprised 10 to 15 seconds of speech samples that could be played twice. To ensure the listener's choice was unbiased, the audio transcription was not provided. For this accentedness rating task, a total of 42 samples from

42 different speakers were used, with 6 samples representing each accent, including an equal number of native Italian speakers. Regarding the speech modalities, 21 samples (50.00% of the entire dataset used for this task) consisted of read speech excerpts, while the remaining 21 samples were elicited as spontaneous speech.

To address the research question concerning the influence of speaker- and style-related factors on the degree of perceived foreign accent, a comprehensive control of all relevant variables was implemented in this task. In terms of gender distribution, the accentedness rating task followed a pattern consistent with the overall corpus and the first task. Specifically, 61.90% of the stimuli featured female voices, while 38.10% were represented by male voices. Among the speakers, 14.29% were native Italian speakers who had been exposed to the language since birth. The remaining informants had different ages of onset: 21.43% started learning Italian during infancy, 23.81% during adolescence, and 40.47% during adulthood. Regarding the method used to learn Italian, the distribution among non-native speakers was evenly split. Specifically, 50.00% acquired the language through naturalistic means, while the other 50.00% used a combination of guided instruction and natural exposure. Then, 38.89% of learners received targeted pronunciation training, while the remaining 61.11% did not receive such training. Next, 44.44% of the speakers had spent less than a year in Italy by the time they were recorded, 19.44% between one and two years, and 36.12% more than two years, and generally less than five years. Just like in the previous task, only samples produced by speakers possessing at least a B1 level in Italian were used for this accentedness rating experiment. Therefore, the proficiency level among non-native speakers is distributed as follows: threshold/intermediate level - B1 (30.56%), vantage/upper-intermediate level - B2 (22.22%), advanced level – C1 (19.44%), and proficiency level – C2 (27.78%). All in all, the data used tend to display a rather balanced group distribution.

In the third and final phase of this experiment, the respondents were asked to provide their subjective feedback regarding the most prominent foreignism features they detected in the audio samples they listened to. These features could be freely inserted in a given blank space. It is worth noting that this particular task has not been extensively used in previous studies exploring foreign accent, particularly not with such a diverse range of non-native accents. The speech samples had an average length of 15 seconds, and they were excerpts of read speech. The transcription of the audio was not provided. For this open-cloze task, a total of 21 samples from 21 distinct speakers were employed, with 3 samples representing each accent, including an equal number of native Italian speakers. Just like in the other

two tasks, attention was given to balancing all the variables discussed earlier in this chapter to guarantee a representative sample encompassing all the accents considered. It is important to note that this task primarily focused on gathering subjective qualitative judgments, so no model was derived from these data. Descriptive statistics will be provided in Chapter 6 to characterise each accent globally, without specifically considering speaker-related factors.

5.2 Accuracy in identifying foreign accents

In the first task of the perceptual experiment, the performance of 288 Italian listeners was assessed in terms of their ability to correctly identify the L1 of 49 speakers. The Italian raters achieved an overall identification accuracy of 0.47, with Cohen's kappa coefficient measuring 0.38[6]. This accuracy level is over three times higher than the chance expectation of 0.14.

The confusion matrix presented in Table 5.3 summarises the listeners' performance for each accent class. Unsurprisingly, the Italian listeners displayed the highest accuracy in recognising samples of native Italian speech, achieving an accuracy/recall of 0.95. In general, the respondents displayed considerable ability in discerning non-native speakers of Italian, successfully identifying them in most of the cases (0.89). However, determining the precise origin of a speaker proved challenging for most accents, with the exception of Spanish-accented Italian, which was correctly identified with a 0.78 recognition rate. It is interesting to note that the Spanish accent was rarely confused with other accents (notably 0.09 labelled as Romanian). For all other accents, the accuracy scores for recognition were significantly lower. In fact, the Russian-accented voices were identified with an accuracy of 0.39, and they were frequently mistaken for the Romanian accent (0.28). As discussed in Chapter 2, this confusion can be attributed to a general mix-up between these two accents, possibly influenced by media portrayals. It is common for media references to use phrases like 'Slavic accent' or 'Eastern-Europe accent' to denote someone from an unspecified country in Eastern Europe, generating confusion among the naïve public. The German accent had a recognition rate of 0.38, and it is interesting that a quarter of the listeners misidentified it as French (0.23). This confusion could potentially be attributed to the fact that the French and German speakers whose samples were played in this task faced similar pronuncia-

6. The reliability of a Cohen's kappa of 0.38 is considered 'fair' (see Cohen, 1960; Landis and Koch, 1977).

tion challenges, such as the production of the Italian alveolar trill /r/ (see Chapter 6). The German accent was also occasionally labelled as English (0.12). Surprisingly, even if almost all listeners claimed that they spoke English and that they were familiar with the English-accented Italian, the English accent was identified correctly only in a limited number of cases (0.30). A significant number of respondents confused it with the Romanian accent (0.18) and with the German accent (0.15). The French accent, as well, was recognised with a low recognition rate (0.30), although most speakers claimed that they were familiar with it. Many participants believed that the voices they heard were Romanian (0.20) or Spanish (0.14). Finally, recognising the Romanian accent proved to be extremely challenging. In fact, it was identified correctly only on a few occasions (0.22), while there were frequent cases in which Romanian speakers were confused with Italian speakers (0.27). The Romanian accent was also labelled as Spanish (0.16) or English (0.13).

Table 5.3: Confusion matrix for the results achieved by Italian listeners[7]

	ENG	FRA	GER	ITA	ROM	RUS	SPA
ENG	**0.30**	0.04	0.15	0.07	0.18	0.13	0.12
FRA	0.07	**0.30**	0.09	0.09	0.20	0.12	0.14
GER	0.12	0.23	**0.38**	0.07	0.08	0.07	0.05
ITA	0.01	0.01	0.00	**0.95**	0.01	0.00	0.02
ROM	0.13	0.05	0.08	0.27	**0.22**	0.09	0.16
RUS	0.06	0.02	0.06	0.15	0.28	**0.39**	0.04
SPA	0.03	0.02	0.03	0.02	0.09	0.03	**0.78**

One of the hypotheses of this study is that the perception and the identification of foreign accents depends on listener-related factor. To examine this claim, various statistical analyses were conducted. Firstly, the impact of listener variables – such as gender, age group, level of education, occu-

7. ENG = English accent, FRA = French accent, GER = German accent, ITA = Italian accent, RUS = Russian accent, SPA = Spanish accent. The actual class is displayed on the left with respect to the table intersection axes, while the predicted class is displayed on the right. Therefore, the classification matrix should be read on the rows (e.g. English-accented samples were misinterpreted as French in 4% of the cases). The diagonal of the matrix displays the identification score for each accent (e.g. English-accented samples were identified correctly in 30% of the cases).

pation, region of origin, and background in linguistics – on the overall accuracy of foreign accent identification was assessed. Subsequently, the focus shifted to analysing each accent individually. The Italian group of speakers was excluded from the models due to perfect multicollinearity, resulting in aliased coefficients. Additionally, considering that the native Italian varieties were correctly recognised most of the time (0.95), their exclusion from the statistical modelling was deemed appropriate.

Before conducting regression models, a preliminary analysis was performed using descriptive statistics to gain initial insights into the median and mean values of accent recognition rates. When examining gender, both male and female listeners displayed similar performances, as indicated by the identical medians and means of 0.47. Therefore, it appeared unlikely that gender would significantly contribute to estimating the overall accent recognition rate. Regarding age groups, the overall performance was relatively consistent. However, the two age groups below 45 achieved slightly better results, with means and medians around 0.48. In contrast, older respondents exhibited a mean of 0.46 and a median of 0.43 in correctly identifying the foreign accent. Moreover, the descriptive analyses on the accuracy of accent identification revealed an interesting trend related to the listeners' level of education and occupation. This first exploration suggested that higher qualification levels corresponded to better recognition of the correct accent. Listeners with a Master's degree showed a mean accuracy value of approximately 0.51, whereas individuals with a Bachelor's degree had a mean accuracy value of 0.49, and those with a High-school diploma a mean accuracy value of 0.46. In terms of occupation, the mean values for accent identification score varied as well. Employed respondents achieved a mean accuracy score of 0.51, while unemployed respondents and students exhibited mean accuracy scores of 0.48 and 0.47, respectively. This preliminary prompted further investigation into the interaction between education level and occupation. Interpreting the average accent identification scores based on the listener's region of origin posed challenges due to the disparity in the number of respondents within each group. For instance, the Marche group, which exhibited one of the highest recognition rates (0.60), comprised only five listeners. Finally, the accuracy score demonstrated variation between listeners who attended linguistics courses and those who did not. Unsurprisingly, the group with a background in linguistics performed better, displaying a mean value of 0.49, whereas the other group achieved a mean value of 0.40.

Following a treatment coding of the categorical variables, the statistical modelling examined the impact of listener-related factors on estimating the overall accent identification score. This was accomplished by employing

Ordinary Least Squares (OLS) multiple linear regression models. The initial model (Model 1) included all possible explanatory variables, namely the listener's gender, region of origin, age group, level of education, occupation, and background in linguistics. The response variable in this model was the overall accent identification score. In Model 2, the gender variable was dropped as its contribution to estimating the accent identification score was found to be insignificant. Similarly, in Model 3, the age group variable was removed, followed by the exclusion of the occupation variable in Model 4, as their respective effects on the response variable were not statistically significant. Additionally, the interaction between the level of education and occupation was tested, but it did not yield any improvement to the model. Table 5.4 summarises the results obtained for the four models.

Table 5.4: Summary of the four multiple linear regression models

	Model 1	*Model 2*	*Model 3*	*Model 4*
Response variable	accent identification score	accent identification score	accent identification score	accent identification score
Explanatory variables	the listener's gender, age group, occupation, region, education, background in linguistics	the listener's age group, occupation, region, education, background in linguistics	the listener's occupation, region, education, background in linguistics	the listener's region, education, background in linguistics
Observations	288	288	288	288
Residual standard error	0.0877 on 261 degrees of freedom (sigma: 0.185)	0.0876 on 262 degrees of freedom (sigma: 0.185)	0.0875 on 264 degrees of freedom (sigma: 0.185)	0.0872 on 266 degrees of freedom (sigma: 0.184)
Multiple R-squared	0.217	0.217	0.212	0.211
Adjusted R-squared	0.139	0.142	0.143	0.148
F-statistic	2.78 on 26 and 261 degrees of freedom, p-value: 2.03e-05	2.9 on 25 and 262 degrees of freedom, p-value: 1.17e-05	3.09 on 23 and 264 degrees of freedom, p-value: 6.36e-06	3.38 on 21 and 266 degrees of freedom, p-value: 2.12e-06

To sum up, Model 4 included the listeners' region of origin, education level, and background in linguistics as predictor variables. The adjusted R-

squared value for Model 4 was 0.148, indicating that 14.80% of the variance in the accent identification scores could be predicted by these three variables. Interestingly, employees holding a Master's degree and having a background in linguistics outperformed all other participants. To assess the linear model assumptions, a global test was conducted with 4 degrees of freedom, considering a significance level of 0.05. The assumptions for Global Stat, Skewness, Kurtosis, Link Function, and Heteroscedasticity were found to be acceptable. As shown in Table 5.5, the Anova Type II Test conducted on this model revealed significant effects of the region of origin, education level, and attending linguistics courses on the listeners' performance in recognising non-native accents in Italian. However, it is important to note that while these factors accounted for some of the variation in the data, their explanatory strength was limited. At the same time, the observed high significance of the region of origin should be interpreted with caution due to the uneven distribution of listeners across the levels of this variable. It is possible that certain regions, such as Marche, Trentino Alto-Adige, or Valle D'Aosta, which also had a smaller number of respondents, displayed higher identification scores purely by chance, rather than due to any underlying patterns.

Table 5.5: Anova Type II Test for Model 4

	Sum of squares	*Degrees of freedom*	*F value*	*Pr (>F)*
Region	0.371	18	2.71	0.00029 ***
Education	0.059	2	3.88	0.02174 *
Linguistics	0.083	1	10.85	0.00112 **
Residuals	2.024	266	-	-

Besides the analyses above, the role of listener-related factors in predicting scores for each of the six foreign accents was examined as well. While regression models were used to assess the impact of all variables, for the sake of brevity, only significant results or those that hold linguistic significance will be discussed below.

Speaking the Russian language, having some familiarity with the Russian accent, and having attended courses in linguistics collectively accounted for 12% of the variation in the identification scores for the Russian accent, $(F_{(3,284)} = 13.18, p = 0.00; R^2 = 0.13;$ Adj. $R^2 = 0.12)$. This finding was further supported by the Anova Type II Test, that indicated significant

effects for all three variables (p < 0.05). Similarly, proficiency in German, a strong familiarity with the German accent, and a background in linguistics explained 15% of the variation in the identification scores for the German accent (F(3,284) = 16.96, p = 0.00; R² = 0.15; Adj. R² = 0.15). This finding was also corroborated by the Anova Type II Test, with significant p-values (p < 0.05) observed for all three predictors. None of the variables tested reached statistical significance in predicting the variation in the identification scores for the English, French, Romanian, and Spanish accents. While respondents proficient in these languages or with familiarity in the respective accents, as well as those with a background in linguistics, tended to perform better, these associations lacked strong statistical significance (p > 0.05).

Moving forward with the analysis, the accent identification scores were also examined with regards to speaker-related variables, first by means of descriptive statistics and afterwards with statistical modelling. As previously mentioned, Italian voices were excluded from these analyses due to both perfect multicollinearity and their high recognition rate of 0.95.

The variation in accent identification scores based on the speaker's L1/accent was presented in a confusion matrix (Table 5.3) and discussed earlier in this section. It was observed that certain accents, such as the Spanish accent, were more easily recognisable compared to others. Next, the mean accent recognition rates for male and female speakers showed a slight difference, with males having a mean value of 0.42 and females a mean value of 0.38. Male speakers tended to be more accurately identified as native speakers of their respective languages. Then, the L1 of the speakers who were exposed to Italian later in life (during adulthood) was identified more accurately compared to the L1 of the speakers who learned Italian at an early age (mean values of 0.32, 0.25, and 0.48 for infancy, adolescence, and adulthood, respectively). Moreover, it appears that speakers who learned Italian uniquely in a naturalistic setting, without any guided instruction or pronunciation training, were more easily identified as non-native speakers of Italian and had their correct accent recognised more accurately (mean value of 0.48 and a median of approximately 0.45) compared to those who underwent a mixed method of language acquisition. For the latter group, the median score was 0.40 for speakers who did not attend any pronunciation classes, while it was 0.30 for those who did (with mean values around 0.30 for both subgroups). At the same time, speakers who stayed in Italy for shorter periods (i.e., less than 12 months) had their respective accents recognised more easily compared to those who spent longer periods in

Italy (mean values of 0.51, 0.35, and 0.28 for 6-12 months, 12-24 months, and more than 24 months, respectively). Lastly, there was a relationship between the level of proficiency in Italian and the recognition of the speaker's foreign accent, as lower proficiency levels corresponded to higher recognition rates (mean values of 0.57, 0.30, 0.33, and 0.21 for B1, B2, C1, and C2 levels, respectively).

After the treatment coding of the categorical variables, OLS multiple linear regression models were employed to analyse the effect of speaker-related factors on the estimation of the accent identification score. The initial model (Model 1) included all possible explanatory variables such as the speaker's L1/accent, gender, age of onset, method of learning Italian, pronunciation training, length of stay in Italy, and proficiency level in Italian. The response variable was the accent identification score. Model 1 accounted for 56.60% of the variation in accent recognition rates. In Model 2, the variables that did not significantly contribute to estimating the response variable were excluded. Consequently, the revised model included only the speakers' L1/accent and proficiency level in Italian as predictors, which accounted for 58.60% of the variation in the accent identification scores. The linear model assumptions, including Global Stat, Skewness, Kurtosis, Link Function, and Heteroscedasticity, were assessed using a global test on 4 degrees of freedom at a significance level of 0.05. The results demonstrated that these assumptions were acceptable. The summary of the multiple regression model is displayed in Table 5.6.

Table 5.6: Summary of the two multiple linear regression models

	Model 1	*Model 2*
Response variable	accent identification score	accent identification score
Explanatory variables	the speaker's gender, age of onset, method of learning Italian, pronunciation training, length of stay in Italy, and proficiency level in Italian, L1/accent	the speaker's proficiency level in Italian, L1/accent
Observations	42	42
Residual standard error	0.175 on 26 degrees of freedom (sigma: 0.371)	0.171 on 33 degrees of freedom (sigma: 0.362)
Multiple R-squared	0.725	0.667
Adjusted R-squared	0.566	0.586
F-statistic	4.56 on 15 and 26 degrees of freedom, p-value: 3.62e-04	8.25 on 8 and 33 degrees of freedom, p-value: 4.54e-06

The Anova Type II Test conducted on Model 2 indicated that both the speaker's L1/accent and her/his proficiency level in Italian had statistically significant effects on the accent recognition rates (Table 5.7). However, it is important to note that while these variables had indeed a significant effect, they did not fully account for all the observed variation in the data.

Table 5.7: Anova Type II Test for Model 2

	Sum of squares	Degrees of freedom	F value	Pr (>F)
L1/accent	0.969	5	6.60	0.00023 ***
Proficiency	0.560	3	6.36	0.00160 **
Residuals	0.968	33		

5.3 Degree of perceived foreign accent[8]

In the previous chapters, the discussion touched upon how foreign accents can serve as indicators of identity, providing insights into the speakers' origins, as well as conveying information about their socio-demographic and socio-cultural background. Furthermore, the degree of accentedness was suggested to indicate one's receptiveness towards the language and culture of the host country, or it could serve as a predictor of the quality and quantity of input for the specific language being learned (Flege, 2008).

This section takes an innovative approach compared to previous studies on L2 Italian and draws inspiration from similar research on non-native English (e.g., Piske et al., 2001). The objective is to examine the influence of listener-, speaker-, and style-related factors on the perceived degree of foreign accent in Italia. Rigorous control measures were implemented to ensure the suitability of the stimuli for the statistical analyses, encompassing variables such as the style of the speech sample (i.e., read and spontaneous), the speaker's L1/accent, gender, self-assessed proficiency level in Italian, length of stay in Italy, age of onset, predominant language learning method, and the presence or absence of specific pronunciation training. The assessment of sociopsychological factors is beyond the scope of this study, as the recorded sample consisted of Bachelor's, Master's, Ph.D., and spe-

---

8. Part of this section was published as Combei and Marotta (2019) in an edited volume.

cialising students, all young adults, resulting in limited variation. To thoroughly explore the impact of these variables on pronunciation skills in L2 Italian, future research, incorporating a more diverse speaker sample, is needed.

As mentioned at the beginning of this chapter, the accentedness rating task involved samples of read and spontaneous speech from a group of 42 speakers, with 6 speakers representing each L1/accent. Prior to the task, the stimuli underwent validation by five expert listeners. Using a 6-point Likert scale, the experts were asked to assess the degrees of accentedness of the selected stimuli. The inter-rater reliability was calculated, resulting in a high level of agreement that confirmed the adequacy of the samples chosen and ensured a balanced representation across various degrees of accentedness: ICC (A, 1) = 0.76[9]; p-value < 0.05; 95% – confidence interval. Subsequently, the 42 speech samples were presented to the 288 respondents, who were instructed to evaluate the level of foreign accent using the same 6-gradients Likert scale. Before discussing the findings of this experiment, it is important to note that the results of the inter-rater reliability test conducted on the listeners' judgments yielded favourable outcomes, ICC (A,1) = 0.68[10]; p-value: 0; 95% – confidence interval, indicating a substantial level of agreement among the listeners' ratings.

First, the descriptive statistics highlighted noticeable variations in the accentedness ratings among the six accents, with some accents, such as Spanish and German, being perceived as stronger than others, even when controlling for all the other variables. Also, the analysis revealed that the spontaneous speech samples received lower accentedness ratings compared to the read speech samples.

The linear regression model incorporating all listener-related predictors (gender, age group, education level, region of origin, occupation, and background in linguistics) did not provide significant results for explaining the variation in the accentedness scores: $F_{(26,261)} = 1.07$, $p = 0.37$; $R^2 = 0.096$; Adj. $R^2 = 0.006$; Residual standard error: 0.4601 on 261 degrees of freedom. The only factor that exhibited some effect on the response variable was the listeners' education level (p-value: 0.03). Generally, higher education levels were associated with more critical evaluations of the speakers' degree of accentedness (mean scores of 2.42, 2.52, and 2.59 for high school, Bachelor's, and Master's, respectively). The final model's results can be summarised as follows: $F_{(4,283)} = 2.35$, $p = 0.05$; $R^2 = 0.03$; Adj.

---

9. According to Cicchetti's (1994) guidelines for the interpretation of inter-rater agreement measures, an ICC value of 0.76 is considered 'excellent'.

10. According to Cicchetti's (1994) guidelines for the interpretation of inter-rater agreement measures, an ICC value of 0.68 is considered 'good'.

R² = 0.02; Residual standard error: 0.46 on 283 degrees of freedom. All in all, despite removing several listener-related factors (region, age group, gender, and linguistic background), the model did not improve its predictive strength.

This study also examined the degrees of accentedness among the six accents, analysing them in relation to three variables associated with the listeners: background in linguistics, proficiency level in foreign languages, and familiarity with the accents considered. The various levels of these listener-related variables did not display significantly different values for the degree of accentedness of the Russian speakers (all of them reported scores around 2.00, which according to the rating scheme at hand suggests a moderate accent). The means and the medians of the ratings given by the informants that attended linguistics courses were similar to those given by informants that did not study it. ANOVA tests for each of these factors were performed yielding significant results only for the proficiency level in Russian (p-value: 0.03).

Then, the proficiency level in English and familiarity with the English accent did not show significant differences in the perceived degrees of English-accented Italian. The ratings for all levels of these variables were consistently around 2.70, indicating a marked accent. However, informants with a background in linguistics tended to judge the accentedness levels of English speakers less severely, although these differences were not statistically significant. The ANOVA analyses conducted on the three variables did not yield statistically significant results (p-values > 0.05) for any of the factors.

Similarly, the proficiency level and familiarity with the German accent did not significantly influence the perception of German-accented Italian. Ratings for all levels of these variables were around 3.00, indicating a marked foreign accent. Informants with a background in linguistics showed slightly less severity in judging the German accent, but these differences were not statistically significant (p-values > 0.05).

Next, proficiency level in French and familiarity with the French accent did not significantly affect the perception of French-accented Italian. Overall ratings for all the levels of these variables were around 2.70, indicating a marked foreign accent. Informants with a background in linguistics and those without provided similar ratings, as confirmed by non-statistically significant ANOVA results (p-values > 0.05).

The perception of Romanian-accented Italian did not significantly vary based on the informant's proficiency level in Romanian, and no particular pattern was observed in the values. Familiarity with the Romanian accent led to slightly more severe judgments among the few informants that were

completely familiar with it, but an ANOVA indicated that the difference was not statistically significant (p-value > 0.05). Similarly, there were no significant differences between the ratings given by respondents with or without a background in linguistics, with all ratings averaging around 1.50. This suggests that all respondents perceived the speech samples as having mild to moderate degrees of foreign accent.

Finally, the degree of perceived Spanish accent remained consistent regardless of the informant's proficiency level in Spanish, with ratings averaging around 3.00, indicating a marked accent. However, informants who were completely familiar with the accent tended to provide slightly more critical judgments. Additionally, respondents with a background in linguistics showed slightly less severity in evaluating the degree of the Spanish accent, but none of these differences were statistically significant according to the ANOVA analyses (p-value > 0.05).

Descriptive analyses and statistical models were conducted to examine the impact of style (spontaneous speech vs. read speech) and various speaker-related factors (L1, gender, age of Italian onset, Italian learning method, length of stay in Italy, presence or absence of specific pronunciation training, and proficiency level in Italian) on the perception of foreign-accented Italian. As expected, Italian speakers were consistently perceived as native speakers, with a low average accentedness score of 0.23 for read speech samples and 0.22 for spontaneous speech samples. To ensure that results were not affected by multicollinearity issues, the group of native Italian speakers was excluded from the analyses.

First of all, in terms of gender, female speakers were assigned an average accentedness score of 2.89 for read speech, while male speakers received a score of 2.72. For spontaneous speech, the average score for female speakers was 2.17, compared to 2.01 for male speakers. Regarding the age of Italian onset, there was an evident difference in ratings among the three groups, particularly in read speech, with an approximate difference of 1.00 point. Speakers exposed to Italian in infancy were generally perceived as having a milder accent, while those with different ages of onset received more severe ratings. The average accentedness scores for infancy, adolescence, and adulthood were 1.6, 2.51, and 3.62, respectively. The difference in mean scores was less evident in spontaneous speech, possibly due to some outliers, but the median for the infancy group indicates that some speakers who learned Italian as children received ratings similar to those of native speakers (ranging from 0 to 1). Moreover, there was notable variation in the ratings given to speakers who learned Italian through guided and naturalistic approaches compared to those who did not attend Italian classes. The former group was consistently perceived to have a less noticeable

accent when speaking Italian, both in read speech (mean scores of 3.56 for naturalistic learners and 2.21 for guided and naturalistic learners) and in spontaneous speech (mean scores of 2.67 for naturalistic learners and 1.44 for guided and naturalistic learners). Likewise, the speakers who underwent targeted pronunciation training exhibited a more natural-sounding Italian pronunciation than the other group, both in read speech (with an average accentedness score of 1.58 for those who received pronunciation training, compared to 3.6 for those who did not) and spontaneous speech (with an average accentedness score of 1.08 for trained individuals, compared to 2.79 for those without training). In terms of the time spent in Italy, speakers who stayed in the country for more than two years received more favourable ratings (with scores of 2.33 for read speech and 1.20 for spontaneous speech). However, no noticeable distinction was found between speakers who stayed for less than one year and those who stayed for two years or less. Lastly, in relation to the Italian proficiency levels of the speakers, those who self-assessed their skills at B1, B2, and C1 levels received similar scores for accentedness in their read speech productions (scoring above 3.00, indicating a marked to strong accent), whereas those who claimed to have a C2 level in Italian were rated as speaking with a native-like or slight accent. When it came to spontaneous speech, speakers with C1 and C2 levels displayed slightly less accented speech compared to the other two groups.

Following the methodology outlined for the preceding statistical analyses in this chapter, the impact of style- and speaker-related variables on predicting accentedness scores was evaluated. First, Model 1 (Table 5.8) was constructed, incorporating various explanatory variables such as the style, the speaker's L1/accent, gender, age of onset, method of learning Italian, pronunciation training, length of stay in Italy, and proficiency level in Italian. The response variable was the accentedness score assigned by the listeners. This model demonstrated the ability to explain 77.00% of the variance observed in the ratings of accentedness. Based on the findings of Model 1, which align with the existing literature on this subject as discussed in Chapter 2, the gender variable was found to be an ineffective predictor for the degree of accentedness. Consequently, the variable was excluded from the analysis, so Model 2 (Table 5.8), accounted for 78.1% of the variation observed in the accentedness scores. In the final Model 3 (Table 5.8), the age of onset was removed as it emerged as the second weakest predictor. While there was observed variation in the ratings across the three age groups (infancy, adolescence, and adulthood), as seen from the descriptive analyses, this difference did not reach statistical significance. This suggests that starting to learn Italian at a younger age (e.g., adolescence vs. adulthood) does not necessarily guarantee the ability to produce native-like

speech. Therefore, Model 3 accounted for 79.3% of the variation in the accent rating, and the relevant fit information is summarised in Table 5.8. The linear model assumptions were evaluated through a global test on 4 degrees of freedom, with a significance level of 0.05. The results indicated that the assumptions for Global Stat, Skewness, Kurtosis, Link Function, and Heteroscedasticity were satisfied, implying that the model met the necessary assumptions for reliable analysis.

Table 5.8: Summary of the three multiple linear regression models

	Model 1	*Model 2*	*Model 3*
Response variable	level of accentedness	level of accentedness	level of accentedness
Explanatory variables	style, the speaker's gender, L1/accent, age of onset, method of learning Italian, pronunciation training, length of stay in Italy, and proficiency level in Italian	style, the speaker's L1/accent, age of onset, method of learning Italian, pronunciation training, length of stay in Italy, and proficiency level in Italian	style, the speaker's L1/accent, method of learning Italian, pronunciation training, length of stay in Italy, and proficiency level in Italian
Observations	36	36	36
Residual standard error	0.626 on 19 degrees of freedom (sigma: 0.254)	0.611 on 20 degrees of freedom (sigma: 0.247)	0.594 on 22 degrees of freedom (sigma: 0.240)
Multiple R-squared	0.875	0.875	0.870
Adjusted R-squared	0.770	0.781	0.793
F-statistic	8.33 on 16 and 19 degrees of freedom, p-value: 1.678e-05	9.34 on 15 and 20 degrees of freedom, p-value: 5.346e-06	11.34 on 13 and 22 degrees of freedom, p-value: 6.578e-07

The Anova Type II Test conducted on Model 3 (Table 5.9) revealed a significant difference in the degree of perceived accentedness between read speech and spontaneous speech style (p-value: 0.04). This finding suggests that read speech samples were generally perceived as more foreign-accented than spontaneous speech samples. One possible explanation for

this result could be the occurrence of reading errors resulting from the differences in orthographic norms between the speakers' native language and Italian. Additionally, it is worth noting that not all non-native speakers in the study had achieved full proficiency in written Italian by the time they were recorded; this observation is particularly true for those who had been exposed to the language for less than 12 months, such as exchange students. This finding aligns with previous research on foreign-accented speech, including studies by Oyama (1976), Thompson (1991), and Kolly and Dellwo (2013), that consistently reported that read speech tended to be perceived as more accented compared to spontaneous speech. However, this goes against the stylistic corollary in Major's (2001) OPM, that claims that as style becomes more formal, L2 increases, while on the contrary, L1 decreases.

Table 5.9 – Anova Type II Test for Model 3

	Sum of squares	Degrees of freedom	F value	Pr (>F)
L1/accent	8.143	5	4.61	0.005 **
Style	1.659	1	4.70	0.040*
Italian learning method	1.526	1	4.30	0.059
Specific pronunciation training	4.357	1	12.33	0.001**
Length of stay in Italy	1.287	2	1.82	0.181
Proficiency level in Italian	5.557	3	5.24	0.010*
Residuals	7.775	22		

When controlling for all other grouping variables and levels, Romanian speakers consistently received ratings of 'no accent' or 'mild accent' and rarely ratings of 'marked accent'. Conversely, the samples produced by German speakers were perceived as the most foreign-accented. On a scale of accentedness, the German speakers were followed in this order by Spanish, French, English, and Russian speakers. These findings are consistent with the SLA literature that has linked pronunciation deviations to blocking mechanisms ascribable to L1 (Flege, 1995; Best and Tyler, 2007; Kuhl et al., 2008). Indeed, based on these results, it appears that the speaker's L1

may predict the degree of foreign accent (p-value: 0.005). Additionally, speakers who had received specific pronunciation training were perceived as having a lesser degree of foreign accent compared to those who had not (p-value: 0.001). This finding supports the argument for shifting focus towards the role of input in SLA, advocated, among others by Flege (2008, 2018), emphasising the significance of appropriate training to overcome pronunciation issues. While the method of learning Italian (guided and naturalistic vs. naturalistic) appeared to have only a limited effect on perceived accentedness (p-value: 0.06), the inclusion of targeted pronunciation training resulted in improved pronunciation and reduced foreign accent. This result reinforces the notion that input plays a crucial role in SLA and highlights the importance of targeted pronunciation instruction. The length of stay in Italy had a limited effect on predicting the degree of accentdness (p-value: 0.18). This could be attributed to the relatively small difference in duration among the groups (i.e., only 12 months), which may not have been substantial enough to produce a noticeable effect. However, it is worth mentioning that speakers who stayed in Italy for more than two years tended to receive less severe judgments for their accent, compared to the other two groups. The self-assessment of speakers' proficiency level in Italian may introduce limitations to the reliability of this predictor in relation to perceived foreign-accentedness. However, it is noteworthy that speakers who claimed to have a C2 level in Italian received higher scores ranging from 0.00 to 1.00, suggesting a pronunciation closely resembling that of a native speaker (p-value: 0.01). On the other hand, speakers with a B1 proficiency level in Italian obtained similar ratings to those with a B2 level, indicating that the difference between these two levels is not statistically significant.

Overall, these empirical findings indicated that various factors significantly influenced the perceived degree of accentedness in non-native Italian speech. Specifically, these were the speakers' L1 (with certain accents being perceived as more pronounced than others), their proficiency level in Italian (speakers with a C2 proficiency level exhibited a milder or negligible accent), the presence of specific pronunciation training (speakers who received such training tended to receive lower ratings of accentedness), and the speaking style (spontaneous speech consistently received lower accentedness ratings compared to read speech). These predictors demonstrate greater significance compared to other variables investigated in this study, such as gender, age of onset, and length of stay in Italy.

6. Decoding foreign-accented Italian

As discussed in the previous chapters of this book, healthy infants have the ability to perceive, differentiate, and acquire the sounds and prosody of all languages during the first few months of their lives (Eimers et al., 1971; Griever and Kuhl, 1989; etc.). By the age of one, children begin to develop an awareness of the phonetically relevant speech contrasts in their native language, while potentially disregarding others (Werker et al., 1981; Kuhl, 1992, 1994; Werker and Tees, 2002; etc.). However, as argued by Flege (1995) and Piske (2008), I believe that sensitivity towards subtle phonetic cues of the L2 is not entirely lost over time, but rather becomes overshadowed.

When acquiring a new language, individuals may unconsciously apply the phonological processes and orthographic norms of their L1when speaking or reading in the L2 (as discussed in Chapter 1). For instance, adult learners of L2 Italian who are native Romanian speakers and who have received little input for L2 Italian may struggle to perceive and produce the palatal nasal, which is absent from their L1's phonemic inventory. As a result, they may substitute it with the sequence [n+j] in words such as "*montagna*"[1]. This production may be perceived as non-native by Italian L1 speakers, potentially leading them to infer that the person they are hearing is an L2 speaker of Italian.

In fact, nearly everyone can recognise when their language is spoken with an unfamiliar accent, and typically, non-native speech is perceived within a matter of milliseconds (Flege, 1984). However, as demonstrated by the experiments in Chapter 5, it is often challenging to pinpoint a person's precise origin solely through listening to their speech. Confusions and uncertainties arise when attempting to differentiate among various accents. Furthermore, Chapter 5 showed that the identification rates of non-native accents and their perceived level of markedness are not necessarily influenced by factors specific to the listener, such as linguistic skills or level of

1. Translation into English: 'mountain'.

education. Similarly intriguing is the finding that the listener's familiarity with the speaker's accent does not always appear to have a significant impact on accent identification scores. As previously noted in the literature, in cases where the accent cannot be correctly recognised, a general perception of foreignness may be attributed to the speech production (Gluszek and Dovidio, 2010; Roessel et al., 2019).

Thus, being able to accurately identify a specific foreign accent requires attention. Beginning with the responses provided by the participants in the third task of the accent perception experiment, the first part of this chapter (§6.1) focuses on mapping the foreignism features of the six accents under investigation. Next, the second part of the chapter examines the performance of several straightforward machine learning accent classification methods. Specifically, one method is based on spectral features (§6.2), while others are developed around prosodic features (§6.3 and §6.4). The objective is to determine whether these methods return results comparable to those obtained by native Italian speakers when it comes to the identification of foreign accents. Furthermore, clustering experiments will be conducted in §6.5. Finally, the findings of all the experiments will be thoroughly discussed in §6.6.

6.1 Salient features of perceived foreign accent

Numerous segmental and suprasegmental cues contribute to the perception of foreign-accented Italian. Chapter 2 presented several studies that investigated this phenomenon. Previous research consistently suggests that both segmental and suprasegmental features play an equally significant role in generating the perception of foreign-accented Italian, but that one prominent characteristic of foreign-accented Italian is the improper use of consonant gemination. In terms of acoustics and auditory perception, geminated consonants in 'standard' Italian are roughly twice as long as their corresponding single consonants (Bertinetto and Loporcaro, 2005). Therefore, uttering geminated consonants as singletons is perceived as deviant, since they are contrastive, as for instance in *farro*[2] /ˈfarro/ vs. *faro*[3] /ˈfaro/. Empirical studies have shown that the most salient phonetic correlate of the phonological contrasts between singletons and geminates is the closure duration. For instance, it seems that, on average, long stops have between 1.5 to 3 times the closure duration of short stops (Ladefoged and Maddieson,

2. Translation into English: 'hulled wheat'.
3. Translation into English: 'lighthouse'.

1996). Besides the lack of consonant gemination, other features have been mentioned for the perceived foreign-accented Italian (e.g. vowel lowering; rolled alveolar trills, non-native rhythmic patterns; etc.).

It was mentioned that an L1 speaker is able to perceive whether her/his language is spoken with a foreign accent in a few milliseconds. Chapter 5 further reinforced this claim, revealing that the differentiation between native and non-native speech is often readily apparent to native speakers. In fact, the listeners involved in this study successfully identified native varieties of Italian with a 0.95 accuracy. Additionally, they showed an aptitude for recognising non-native varieties, achieving an accuracy rate of 0.89, even though pinpointing the exact origin of a speaker proved to be a challenge. This section aims to elucidate the factors that contribute to this distinct differentiation between native and non-native Italian speech. The task was entrusted to 288 native speakers of Italian that had to freely report which features contributed to the perception of foreignness (see §5.1, for the experimental design).

While the primary emphasis of this monograph revolves around the perception of foreign-accented Italian, it is worth noting that the responses provided by native Italian speakers could potentially offer insights into how learners approach new and similar (yet not identical) sounds. In fact, upon commencing this work, one of the expectations was that at least some non-native speakers would face difficulties in realising the consonant gemination in Italian, a non-native feature for many of the speakers considered here. At the same time, due to 'equivalence classification' (Flege, 1987, 1995, 2007), it was hypothesised that similar sounds between the speakers' L1s and Italian could be problematic both in terms of perception and production. Similarly, regarding prosody, it was assumed that the prosodic patterns of non-native speakers would be perceived as somewhat divergent from those of native varieties of Italian. This expectation arises from the basic understanding that the prosody of one's L1 can influence, at least to some extent, the prosodic production in the L2. For example, it was expected that the potential errors in lexical stress placement in Italian for French speakers would have been noted by the listeners; similar reactions were expected concerning the rhythm of Russian speakers in Italian. In general, the influence of L1 on L2 prosody was likely to manifest in the perception and evaluation of the speech of non-native Italian speakers.

Since the evidence presented in this section is gathered by means of subjective assessments provided by native speakers, I do not claim to test any model of L2 acquisition. However, this type of exploration of features might be useful as a starting point for future instrumental analysis on non-native Italian speech. In order to better control the data and to make it com-

parable between the accents and the speakers, for this task, only samples of read speech were used. I acknowledge that read speech, unlike spontaneous speech, might allow to gain less insights, for instance, into prosody. At the same time, however, read speech facilitates the detection of errors arising from conflicting orthographic norms and conventions between the speakers' native languages and Italian (Wottawa and Adda-Decker, 2016).

Detailed theoretical descriptions of L2 Italian speech have already been provided by Canepari (2007), but the novelty of this research lies in its empirical multi-method approach. Thus, the study encompasses the use of authentic speech samples from a diverse range of speakers, as well the involvement of listeners tasked with describing the features associated with foreign-accented Italian in the voices they heard. The experiment yielded a substantial number of responses, as will be demonstrated shortly. However, due to the nature of the data, any predictions should be approached cautiously. It is worth noting that this task was mandatory, similar to other sections of the experiment. Nevertheless, as it was presented in the form of an open-cloze exercise, some participants opted to input nonsensical characters or letter sequences to evade providing a meaningful response. Also, some answers consisted in single words that appeared in the speech sample, without any further clarification (e.g. *comitato*[4]), or they were simply ironic comments (e.g. *Parla come Papa Francesco*[5]). Consequently, the 6,048 responses were carefully filtered to exclude any ironic or nonsensical comments. A total of 3,486 meaningful responses remained, each addressing one or more characteristics of foreign-accentedness. These responses collectively constituted a corpus of over 30,000 words.

In Chapter 5, when describing the sample of listeners, it was shown that an appropriate number of participants was reached for both levels based on their background in linguistics. This includes individuals who received formal education in linguistics, as well as those who did not. This meant that the way listeners responded in the third task of the experiment varied to some extent. In fact, naïve listeners tended to provide rather vague interpretations of the features that they perceived as the most salient (e.g. *Gli accenti delle parole sbagliati o un po' indecisi, la mancanza di espressività della frase perché probabilmente fa ancora troppa fatica a pensare a cosa dice per preoccuparsi di come lo dice; alcune lettere come c e t hanno un lieve pronuncia straniera, forse inglese*[6]). On the contrary, the participants

---

4. Translation into English: 'committee'.
5. Translation into English: 'He speaks like Pope Francis'.
6. Translation into English: 'The word accents are wrong or a little bit uncertain, the lack of the sentence expressivity because probably the speaker still struggles to think about

that had studied linguistics were generally more precise and their responses could be interpreted more easily (e.g. *Geminate non precise, accento sbagliato su "celebri"*[7]), even if this was not always the case.

As the aim was to identify emerging feature patterns from the data, a thorough analysis of the 3,486 responses was performed. In most cases, each comment described multiple features (e.g. *Pause e ritmo diversi, poco marcata la doppia "m" di "cammino"*[8]). Therefore, each feature was reported separately, categorising comments into those referring to segmental features and those referring to suprasegmental features, each with varying degrees. Fluency-related features were infrequent and thus, to make things simpler, they were included in the suprasegmental set. Comments that reported no clear or apparent foreign accent features in the speech sample were coded as *niente*[9]. Therefore, for each of the 21 speakers (7 for each L1/accent) the number total of comments and features was counted. Afterwards, for each subcategory of feature type, the number of occurrences was converted into percentages that added to 100% for each accent. Almost all the samples produced by native speakers of Italian were rated as having no cue of foreign-accentedness, with a few exceptions, where some listeners found that one stimulus displayed a rather reduced use of consonant gemination; in fact, 2.37% of the comments described this issue (e.g. *Penso che sia italiana, ma non pronuncia bene tutte le doppie*[10]). The percentages of all comment types are summarised, for each accent, in the Table 6.1 below.

The first thing that clearly emerged from the analysis of the listeners' judgments was the fact that the most frequent feature they reported was the improper use of consonant gemination (e.g. *Problemi con tutte le doppie*[11]; *Non usa proprio le doppie*[12]; *Non ci sono le geminate*[13]; etc.). Italian listeners frequently demonstrated their ability to promptly identify instances of incorrect consonant gemination in non-native speech. In fact, in the case of Russian-accented Italian, 22.52% of the comments focused on the improper use of consonant gemination. The French accent garnered 17.25% of such

---

what he is saying instead of worrying about how he is saying it; some letters, such as c and t have a slightly foreign pronunciation, maybe English'.

7. Translation into English: 'Imprecise use of consonant gemination, wrong accent on "celebri"'.

8. Translation into English: 'Different pauses and rhythm, not very marked the geminated consonant "m" in "cammino"'.

9. Translation into English: 'none'. I will report this label as 'none' in Table 6.1.

10. Translation into English: 'I think she is Italian, but she does not pronounce well all the geminates.'

11. Translation into English: 'Problems with all geminate consonants'.

12. Translation into English: 'He does not use geminate consonants at all'.

13. Translation into English: 'There are no geminate consonants'.

remarks, followed by the Spanish accent with 16.03%, the English accent with 12.15%, the German accent with 10.06%, and the Romanian accent with 5.59%. Besides that, another interesting finding is that a high number of segmental deviations (i.e. phoneme alterations, substitutions, or insertions) was reported for the samples of German accent (50.49%). Similar patterns were observed for the Spanish accent (45.97%). There followed the Russian (32.74%), the English (26.54%), the French (17.74%), and the Romanian (10.20%) accents.

Table 6.1: The distribution of judgements (as percentages) for each accent

ACCENT	SEGMENTAL (%)		SUPRASEGMENTAL (%)		NONE (%)
	Improper use of consonant gemination	Phoneme alterations, substitutions, insertions	Word stress misplacement	Intonation, rhythm, speech rate, pauses, fluency	Native-like pronunciation
RUSSIAN	22.52	32.74	15.50	26.92	2.32
ENGLISH	12.15	26.54	40.48	19.68	1.15
GERMAN	10.06	50.49	26.06	13.39	0.00
FRENCH	17.25	17.74	38.98	16.84	9.19
ROMANIAN	5.59	10.20	13.24	14.86	56.11
SPANISH	16.03	45.97	15.36	22.64	0.00
ITALIAN	2.37	0.00	0.00	0.00	97.63

In contrast to the segmental cues, which were consistently reported in clear terms by both naïve listeners and those with a linguistic background, the suprasegmental features were described using more ambiguous and vaguer terms by nearly all listeners, particularly when it came to intonation cues. (e.g. *Musicalità non prettamente italiana*[14]; *Pronuncia meno melodica di quella italiana*[15]; etc.). This may be attributed to the difficulty of explaining prosody, which was acknowledged by both experienced and inexperienced listeners (e.g. *Non essendo esperta in materia, non so come spie-*

14. Translation into English: 'Not a typically Italian musicality'.
15. Translation into English: 'The way he pronounces is less melodious than that of an Italian speaker'.

garlo, ha una strana inflessione a fine parola[16]; *Non è italiana ma non so spiegare come lo capisco, non ha la nostra espressività*[17]; etc.). Nonetheless, discernible patterns emerged from the listeners' comments, highlighting consistent trends. In relation to the English- and French-accented Italian, the most commonly cited suprasegmental feature was the misplacement of word stress. This accounted for 40.48% of instances for the English accent and 38.98% for the French accent. Issues at this level were also frequently reported for the German accent (26.06%), while they were less frequent for the Russian (15.50%), the Spanish (15.36%), and the Romanian accents (13.24%). Cues related to intonation, rhythm, speech rate, pauses, and fluency were particularly common for the Russian (26.92%) and Spanish (22.64%) accents, followed by the English (19.68%), French (16.84%), Romanian (14.86%), and German (13.39%) accents.

Finally, some listeners noted that certain speech samples exhibited no discernible features of a foreign accent. As expected, among the Italian speakers, 97.63% of cases were reported to lack any foreign accent cues. Interestingly, the Romanian-accented speech was judged to be devoid of such features 56.11% of the time. The French accent obtained this judgment in 9.19% of instances, followed by the Russian accent with 2.32% and the English accent with 1.15%. No instances of such comments were reported for the German and Spanish accents. From §6.1.1 through §6.1.6, a detailed breakdown of the cues that distinguish each of the six foreign accents will be provided.

6.1.1 Russian-accented Italian

This paragraph will delineate the prevalent segmental and suprasegmental features that Italian listeners commonly observed for the Russian accent. The reinterpretation of the judgments was undertaken to offer a linguistically meaningful categorization, while maintaining the scheme presented in Table 6.1. It also incorporates original comments provided by the listeners. While I identified additional features of the Russian accent not reported by native speakers, this discussion focuses only on the cues perceived as the most prominent by L1 speakers, to avoid issues arising from confirmation bias.

---

16. Translation into English: 'Not being an expert on this topic, I don't know how to explain this, she has a strange inflection in word final position'.
17. Translation into English: 'She is not Italian, but I don't know how to explain how I understand it, she doesn't have our expressivity'.

1. Segmental features:

a) Improper use of consonant gemination:
- Failure to produce consonant gemination, as observed in "*accompagnano*", "*accorrono*", "*ammirare*", "*arrivare*", "*attraverso*", "*bellezza*", "*cammino*", "*dalla*", "*della*", "*Geretti*", "*ovvero*", "*quello*", "*terreno*". As already mentioned in the previous section, this was the cue that listeners noticed the most (e.g. *Problemi nella pronuncia delle doppie*[18]). Even if in Russian the consonant gemination is not frequent, it may be encountered in loanwords, in highly specialised professional vocabulary, and in derivates, the latter are often the result of morpheme concatenation, hence the name concatenative gemination (Dmitrieva, 2017). However, the phonological status of Russian geminates is still an open topic. At the same time, additional empirical research is necessary to explore the acoustic differences between singletons and geminates in this language. What the results of this experiment strongly indicate is that Russian speakers were unable to produce consonant gemination in Italian, raising the possibility that they may not perceive this particular feature, after all.
- Unnecessary gemination in the following context: vi[tʧ]inissimo instead of vi[ʧ]inissimo in "*vicinissimo*", and sca[tt]urito instead of sca[t]urito in "*scaturito*". This was reported several times (e.g. *Errori di geminate: SCATTURITE invece di scaturite*[19]). As mentioned above, the random use of gemination among some of the Russian speakers suggests a potential lack of awareness regarding its appropriate usage.

b) Phoneme alterations, substitutions, insertions:
- Non-native production of the alveolar trills, which were rendered as apical dental flap, often palatalised, [r] as [ɾʲ]. This realisation was mentioned to have occurred in "*persone*", "*prestare*"; "*sempre*". The native speakers reported this cue often (e.g. *Le R sono troppo lievi*[20]). Palatalization is a common feature of Russian (Padgett, 2001), and as far as the trill-flap cue is concerned, according to Ladefoged and Maddieson (1996), in Russian, the flap is more common than the trill.
- Listeners reported also that the voiced alveolar lateral approximant [l] was produced as velarised [ɫ]; this was referred to as Russian dark

18. Translation into English: 'Pronunciation issues with the consonant gemination'.
19. Translation into English: 'Errors with geminates: SCATTURITE instead of scaturite'.
20. Translation into English: 'The Rs are too soft'.

lateral or hard lateral (Padgett, 2001). In fact, the voiced alveolar lateral appeared to be often realised as the pharyngealised laminal denti-alveolar [ɭ̪] and that it occurred in the following context: "*castello*", "*del*", "*il*", "*filosofia*" (e.g. *Ha la l scura*[21]).

- Non-native realisation of the palatal lateral approximant: [ʎ] as [l+i], in the context of "*meraviglia*" (e.g. *Meraviglia sembra meravilia*[22]).
- Similarly, non-native production of the palatal nasal: [ɲ] as [n+j] in "*accompagnano*" (e.g. *Pronuncia di gn in "accompagnano"*[23]). The two palatals are not attested in the phoneme inventory of Russian (Hamilton, 1980).
- Non-native realisation of the close back rounded vowel, either due to the centralization [u] as [ʉ], as for instance in "*scaturito*", or due to a transformation in a near-close near-back rounded vowel: [u] as [ʊ], such as the case of "*stupore*". The listeners mentioned these deviations frequently (e.g. *Non pronuncia benissimo la u in "stupore"*[24]). Both [ʉ] and [ʊ] occur frequently in Russian, the former is the production of /u/ between soft consonants, while the latter is its realisation when unstressed (Jones and Ward, 1969).
- Failure to produce the close-mid back rounded vowel in word-final position that auditorily seems to be reduced to a schwa; examples of the vowel reduction of [o] as [ɵ]/[ə] were found by native speakers in "*pensiero*" and "*Pietro*" (e.g. *Pronuncia dura della vocale o in "pensiero"*[25]). In fact, in Russian, after a soft consonant, /o/ is raised and centralised to [ɵ] (Jones and Ward, 1969).
- Non-native realisations of the open-mid front unrounded vowel, produced, instead, as a close-mid unrounded vowel, thus, [ɛ] as [e], in contexts such as "*tempo*" and "*castello*" (e.g. *Accento russo dato dalle "e" chiuse*[26]), or almost deleted in '*Pietro*'. This open-mid vowel is to be found in Russian, and it is used after 'soft consonants' (Hamilton, 1980).
- Non-native productions of the close-mid front unrounded vowel, due to vowel raising, so [e] as [i], in the case of "*bellezza*". This was reported often by the Italian listeners with approximative orthographic transcriptions (e.g. *Dice "bilezza" invece di "bellezza"*[27]).

---

21. Translation into English: 'She has the dark l'.
22. Translation into English: 'Meraviglia seems meravilia'.
23. Translation into English: 'The way she pronounces gn in "accompagnano"'.
24. Translation into English: 'She doesn't pronounce well the u in "stupore"'.
25. Translation into English: 'Hard pronunciation of o in "pensiero"'.
26. Translation into English: 'Russian accent due to the closed "e"'.
27. Translation into English: 'She is saying "bilezza" instead of "bellezza"'.

- Listeners reported the deletion of the final close front unrounded vowel [i] in "*tormentatissimo*" (e.g. *Tormentatismo invece di tormentatissimo*[28]), but more realistically it might have been uttered as a near-close central unrounded vowel [i] as [ɨ], attested in Russian (Hamilton, 1980).
- Non-native realisation of the rising diphthong /wɔ/ in "*cuore*", which is instead uttered as a hiatus, and at the same time the open-mid back rounded vowel is produced as a close-mid back unrounded vowel, [wɔ] as [uo] (e.g. *Profondità di pronuncia del "U" in "cuore" e chiusura della "O"*[29]).

2. Suprasegmental features:

a) Word stress misplacement in various instances: "*parroco*" ([ˈparroko] as *[parˈroːko]), "*allestite*" ([allesˈtiːte] as *[ˈallestite]) were reported by most listeners (e.g. *Accento sulle parole sbagliato*[30]). One possible reason why Russian speakers stressed the penultimate syllable so often, as for instance in the word "*parroco*", is because in Italian, the paroxytone is the most frequently distributed type of lexical stress (Nespor, 1993; D'Imperio and Rosenthall, 1999; Marotta, 2016), and therefore they generalised this rule for words they might have not known or used frequently.

b) Non-native intonation, pausing, fluency issues and non-native rhythm patterns were mentioned quite often for the Russian-accented voices. However, most of the times, these comments were too generic to be properly interpreted (e.g. *Parole stentate, nessuna fluidità*[31]; *Sbaglia il tono della frase*[32]; *Sbaglia pause*[33]). Unlike the lexical stress or the segmental cues, signalling intonation is a rather challenging task, as it somehow lacks tangible reference points. Furthermore, explaining prosody becomes arduous when the listener lacks familiarity with the topic. Nonetheless, the comments received indicated that listeners were indeed aware of something happening at the prosodic level, but they were not able to name it properly. Most comments described especially rhythmic aspects, which were perceived as distant from the

---

28. Translation into English: 'Tormentatismo instead of tormentatissimo'.
29. Translation into English: 'Deepness of pronunciation of "U" in "cuore", and the closure of "O"'.
30. Translation into English: 'Wrong lexical stress'.
31. Translation into English: 'Faltering words, no fluidity'.
32. Translation into English: 'The tone used in this sentence is incorrect'.
33. Translation into English: 'The pauses are incorrect'.

Italian rhythmic patterns (e.g. *Il tentennamento di alcune parole*[34]; *Parlata meccanica*[35]; *Andamento macchinoso*[36]). If one accepts the conventional differentiation between syllable-timed and stressed-timed languages, the discrepancy in rhythm patterns can potentially be elucidated using that framework (Hoequist, 1983). However, this trait is deemed to be more intricate, prompting the exploration of this feature instrumentally in §6.3 and §6.4. The goal is to extract rhythmic features for all the accents examined in this study, providing a more comprehensive explanation based on an empirical exploration of the prosody.

6.1.2 English-accented Italian

To interpret the feature patterns of the English-accented voices, an analysis was conducted on judgments provided by native Italian listeners. It is worth noting that the cues listed by the listeners are not exhaustive, as additional cues could be identified. However, only the cues listed by the native Italian listeners will be reported, as the focus lies on the features perceived as the most salient.

1. Segmental features:

   a) Improper use of consonant gemination:
   - Listeners reported the lack of consonant gemination in "*accompagnano*", "*alla*", "*arrivare*", "*attraverso*", "*bellezza*", "*colleghi*", "*condurre*", "*dalla*", "*della*", "*fatto*", "*Geretti*", "*hanno*", "*Illegio*", "*minaccia*", "*ovvero*", "*parroco*", "*terreno*". Some of their comments were particularly detailed in this respect (e.g. *Alcune doppie non sono marcate quindi si percepiscono poco, colleghi pare coleghi*[37]). In English consonant length is not a distinctive trait within root words (Kaye, 2005).
   b) Phoneme alterations, substitutions, insertions:
   - Non-native realisation of the alveolar trill, rendered, instead, as post-alveolar approximants, sometimes labialised: [r] as [ɹ]/[ɹʷ] in "*arte*", "*Carnia*", "*cerca*", "*condurre*", "*cuore*", "*forme*", "*Geretti*", "*inferno*", "*meraviglia*", "*ovvero*", "*pensiero*", "*per*", "*perciò*", "*realtà*",

---

34. Translation into English: 'Hesitation in some words'.
35. Translation into English: 'Mechanical speech'.
36. Translation into English: 'She speaks as if she were a machine'.
37. Translation into English: 'Some geminates are not marked, so they are hardly perceivable, "colleghi" seems "coleghi"'.

"*sempre*", "*scaturite*", "*scaturito*". This feature was listed often in the listeners' comments (e.g. *Suono della lettera "r", sicuramente meno marcato e più "morbido" rispetto a come lo pronuncerei io di madrelingua italiana*[38]). Generally, in English, the postalveolar approximant is the most common realisation of the /r/ phoneme, in most dialects, as well as in RP and GA (Ladefoged, 2001).

- Non-native production of the palatal lateral approximant: [ʎ] as [l+i] were also reported in "*meraviglia*" (e.g. *Pronuncia sbagliata di meraviglia*[39]).

- Similarly, the non-native realisation of the palatal nasal: [ɲ] as [n+j] in "*accompagnano*" (e.g. *Gn nella parola accompagnano è pronunciato male*[40]). The two palatals are not attested in the phonemic inventory of RP and GA (McCully, 2009).

- The realisation of a voiceless palatal alveolar sibilant fricative [ʃ] instead of a voiceless palatal alveolar sibilant affricate [ʧ] in "*decina*" (e.g. *Fricativa al posto dell'affricata in "decina"*[41]).

- Just like for their L1 (McMahon, 2002), English speakers used aspiration in the case of voiceless stops. This was reported for [k] in "*coniugare*", [p] in "*paese*", "*per*" and "*porta*", and [t] in "*comitato*"; aspirated [t] is also produced as an alveolar stop rather than a dental stop (e.g. *La parola "comitato" è pronunciata come se la t fosse aspirata*[42]).

- English speakers were reported to utter the word "*forme*" with a long open-mid back rounded vowel [ɔ] instead of a close-mid back rounded vowel [o] (*[ˈfɔːme] or *[ˈfɔːˈme]). This cue of vowel lengthening, and openness was reported frequently by the listeners (e.g. *Pronuncia inglese con la o aperta in "forme"*[43]). In most varieties of English, the vowel length is allophonic: a vowel is generally shorter in case it occurs before a voiceless consonant in the same syllable; while, however, in Australian English, the length feature is contrastive (Collins and Mees, 2013).

- Another example of a substitution that was frequently perceived among English speakers of Italian concerns once again the final

---

38. Translation into English: 'The r sound less definitely less marked and softer than how I would pronounce it as a native speaker'.
39. Translation into English: 'Incorrect pronunciation of the word meraviglia'.
40. Translation into English: 'Gn in the word accompagnano is not pronounced correctly'".
41. Translation into English: 'A fricative instead of an affricate in "decina"'.
42. Translation into English: 'The word "comitato" is pronounced as if the t were aspirated'.
43. Translation into English: 'English pronunciation of "forme" with an open o'.

close-mid back rounded vowel [o] in "*ovvero*" and in "*pensiero*", which was diphthongised as [əʊ]. (e.g. *La o finale di pensiero*[44]).

- Furthermore, the first close-mid back rounded vowel [o] in "*minuscolo*" was uttered as the close back rounded vowel [u] (e.g. *Pronuncia di o in "minuscolo"*[45]).
- Listeners mentioned the lengthening of the close back rounded vowel, [u] as [uː], with a possible centralization [ʉː] in "condurre" or fronting [yː], in words such as "*coniugare*" (e.g. *La u di "coniugare" è lunga e sembra staccata, pronunciata all'inglese*[46]).

2. Suprasegmental features:

a) Word stress misplacement in various instances, such as "*celebri*" ([ˈtʃɛlebri] as *[tʃɛˈlebri]), "*decina*" ([deˈtʃina] as *[ˈdetʃina]), "*ovvero*" ([ovˈvero] as *[ˈoverəʊ]), "*parroco*" ([ˈparroko] as *[parˈroko]), "*perciò*" ([perˈtʃɔ] as *[ˈpertʃɔ]), were reported by most listeners (e.g. *Sbaglia l'accento di parola in "parroco"*[47]; *L'accento sbagliato sulla parola "celebri" indica che non è madrelingua*[48]).

b) Various comments referred to non-native intonation, pausing, fluency issues and non-native rhythm patterns (e.g. *Un effetto "singhiozzo" nella percezione della frase intera*[49]; *Pause errate*[50]; *La lentezza della frase*[51]; *Intonazione complessiva della frase*[52]; *Intonazione completamente diversa da quella italiana*[53]). In spite of the fact that these comments tend to be generic, they imply that listeners are aware of deviations at the prosodic level. A lot of comments reported the unusual rhythmic patterns of the English speakers, which were perceived as distant from the Italian model (e.g. *Il ritmo e le pause della frase*

---

44. Translation into English: 'Final o in "pensiero"'.
45. Translation into English: 'The way he pronounces o in "minuscolo"'.
46. Translation into English: 'The u in "coniugare" is long and it seems to be detached, pronounced the English way'.
47. Translation into English: 'The lexical stress in "parroco" is incorrect'.
48. As is' translation into English: 'The incorrect lexical stress on the word "celebri" indicates that it is not a native speaker of Italian'.
49. Translation into English: 'The perception of an overall hiccup when listening to the entire sentence'.
50. Translation into English: 'Wrong pauses'.
51. Translation into English: 'The slowness when uttering the sentence'.
52. Translation into English: 'The entire intonation of the sentence'.
53. Translation into English: 'The intonation is completely different from the Italian intonation'.

sono come in inglese[54]; *Ritmo diverso dal nostro*[55]; *Il ritmo della frase non è naturale*[56]). Again, this might be explained in terms of isochrony, since traditionally, English is classified as a stress-timed language, while Italian is syllable-timed (Hoequist, 1983).

6.1.3 German-accented Italian

Just like with the previous accents, this section provides a summary of the comments made by the listeners regarding the salient features of the German-accented Italian.

1. Segmental features:

a) Improper use of consonant gemination:
- The failure to produce the consonant gemination was less frequent than in the case of other accents and the listeners mentioned it in the following contexts: "*accompagnano*", "*arrivare*", "*arrivato*", "*attraverso*", "*bellezza*", "*cammino*", "*castello*", "*fatto*", "*fiamme*", "*Geretti*", "*minaccia*", "*parroco*", "*rappresenta*", "*siccome*", "*terreno*" (e.g. *Difficoltà con le geminate in "fiamme"*[57]). As previously seen with English, modern standard German does not have a productive length contrast in consonants, but it allows sequences of identical consonants in morphological contexts, such as compounds. However, according to Kotzor et al. (2016: 356), English and German speakers articulate this type of geminates "with acoustic characteristics similar to those found in languages with an underlying length contrast, despite no longer displaying the contrast morpheme-internally."

b) Phoneme alterations, substitutions, insertions:
- All German speakers selected for this experiment were reported to have encountered difficulty in producing the alveolar trill. In fact, they produced it as an uvular trill [ʀ] or as the voiced uvular fricative [ʁ], both more common in their L1 (Grantham O'Brien and Fagan, 2016), in the following words: "*affabulatore*", "*arrivare*", "*attraverso*", "*bravissimo*", "*Carnia*", "*celebri*", "*cerca*", "*condurre*", "*cuo-*

---

54. Translation into English: 'The rhythm and the pauses in this sentence are just like in English'.
55. Translation into English: 'The rhythm is different from ours'.
56. Translation into English: 'The rhythm of this utterance is not natural'.
57. Translation into English: 'Difficulties uttering the geminates in "fiamme"'.

re", "*deriva*", "*dire*", "*forme*", "*Geretti*", "*grande*", "*inferno*", "*internazionale*", "*meraviglia*", "*ovvero*", "*pensiero*", "*persone*", "*perciò*", "*Pietro*", "*prete*", "*predicatore*", "*rappresenta*", "*realtà*", "*romano*", "*sempre*", "*scaturite*", "*scaturito*", "*stupore*", "*terminò*", "*terreno*", "*tormentatissimo*", "*verso*", "*vetrina*". It might also have been uttered as the near-open central vowel [ɐ]. In general, the improper production of the alveolar trill was frequently listed in the Italian listeners' comments (e.g. *Pronuncia delle "r", sia singole che doppie*[58]; *Marcata pronuncia della erre arrotolata*[59]; *R tedesca*[60]).

- Sometimes the voiceless alveolar fricative was produced as a voiced alveolar fricative, as for instance in "*insieme*", "*consiste*", and "*persone*", the latter being uttered as the German counterpart [pɛrˈzoːn]. Native speakers mentioned this cue frequently (e.g. *"Persone" con la "s" sonora*[61]).

- Non-native realisation of the palatal lateral approximant: [ʎ] as [l+i], in "*dagli*" (e.g. *"Dagli" sembrerebbe quasi "dali"*[62]). The phoneme is not attested in the phonemic inventory of German (Ladefoged and Maddieson 1996).

- The use of aspiration in the voiceless stops, attested in German (Mangold, 2005), was also mentioned for [p] in "*passo*" (e.g. La *[p] iniziale sembra aspirata*[63]) was also reported.

- In German, [ə] occurs in closed final syllables (Kohler, 1999), and this feature was transferred by the German speakers in their L2 Italian. In fact, listeners reported the alteration of the close-mid front unrounded vowel [e] that systematically would become the mid-central vowel [ə] in word-final position, for instance in "*prete*" and "*predicatore*". Native speakers of Italian perceived this as deviant (e.g. *La e finale di "predicatore" non sembra proprio una e*[64]).

- Insertion of the close front unrounded vowel after the geminated voiceless palate-alveolar affricate in "minaccia", shifting also the word stress (e.g. [miˈnattʃa] as *[minatˈtʃiːa]). Native speakers of Italian promptly noticed this feature (e.g. *Pronuncia della "i" grammaticale in "minaccia"*[65])

---

58. Translation into English: 'The way he pronounces "r"s, both singletons and geminates'.
59. Translation into English: 'Marked pronunciation of the rolling r'.
60. Translation into English: 'German R'.
61. Translation into English: '"Persone" with voiced "s"'.
62. Translation into English: '"Dagli" seems more like "dali"'.
63. Translation into English: 'The word-initial [p] is aspirated'.
64. Translation into English: 'The final e in "predicatore" doesn't even sound like an e'.
65. Translation into English: 'Pronunciation of grammatical "i" in "minaccia"'.

- Native speakers reported the deletion of the close front unrounded vowel [i] in word-final position in "*suoi*" (e.g. *Difficoltà a scandire vocali in "suoi", la "i" è talmente chiusa che quasi non si sente*[66]).
- The vast majority of listeners reported a pronunciation deviation in the word "*quello*" ([ˈkwello] as [ˈkvɛllo]), which could be attributed to differences in orthographic norms between German and Italian (e.g. "*Quello" pronunciato "qvello*"[67]).
- Italian listeners reported a similar issue with the word "*scelta*", where German speakers, potentially due to the absence of similar grapheme combinations in their L1, pronounced it as *[ˈskɛlta] instead of [ˈʃelta] (e.g. *Non sa leggere "scelta"*[68]).

2. Suprasegmental features:

a) Word stress misplacement in "*celebri*" ([ˈtʃɛːlebri] as *[tʃeˈleːbri]), "*internazionale*" ([internattsjoˈnaːle] as *[ˈinternattsjonale]), "*minaccia*" ([miˈnattʃa] as *[minatˈtʃiːa], with vowel insertion), "*parroco*" ([ˈparroko] as *[parˈroːko]), "*terminò*" ([termiˈnɔ]] as *[ˈtɛrmino]) were reported by most listeners (e.g. *L'accento in "parroco" è sbagliato*[69]; *Accento sbagliato di alcune parole: minaccia, celebri*[70]). As seen with other non-native Italian accents discussed above, it seems that some German speakers stressed the penultimate syllable, as for instance in "*celebri*" and "parroco", overgeneralising the fact that the paroxytone is the most common type of lexical stress in Italian.

b) Unlike most of the other accents, there were relatively fewer comments on the prosodic features of German-accented Italian, and generally they were quite broad, reported non-native intonation in general, pausing, and non-native rhythm patterns (e.g. *La sua intonazione è molto meno "melodica" della nostra*[71]; *Intonazione inusuale*[72]; *Pause troppo cadenzate*[73]). Most comments referred to the rhythm, which was perceived as different from the Italian rhyth-

---

66. Translation into English: 'Difficulty of uttering the vowels in "suoi", the "i" is so closed that one cannot hear it at all'.
67. Translation into English: '"Quello" is pronounced "qvello"'.
68. Translation into English: 'He can't read "scelta"'.
69. Translation into English: 'The lexical stress in "parroco" is incorrect'.
70. Translation into English: 'Incorrect lexical stress of some words: minaccia, celebri'.
71. Translation into English: 'His intonation is far less melodic than ours'.
72. Translation into English: 'Unusual intonation'.
73. Translation into English: 'The pauses are too evenly marked'.

mic pattern (e.g. *Ritmo lento*[74]; *Il ritmo dato alla frase*[75]; *Ritmo e pause diverse*[76]; *Il ritmo e le pause della frase*[77]).

6.1.4 French-accented Italian

The following section will summarise the Italian listeners' comments regarding the features they perceived as being more salient for the French accent.

1. Segmental features:

   a) Improper use of consonant gemination:
   - The lack of consonant gemination was mentioned for all the samples of French accented-Italian: "*accorrono*", "*allestite*", "*ammirare*", "*arrivare*", "*arrivato*", "*attraverso*", "*bellezza*", "*cammino*", "*castello*", "*fatto*", "*fiamme*", "*Geretti*", "*minaccia*", "*parroco*", "*rappresenta*", "*terreno*", "*villaggio*" (e.g. *Non dice alcune doppie, come quelle in "bellezza"*[78]). This finding is not surprising; graphemically, there are geminated consonants in French, but in words such as "la guerre" auditorily and acoustically there is no gemination. In fact, the geminate consonants are rare in French, and the consonant length is generally not distinctive (Tranel, 1987; Walker, 2001). Traditionally, in case of the uvular fricative, the pronunciation of conditional forms would contrast with the imperfect forms ("courrait" vs. "courait", [kuʁʁɛ] vs. [kuʁɛ]). However, the contrast is not marked in case of other consonants ("mettre" vs. "maître", [mɛtʁ] in both cases).
   b) Phoneme alterations, substitutions, insertions:
   - Most French speakers selected for the experiment were reported to have had difficulties pronouncing the alveolar trill. In fact, they rendered it either as the voiced uvular fricative [ʁ], or as the uvular trill [ʀ], the most common realisations of the French rhotic (Fougeron and Smith, 1993). This issue was reported for "*arrivare*", "*attraverso*", "*bravissimo*", "*celebri*", "*cerca*", "*condurre*", "*cuore*", "*deriva*", "*dire*", "*forme*", "*Geretti*", "*grande*", "*inferno*", "*internazionale*", "*meraviglia*", "*ovvero*", "*pensiero*", "*persone*", "*perciò*", "*Pietro*",

---

74. Translation into English: 'Slow rhythm'.
75. Translation into English: 'The rhythm given to the sentence'.
76. Translation into English: 'Different rhythm and pauses'.
77. Translation into English: 'The rhythm and the pauses'.
78. Translation into English: 'He doesn't utter some geminates, like for instance those in "bellezza"'.

"*presentare*", "*prete*", "*predicatore*", "*rappresenta*", "*realtà*", "*romano*", "*sempre*", "*scaturite*", "*scaturito*", "*stupore*", "*terminò*", "*terreno*", "*tormentatissimo*", "*verso*", "*vetrina*". Almost all listeners mentioned this cue (e.g. *La R moscia in tutte le parole*[79]; *La pronuncia della r, sebbene somiglia un po' alla "r veneziana", non è comune tra i madrelingua italiani*[80]; *Le R sono troppo francesi*[81]; *Attraverso e stupore hanno una R quasi retroflessa*[82]).

- In one instance – "Colosseo" – the voiceless alveolar fricative [s] was produced as the voiced alveolar fricative [z]. Native speakers mentioned this cue frequently for this speaker (e.g. *Pronuncia z invece di s in Colosseo*[83]), but more realistically this was a mere reading error.

- The voiceless alveolar affricate [ts] was reported to have been uttered as the voiceless alveolar fricative [s] on several occasions in "internazionale" (e.g. *Pronuncia in maniera sbagliata la z di "internazionale"*[84]).

- Substitutions of the voiced palatal alveolar affricate [dʒ] with the voiceless palate-alveolar affricate [tʃ] in "Angelo" were reported several times by the Italian listeners (e.g. *Sento che ge è pronunciata come ce in Angelo*[85]).

- Listeners mentioned that French speakers inserted the close front unrounded vowel [i] after the geminated voiceless palate-alveolar affricate [tʃ] in "minaccia", also shifting the word stress (e.g. [miˈnattʃa] as *[minatˈtʃiːa]). There were several comments regarding this issue (e.g. *Si sente la "i" in "minaccia"*[86]).

- French speakers were reported to lengthen their open-mid front unrounded vowel [ɛ] in "castello" and "verso" (*Allunga di più la e in "castello"*[87]). In fact, this may be attributed to the fact that in French it determines some instances of minimal pairs (e.g. mètre vs. maître, [mɛtʁ] vs. [mɛːtʁ]), but otherwise, variation in vowel length is purely allophonic (Tranel, 1987; Walker, 2001).

79. Translation into English: 'The way she pronounces "r"s, both singletons and geminates'.
80. Translation into English: 'The way he pronounces the r, even if it sounds a bit like the Venetian "e", it is not common among native speakers of Italian'.
81. Translation into English: 'The Rs are too French'.
82. Translation into English: '"Attraverso" and "stupore" have almost a retroflexed R'.
83. Translation into English: 'He pronounces z instead of s in Colosseo'.
84. Translation into English: 'She pronounces the z in "internazionale" the wrong way'.
85. Translation into English: 'I hear that ge is uttered like ce in Angelo'.
86. Translation into English: 'One can hear an "i" in "minaccia"'.
87. Translation into English: 'The e in "castello" is longer than usual'.

- Additionally, Italian listeners perceived that the vowels in word final position were longer than what they would have expected, and at the same time the stress would be placed on the last syllable (e.g. *Vocali allungate a fine parola e la tendenza ad accentare l'ultima vocale delle parole, come per esempio in "attraverso"*[88]).

2. Suprasegmental features:

   a) As anticipated above, since word stress is not distinctive in French (Walker, 1975), it was expected that French speakers would encounter challenges regarding proper stress placement in Italian. Indeed, misplacement of stress was frequently observed in French-accented Italian in "*attraverso*" ([attraˈvɛrso] as *[attravɛrˈso]), "*basilica*" ([baˈziːlika] as *[baziliˈka]), "*celebri*" ([ˈtʃɛːlebri] as *[tʃɛˈleːbri]), "*esitano*" ([ˈeːzitano] as *[eziˈtaːno]), "*Angelo*" ([ˈandʒelo] as *[anˈdʒeːlo]), "*minaccia*" ([miˈnattʃa] as *[minatˈtʃiːa]), "*parroco*" ([ˈparroko] as *[parˈroːko]), "*terminò*" ([termiˈnɔ]] as *[ˈtɛrmino]), "*vicinissima*" ([vitʃiˈniːsimo] as *[vitʃinisiˈmo]) were reported by most listeners (e.g. *Accento non posto sulle sillabe giuste*[89]; *Accento sbagliato su "basilica" e "terminò"*[90]). As already mentioned in the case of the previous accents, some speakers overgeneralised the fact that often in Italian the stress falls on the penultimate syllable.
   b) Various comments referred to the prosodic features of French-accented Italian. Besides the stress, most of the comments regarded pauses, rhythm, and fluency in general (e.g. *Ritmo e pause diverse*[91]; *Ritmo spezzato*[92]; *Ritmo sbagliato*[93]; *Nessuna intonazione - frase piatta*[94]).

6.1.5 Romanian-accented Italian

The summary of the findings gathered from the comments provided by the Italian listeners is presented below.

---

88. Translation into English: 'Long vowels at the end of the word and the tendency to place the stress on the last vocal of the word, like for instance in "attraverso"'.
89. Translation into English: 'The stress is not on the right syllables'.
90. Translation into English: 'Incorrect stress on "basilica" and "terminò"'.
91. Translation into English: 'Different rhythm and pauses'.
92. Translation into English: 'Broken rhythm'.
93. Translation into English: 'Wrong rhythm'.
94. Translation into English: 'No intonation – the utterance is plain'.

1. Segmental features:

a) Improper use of consonant gemination:

- Similarly to the other accents, there were various instances of failure to produce consonant gemination, like for instance in: "*allestite*", "*ammirare*", "*arrivare*", "*attraverso*", "*bellezza*", "*cammino*", "*castello*", "*colleghi*", "*condurre*", "*della*", "*fatto*", "*fiamme*", "*Geretti*", "*hanno*", "*minaccia*", "*parroco*", "*rappresenta*", "*villaggio*" (e.g. *Doppie pronunciate come singole*[95]). This difficulty may be attributed to the fact that Romanian lacks consonantal length as a distinctive feature (Chiṭoran, 2002a).

b) Phoneme alterations, substitutions, insertions:

- Sometimes, the listeners mentioned that the alveolar trill was not produced how native speakers of Italian would produce it, for instance in case the case of "*attraverso*", "*cuore*", "*Geretti*", "*arrivare*", and "*stupore*" (e.g. *La pronuncia non vibrata della "r" in "Geretti" and "arrivare"*[96]). The trill is present in the phoneme inventory of Romanian, but it is generally thought to be rendered as a dental trill [r̪] rather than a typical alveolar trill, especially in word-initial position (Vasiliu, 1968); however, some other scholars, such as Chiṭoran (2002a) consider that it is realised as a flap [ɾ] and only occasionally, in word-initial position, it is realised as the dental trill [r̪]. Additionally, the lack of gemination might have also contributed to the perception of an insufficiently vibrated trill in words such as "*arrivare*".

- Non-native production of the palatal lateral approximant: [ʎ] as [l+i], in "*dagli*" and "*meraviglia*" (e.g. *"Gli" pronunciato come "li"*[97]).

- Similarly, Italian speakers reported the improper realisation of the palatal nasal [ɲ] as [n+j] in "*accompagnano*" (e.g. *Non riesce a pronunciare gn*[98]). None of the two palatals is attested in the phonemic inventory of Romanian (Chiṭoran, 2002a).

- Non-native realisations of the open-mid front unrounded vowel, which was instead produced as a close-mid unrounded vowel, thus [ɛ] as [e], in "*attraverso*", "*colleghi*", "*Illegio*", "*inferno*", "*rappresenta*", "*sempre*", "*tempo*" and "*castello*" (e.g. *Tutto perfetto tranne in "castello": la "e" è troppo chiusa*[99]).

---

95. Translation into English: 'The geminates are pronounced as singletons'.
96. Translation into English: 'Non-vibrated pronunciation of the "r" in "Geretti" and "arrivare"'.
97. Translation into English: '"Gli" seems more like "li"'.
98. Translation into English: 'She cannot pronounce gn'.
99. Translation into English: 'Everything is perfect, except for "castello": the "e" is too closed'.

- Italian listeners mentioned that Romanian speakers failed to produce the open-mid back rounded vowel, which was instead realised as the close-mid rounded vowel, thus [ɔ] as [o], in "*canonica*" and "*dopo*" (e.g. *Diverso suono della "o", molto chiuso*[100]*; Sembra italiano ma chiude la "o"*[101]). The failure to produce open vowels among the Romanian speakers of Italian is supported also by the results of acoustical-instrumental analysis. In fact, the research conducted by Combei et al. (2020) reveals an interesting finding: Romanian learners of L2 Italian produce open-mid vowels [ɔ] and [ɛ] with spectro-acoustic characteristics that closely resemble their closed-mid counterparts [o] and [e].
- Non-native realisation of the rising diphthong in "*cuore*" and "*vuole*", which is instead uttered with a hardly perceivable [w] (e.g. *È molto simile alla pronuncia italiana, a parte l'u di "vuole" che non si sente*[102]). This diphthong, indeed, is not attested in Romanian. The closest sequence – /wo/ – may be found only in rare borrowings, such as "walkman", and it is generally produced as an ascending diphthong (Chițoran, 2002a, 2002b, 2002c).

2. Suprasegmental features:

a) Word stress misplacement was reported on a few occasions: "*celebri*" ([ˈtʃɛlebri] as *[tʃɛˈleːbri]), *dagli* ([ˈdaʎʎi] as *[daʎˈʎi]), "*parroco*" ([ˈparroko] as *[parˈroːko]) but compared to other accents this cue was less frequent (e.g. *Accento sbagliato su "celebri"*[103]).
b) Some judgments referred to the prosodic features of Romanian-accented Italian (e.g. *Intonazione innaturale*[104]; *Tono generale e cadenza della frase*[105]; *Parla bene e velocemente ma senza intonazione*[106]). However, most of these comments regarded pauses and rhythm, (e.g. *Lievissime pause che interpreterei come difficoltà di leggere un testo per la prima volta, ma in realtà potrebbero anche essere dovute al diverso ritmo*[107]; *Troppo veloce, senza una pausa tra*

---

100. Translation into English: 'The "o" sound is different, very closed'.
101. Translation into English: 'He seems Italian, but he closes the "o"'.
102. Translation into English: 'It's very similar to the Italian pronunciation, except for the u in "vuole", that is not perceivable'.
103. Translation into English: 'Incorrect stress on "celebri"'.
104. Translation into English: 'Non-natural intonation'.
105. Translation into English: 'The general tone and the phrase cadence'.
106. Translation into English: 'He speaks well and fast but without intonation'.
107. Translation into English: 'Light pauses that I could interpret as difficulties of reading the text for the first time, but they might as well be due to a different rhythm'.

una parola e l'altra[108]; *Ritmo diverso da quello italiano (molto scandito)*[109]; *Non rispetta le pause della frase*[110]; *Cadenza ritmica anomala*[111]). Unlike the other five accents, quantitatively, the Romanian-accented Italian received less comments regarding the salient foreignness cues. In fact, 56.11% of the judgements reported no feature at all, or the cues were almost unperceivable (e.g. *Sembrerebbe italiana*[112]; *Difficile ma niente*[113]; *Nessun elemento*[114]). Since, comparatively there were not many segmental issues, listeners focused more on the prosodic level, hence, there was a higher number of comments regarding the rhythmic patterns and the intonation of the Romanian accent.

6.1.6 Spanish-accented Italian

The summary of the most salient cues of Spanish-accented Italian, as reported by the native Italian speakers, is presented below.

1. Segmental features:

   a) Improper use of consonant gemination:
   - Many listeners mentioned that Spanish speakers failed to produce consonant gemination, for instance in: "*accompagnano*", "*allestite*", "*ammirare*", "*anni*", "*arrivare*", "*attraverso*", "*bellezza*", "*cammino*", "*castello*", "*colleghi*", "*condurre*", "*della*", "*fatto*", "*fiamme*", "*Geretti*", "*hanno*", "*minaccia*", "*ovvero*", "*parroco*", "*rappresenta*", "*villaggio*" (e.g. "*Geretti" con una T e "anni" con una N*[115]; *Alcune doppie non pronunciate*[116]). The Spanish varieties do not have the trait of consonant gemination (Harris, 1969).
   - However, some Spanish speakers used consonant gemination when it was not necessary, as in "*cammino*", where the final alveolar nasal was geminated, or the case of "*comitato*", this time the bilabial nasal was produced as a geminate (e.g. *Mette le doppie quando non servo-*

---

108. Translation into English: 'Too fast, without any pause between the words'.
109. Translation into English: 'The rhythm is different from the Italian one (very articulated)'.
110. Translation into English: 'She does not respect the pauses'.
111. Translation into English: 'Anomalous rhythmic cadence'.
112. Translation into English: 'She could be Italian'.
113. Translation into English: 'Hard to tell, but nothing'.
114. Translation into English: 'No element'.
115. Translation into English: '"Geretti" with a T and "anni" with a N'.
116. Translation into English: 'Some geminates are not pronounced'.

no: "comMitato"[117]); another such example is "scaturito" (e.g. *Aggiunta di doppie dove dovrebbero essere assenti: ScaTTurito*[118]). This rather random use of gemination might suggest that Spanish speakers were not fully aware of how and when to use this feature.

b) Phoneme alterations, substitutions, insertions:

- Sometimes, the listeners mentioned that the alveolar trill was produced differently, for instance in "*terreno*" and "*realtà*" (e.g. *Le vibranti pronunciate diversamente*[119]; *R ispanica in "terreno" e "realta"*[120]; *Raddoppiamento della "r" ad inizio parola*[121]). In most varieties of Spanish, the trill and the tap are in phonemic contrast in intervocalic contexts, but they are otherwise in complementary distribution; in word-initial position or after the hetero-syllabic consonant the standard Spanish realisation is [r] (Rivera Campos and Boyce, 2003). Interestingly, when describing the differences between the Italian and the Spanish trill, Canepari (2007) points out that the Spanish trill has three "beats" (i.e. it vibrates more), while the Italian sound only two. Therefore, even if there is a contrastive consonant elongation both in Spanish and in Italian, the way Spanish speakers produced the trill when they spoke Italian represented a recognisable trait of Spanish-accented Italian.

- Italian listeners mentioned in various occasions that the voiceless labiodental fricative was not always produced in a native-like manner, for instance in "*filosofia*" and "*fede*" (e.g. *La pronuncial della "f" in "filosofia" è molto marcata*[122]). One possible explanation might be that the sound was realised as a voiceless bilabial fricative [ɸ], a non-standard variant, occurring in various Spanish dialects (Lloyd, 1987; Ladefoged, 2001).

- Probably the most recognisable trait of Spanish-accented Italian, at least as it seemed to emerge from the data of this experiment, was the non-native realisation of the voiced labiodental fricative. Indeed, /v/ does not occur in the phoneme inventory of Spanish; the graphemes ⟨v⟩ and ⟨b⟩ both refer to one phoneme, namely /b/; moreover, [β] and [b] are allophones: in most contexts – excluding the beginning of phrases (after pauses) and after [n] and [m] where /b/ is realised as a [b] – it is realised as [β] (Harris, 1969). Examples of peculiar pro-

---

117. Translation into English: 'He uses geminates when they are unnecessary: "comMitato"'.
118. As is' translation into English: 'Adding geminates where they should be absent: *ScaTTurito*'.
119. Translation into English: 'The trills are pronounced differently'.
120. Translation into English: 'Spanish R in "terreno" and "realtà"'.
121. Translation into English: 'Geminated R at the beginning of the word'.
122. Translation into English: 'The way he pronounces "f" in "filosofia" is very marked'.

153

nunciations of the voiced labiodental fricative and the voiced bilabial stop were reported often by native speakers of Italian, especially in the following instances: "*ovvero*", "*bellezza*", "*meraviglia*", "*vicinissima*", "*vede*" (e.g. *Lettura della "V" come una "b"*[123]; *Confusione tra b e v con la labiale spagnola di cui non ricordo il nome (segno ipa è il beta)*[124]; *Pronuncia della parola "ovvero" con la "v" pronunciata come la "b" tipica degli spagnoli*[125]).

- Non-native realisations of the voiceless and voiced alveolar fricatives were mentioned in several instances for the Spanish-accented Italian: "*scaturito*", "*scaturite*" "*basilica*", "*filosofia*", "*suo*", "*tormentatissimo*". Several phenomena occurred: the voiced alveolar fricative was instead rendered as [s] in "*basilica*" (e.g. "*Basilica*" *ha una "s" strana, tipo sss*[126]); sometimes Italian listeners used the term 'lisp' to describe how Spanish speakers realised the alveolar fricatives (e.g. *Parla con la lisca*[127]), the presence of the voiceless dental fricative /θ/ is, in fact, a trait common to most phonemic inventories of the Spanish varieties in Spain (Hualde, 2005); additionally, some Spanish speakers tended to use a more dentalised alveolar sibilant before dental consonants [s] as [s̪] (e.g. "*sostiene*", "*consiste*", "*vista*"); finally there were also examples of epenthesis, with the prosthetic vowel [e] at the beginning of the word (e.g. *Dice "pensiero escaturito" anziché "scaturito"*[128]).

- One Spanish speaker was reported to utter the voiceless palate-alveolar affricate as a palatalised voiceless alveolar affricate in "*vicinissima*" (e.g. *"Vicinissima" pronunciata con una z*[129]). This type of realisation is common in some Northern varieties of Spanish, even if the standard realisation, just like in Italian, is [tʃ] (Harris, 1969).

- Despite the sound being present in some Spanish varieties, Italian listeners referred that Spanish speakers did not utter correctly the palatal lateral approximant: [ʎ] in "*dagli*" (e.g. *"Dagli spalti" è letto con "li" invece che "gli"*[130]). Considering that in many Spanish varieties this

---

123. Translation into English: 'Reading "V" as if it were a "b"'.
124. Translation into English: 'Confusion between b and v with the Spanish labial the name of which I don't remember (the ipa symbol is the beta)'.
125. Translation into English: 'Uttering the word "ovvero" with the "v" pronounced as the typical Spanish "b"'.
126. Translation into English: '"Basilica" has an odd "s", like sss'.
127. Translation into English: 'She speaks with a lisp'.
128. Translation into English: 'She is saying "pensiero escaturito", instead of "scaturito"'.
129. Translation into English: ' "Vicinissima" is pronounced with a z'.
130. Translation into English: '"Dagli spalti" is read with "li" instead of "gl"'.

sound has merged with [j] (Pharies, 2007), this could indeed mean that some speakers are not able to perceive the contrast between the two sounds and therefore they did not utter the palatal lateral approximant properly.

- Most Spanish speakers did not produce the open-mid front unrounded vowel, realised instead as a close-mid unrounded vowel, thus [ɛ] as [e], in contexts such as "*attraverso*", "*colleghi*", "*Illegio*", "*rappresenta*", "*sempre*", "*tempo*" and "*castello*" (e.g. *Tutte le "e" sono chiuse*[131]).
- Additionally, Italian listeners commented that Spanish speakers failed to produce the open-mid back rounded vowel, which was instead realised as the close-mid rounded vowel, thus [ɔ] as [o], in words such as "*canonica*" (e.g. *O molto chiusa*[132]; *Le vocali aperte sono tutte chiuse*[133]). Spanish has only five vowels /i/, /u/, /e/, /o/ and /a/, therefore none of the two open-mid vowels is present in its phoneme inventory (Martínez Celdrán et al., 2003).

2. Suprasegmental features:

a) Word stress misplacement was reported in a few occasions: "*celebri*" ([ˈtʃɛːlebri] as *[tʃɛˈleːbri]), "*ovvero*" ([ovˈveːro] as *[ˈoβero], failing to utter the geminated consonant), "*parroco*" ([ˈparroko] as *[parˈroːko]) but if we compare Spanish accent to other accents (especially, French and English) this cue was less frequent (e.g. *Accento della parola "ovvero" sulla sillaba sbagliata*[134]).

b) Even if most cues reported for the Spanish accent were at the segmental level, there were various comments that referred to the prosodic features. They mentioned various cues, but especially intonation and rhythm (e.g. *Intonazione spagnola*[135]; *Intonazione della frase*[136]; *Intonazione inusuale*[137]; *In generale il ritmo del parlato*[138]; *Alcune pause tra una parola e l'altra*[139]; *Cadenza strana boh una specie di ritmo nel parlare*[140]; *Ritmo diverso*[141]; *Mancanza di intonazio-

---

131. Translation into English: 'All "e"s are closed'.
132. Translation into English: 'The "o" sound is different, very closed'.
133. Translation into English: 'The open vowels are all closed'.
134. Translation into English: 'The stress in "ovvero" is on the wrong syllable'.
135. Translation into English: 'Spanish intonation'.
136. Translation into English: 'Phrase intonation'.
137. Translation into English: 'Unusual intonation'.
138. Translation into English: 'In general the rhythm of how he speaks'.
139. Translation into English: 'Some pauses between one word and another'.
140. Translation into English: 'Strange cadence, I don't know, a sort of speech rhythm'.

ne della frase come se non sapesse fino a fondo ciò che dice[142]*; Pause in punti inconsueti del discorso, cadenza ritmica anomala*[143]). Upon analysing these comments, it is evident that the Spanish-accented Italian in the sample contains numerous segmental and suprasegmental cues indicating foreignness. This observation offers a plausible explanation as to why native Italian speakers excelled in recognising this accent during the first task of the accent perception experiment, achieving a recognition rate of 78% (see §5.2). At the same time, it also explains why the Spanish accent was regarded as one of the most marked among the six accents investigated in this study (see §5.3).

6.1.7 Converging and diverging features of accentedness

The preceding sections described the characteristics of Italian spoken with Russian, English, German, French, Romanian, and Spanish accents. The analysis was based on the assessments of 288 native Italian speakers who were asked to identify the most distinctive foreign accent cues in 21 speech samples, each from a different speaker. The acknowledged limitations of this method involve its potential simplicity in studying the perception of foreign accents and the subjectivity inherent in the features identified, as they are reliant on the listener's perception rather than objective instrumental analysis. Nevertheless, this listener-oriented approach, derived from perceptual dialectology, remains a valuable foundation for future empirical research on the topic. The obtained findings reflect what native Italian speakers perceived as the most salient features of foreign-accentedness and allowed me to establish a preliminary framework for accent-specific feature patterns. At the same time, this type of exploration has the potential to facilitate forthcoming analyses in the domain of L2 Italian speech.

Based on the previous sections of this chapter, it can be concluded that Italian speakers primarily reported segmental features in their comments. This was particularly true for the German and Spanish accents, that at the same time were also identified as the most marked accents in the experiments outlined in Chapter 5. In particular, the Spanish accent was identified with the highest accuracy of 0.78, making it the most accurately recognised foreign accent. While it is not implied that segmental cues play a more sig-

---

141. Translation into English: 'Different rhythm'.
142. Translation into English: 'Lack of phrase intonation, as if she didn't fully understand what she was saying'.
143. Translation into English: 'Pauses in strange points, unusual rhythmic cadence'.

nificant role in the perception of accentedness, it is worth noting that some of these features are accent-specific, while others are shared across all six accents. The following features, which were reported most frequently by the Italian listeners, are shared among all or some of the accents:

- lack of consonant gemination (all accents);
- non-native production of the alveolar trill (all accents);
- non-native production of the palatal lateral approximant (Russian, English, German, Romanian, and Spanish accents);
- non-native production of the palatal nasal approximant (Russian, English, and Romanian accents);
- confusions between the voiceless and the voiced alveolar fricatives (German, French, and Spanish accents);
- aspiration of the voiceless stops (English and German accents);
- non-native production of the open-mid front unrounded vowel (Romanian and Spanish accents);
- non-native production of the open-mid back rounded vowel (Russian, Romanian, and Spanish accents);
- vowel lengthening (English and French).

The listeners' feedback on suprasegmental features showed a general consistency across accents, although there were instances where comments were more frequent, as in the case of the English accent. These were the most commonly cited non-native prosodic features:

- word stress misplacement (all accents, even if for English and French this feature was the most common);
- non-native use of pauses and intonation (all accents);
- non-native rhythmic patterns (all accents).

Thus, the phenomena discussed in this section primarily relate to non-native contrasts (e.g., consonant elongation) or sounds that are similar but not identical to the native language (e.g., /o/ vs. /ɔ/). The inability to produce these sounds can be attributed to what Flege (1987, 1995, 2003) calls 'equivalence classification'. While, indeed, L1 transfer seems to account for most of these phenomena, in some cases, they may depend on other factors: overgeneralizations, as for instance in the case of unnecessary consonant gemination, or in the tendency of using the paroxytone – the most frequent type of lexical stress in Italian (Nespor, 1993, D'Imperio and Rosenthall, 1999; Marotta, 2016) – even for words with oxytone or proparoxy-

tone stress; the speaker's inexperience with a reading task in Italian; anxiety which could have led to non-native intonation or uncommon rhythm patterns.

Although, overall, more segmental features were reported for the six accents, it remains unclear whether these features were simply easier to identify and describe or if they truly were perceived as more salient than suprasegmental cues. To explore this question further, the second part of this chapter will introduce some computational accent classification experiments. The first experiment will focus on spectral features, while the remaining experiments will be centred around prosodic features. Unlike the feature naming task of the accent perception experiment, where only material produced by 21 speakers was used, in the next sections the speech of 100 speakers will be analysed. For each of them several samples will be used. Finally, the performance of the 288 native Italian speakers in identifying the foreign accent of the non-native speakers will be compared to the results obtained from automated accent classification experiments.

6.2 Spectral features and GMM-SVM[144]

As discussed in Chapter 2, only a handful of studies have so far examined the significance of specific segmental and prosodic features in determining the perceived foreign-accented Italian, such as Boula de Mareüil et al. (2004), Pellegrino (2012a), Romito et al. (2012), and De Meo et al. (2015). However, the influence of these non-native Italian features on the performance of automated accent identification systems, as well as on ASR systems, remains a realm yet to be explored. Therefore, one of the objectives of this study is to investigate the contribution of different feature types to the automated classification of foreign-accented Italian.

Chapter 3 presented several approaches to the automatic identification of foreign and regional accents, generally developed by means of complex speech technology techniques. Most research in this field has, in fact, been motivated by the desire to improve ASR. As previous studies have pointed out, accent variation within a specific language can have negative effects on the performance of ASR (Zheng et al., 2005; Brown, 2016; etc.). Indeed, it has been suggested that by identifying a speaker's accent upfront, it is

possible to build speech recognition models that can adapt more effectively to that particular accent (Weninger et al., 2019).

While this chapter will introduce approaches that may be perceived as somehow elementary, there is a strong belief that the outcomes derived from these experiments will lay the groundwork for more advanced computational explorations in the field of foreign-accented Italian. Moreover, such analyses have the potential to elucidate the differences and similarities among the accents investigated in this study, offering insights into managing foreign accent cues within Italian ASR. This type of study holds potential not only for the field of linguistics but also for forensic sciences. In the last decades, advancements have been made in speaker profiling tasks, especially for English, with several tools being developed to analyse speech samples, typically phone call recordings, and extract information about unknown speakers (Romito, 2013). These tools strive to provide insights into various attributes such as age, gender, and origin (based on the accent). Nevertheless, using automated tools for speaker profiling raises concerns regarding privacy, biases, and the ethical implications of inferring personal information from speech data. Therefore, great attention is needed when applying such tools in automated tools for speaker profiling in real-world contexts (Koenecke et al. 2020).

Continuing with the analyses, for the sake of feasibility and comparability, the accent identification experiments were conducted using exclusively audio material obtained through the reading task technique. As mentioned earlier in Chapter 4, the speakers were instructed to read a 249-word text that was extracted from an Italian newspaper article. The selected reading task encompassed different levels of linguistic complexity and included all Italian phonemes. In order to minimise the impact of unfamiliarity with the text on the speakers' reading performance and, subsequently, on the experimental outcomes, the second reading attempt was used. The objective of the study was to classify speech samples obtained from a diversified group of 100 speakers, which comprised 14 individuals representing each non-native accent and 16 Italian speakers. These speakers exhibited distinct socio-demographic and socio-cultural characteristics, including factors such as gender, age of onset, duration of stay in Italy, method of learning Italian, presence or absence of pronunciation training, and proficiency level in Italian. To conduct automated analysis, ten different utterances for each speaker were included.

The system being presented here was a baseline model that used spectral features. As discussed in Chapter 3, this approach falls under the category of acoustic methods, as it primarily focused on extracting spectral information from the speech sample. In contrast, a phonotactic system would

consider factors such as the presence, type, order, and frequency of phones in the speech sample. This accent classification system consisted of two main processes: feature extraction (also known as speech parametrization) and pattern matching. In the feature extraction process, speech characteristics were extracted from the voice signal, resulting in a compact set of feature vectors that represented the speaker and/or accent. These features were then used to represent the speaker and/or the accent. On the other hand, the pattern matching process involved the techniques used to classify the speaker by comparing their extracted features to the known features for that particular accent.

The first step in setting up the system consisted in feature extraction, therefore retrieving acoustic representations from the speech samples. Scholars in the field of acoustic and experimental phonetics have often opted for fine-grained short-time spectral analyses to investigate cues of native and non-native speech (see Alsteris and Paliwal, 2007, for a review). Even if this type of approach to speech processing offers a lot of rich information, it may only be applied to a limited amount of speech material, because the technique is time-consuming. For this reason, in speech technology tasks, especially in ASR, more compact representations of speech are needed (Hönig, 2017). In fact, due to their compactness, Mel-Frequency cepstral coefficients (MFCC) have become the most popular feature in speech processing.

For the accent identification experiment, 13-dimensional MFCCs features were extracted; short-time energy was substituted for C_0 (i.e. the zeroth cepstral coefficient value in the specified frame); delta and delta-delta features computed over a window of five consecutive frames and then appended, for a total of 39 features. As mentioned above, MFCCs are rather common features in signal processing, and their first uses are attributed to Bridle and Brown (1974) and Marmelstein, (1976). MFCCs are coefficients of the Mel-frequency cepstrum (MFC), namely a representation of the power spectrum of a sound, computed on the Mel-bands, by means of a series of transforms (Formiga and Alias, 2009). Broadly speaking, the Mel scale relates perceived pitch to the actual measured value of its frequency (Stevens et al., 1937; Pedersen, 1965). The reason why this scale was developed is to explain the human auditory system, which does not interpret pitch linearly; on the contrary it perceives it on a logarithmic scale.

The process of MFCC features extraction is described in detail in Furui (1994) and Gordon (1998). In brief, the speech sample is pre-processed with a pre-emphasis filter, then it is split into frames of 25 milliseconds and a Hamming window is applied to each frame. The

Discrete Fourier Transform (DFT) is performed on each frame. Frequencies are linearly spaced into a Mel frequency bank. The logarithm is taken from the filter bank energies and then a Discrete Cosine Transform (DCT) is applied. MFCCs alone do not capture enough information for classification, identification, or verification tasks. For this reason, many accent identification systems based on spectral features generally append delta and delta-delta (or double-delta) cepstral features to static cepstral features in order to exploit temporal and dynamic information (Kumar et al., 2011), similar to those used in other tasks of ASR. In fact, early research in speech technology showed that the performance of a speech recognition system can be improved by delta and delta-delta features to the basic static parameters (Hanson and Applebaum, 1990). Deltas are first order regression coefficients, and the formula used to extract them is the following:

$$d_t = \frac{\sum_{n=1}^{N} n(c_{t+n} - c_{t-n})}{2 \sum_{n=1}^{N} n^2}$$

where d_t is a delta coefficient, from frame t calculated on the corresponding static coefficients c_{t+n} and c_{t-n}. Likewise, delta-delta features are second order regression coefficients, and they are defined in terms of a subsequent delta-operation on the delta features (Kumar et al., 2011).

To sum up, all the features presented above (MFCCs, energy, deltas, and delta-deltas) were used to represent the accent classes in the training phase, and subsequently the single speech samples in the testing phase. In order to model the accents, a text-independent Gaussian Mixture Model (GMM) – Universal Background Model (UBM) – Support Vector Machines (SVM) system was chosen (Hou et. al, 2010; Biadsy et al., 2011; Brown, 2016; etc.). The UBM consisted of 256 Gaussians with diagonal covariance, adapted to each recording using a *maximum a posteriori* (MAP) estimation, and reduced to a supervector consisting in means only (Campbell et al., 2006; Steidl et al., 2011). The idea behind this approach was to identify an input speech by selecting one model from a set of enrolled models. In order to train classifiers for each of the target accents, Chang and Lin's (2011) LibSVM – a library for Support Vector Machines (SVM) – was used. Since the number of features was high, a linear kernel was deployed. After a parameter tuning, the penalty parameter C was set to 10. The training, the enrolment, and the test data were even. Therefore, the UBM was trained using speech material produced by speakers for all the accents investigated. Unlike the GMM-UBM, the GMM-SVM used

speaker-specific enrolment data. Therefore, the models adapted did not correspond to specific accents; on the contrary, they corresponded to the speakers in the enrolment data, resulting in various GMMs for each accent.

Brown (2016: 96) provides a detailed explanation of how the accent identification is generally performed in experiments on accent classification. In brief, these systems employ a speaker-specific GMM for each accent group. The means of the GMMs are concatenated to create a vector representing each speaker, which is used as input for an SVM. Then, the SVM plots these vectors in multi-dimensional space, along with GMM means from all other accent classes to form a 'one-against-the-others' configuration. This allows for the creation of an optimal hyperplane between the accent class and the rest. Each accent class forms an SVM in this manner when rotated. During testing, an unknown speaker's speech sample is used to adapt a model from the Universal Background Model (UBM), and the mean vector is introduced to each SVM formed for each accent class. The accent label is determined based on the clearest margin formed with the hyperplane.

The acoustic GMM/SVM system used in the present study on foreign-accented Italian demonstrated a discrete performance, consistent with previous experiments focusing on accent classification (Vieru et al., 2011; Brown, 2016) and studies pertaining to speaker verification (e.g., Chao et al., 2009). Table 6.2 shows the confusion matrix with absolute values (i.e. the recall/accuracy[145] for the number of utterances identified correctly for that specific accent), while Table 6.3 transforms the data into proportions.

Table 6.2 – Confusion matrix for the acoustic GMM/SVM system (absolute)

	ENG	FRA	GER	ITA	ROM	RUS	SPA
ENG	**40**	12	34	22	12	10	10
FRA	22	**68**	12	8	12	12	6
GER	22	18	**37**	6	14	27	16
ITA	3	1	9	**82**	35	11	19
ROM	12	1	10	45	**37**	29	6
RUS	9	3	9	9	18	**83**	9
SPA	4	10	10	18	15	4	**79**

145. In machine learning and classification, the value for recall indicates what proportion of actual positives was identified correctly (Powers, 2011).

Table 6.3 – Confusion matrix for the acoustic GMM/SVM system (proportions)

	ENG	FRA	GER	ITA	ROM	RUS	SPA
ENG	**0.29**	0.09	0.24	0.16	0.09	0.07	0.07
FRA	0.16	**0.48**	0.09	0.06	0.09	0.09	0.04
GER	0.16	0.13	**0.26**	0.04	0.10	0.19	0.11
ITA	0.02	0.01	0.06	**0.52**	0.22	0.07	0.11
ROM	0.09	0.01	0.07	0.32	**0.26**	0.21	0.04
RUS	0.06	0.04	0.06	0.06	0.13	**0.59**	0.06
SPA	0.03	0.07	0.07	0.13	0.11	0.03	**0.56**

The achieved overall recognition rate (recall/accuracy) of 0.42 (chance expectation: 0.14) is comparable to the performance exhibited by native Italian speakers (0.47). Notably, the system exhibits a particularly robust performance for the Russian accent (0.59), where native listeners did not perform well (see §5.2). However, the rather poor performance for English (0.29) and German (0.26) accents, even if quite comparable to that of listeners, is indeed quite surprising. The GMM/SVM system delivers a good performance for the Spanish accent, achieving a recognition rate of 0.56. However, this falls short of the accuracy achieved by humans, which is 0.78. Similarly, for the Italian accent, the system achieves an identification score of 0.52, compared to the 0.95 obtained by Italian listeners. Lastly, the system identifies the Romanian accent with a recognition rate of 0.26, slightly outperforming human raters who achieved 0.22, but the overall performance remains poor.

Upon analysing the classification errors, it appeared that many English-accented samples were wrongly classified as German, and German speakers were often misclassified as English or Russian. Inaccurately identified French speakers were sometimes confused with English speakers. A considerable number of Romanian speakers were incorrectly classified as Italian, and the same was observed for some Italian speakers who were mistakenly identified as Romanian. Misclassified Russian speakers were primarily classified as Romanian, while misidentified Spanish speakers were often confused with Romanian or Italian speakers. Further details regarding these findings will be presented at the end of this chapter.

6.3 Prosodic features and SVM

During the accent perception experiment, the native speakers of Italian encountered difficulties in providing accurate judgments concerning prosody, despite most of them recognising that there were suprasegmental features at play. Moreover, interpreting an accent classification system based on MFCC-derived features (see §6.2) can be linguistically opaque and challenging. In light of the monograph's focus, the objective was to build a more linguistically meaningful system. A prosody-based machine learning system for accent classification, such as the one that will be outlined in this section, appeared useful in enhancing our understanding of the underling suprasegmental patterns that distinguish the accents under investigation.

Consequently, a set of 961 prosodic features was extracted from the entire read speech material of 100 speakers, comprising 84 non-native speakers (14 for each accent) and 16 Italian speakers, during their second reading attempt. For the sake of brevity, a concise overview, based on the descriptions provided Hönig (2017) will be provided below. Other details can also be found in Hönig's works (2014a, 2014b, 2017), which served as the source of the feature extractor used here. Therefore, the included features were as follows (the descriptions are adapted from Hönig, 2017):

1. Fundamental frequency (f0): The logarithm of the fundamental frequency, estimated using the RAPT algorithm, normalised by subtracting the mean over the speaker. It is undefined outside voiced regions, and the mean is computed only on voiced frames.
2. Energy: The logarithm of the intensity, normalised by subtracting the mean over the speaker.
3. Energy voiced: The logarithm of the intensity for voiced frames (undefined elsewhere), normalised by subtracting the mean over the speaker.
4. Duration: Absolute word duration divided by the number of syllables in speech units. It includes normalised duration of the speech unit with respect to the entire utterance, and other types of normalised values.
5. Pauses: Duration of silent and filled pauses before and after the word.
6. Jitter: A measurement of perturbation, indicating the variation and instability of f0.
7. Shimmer: A measurement of disruption, indicating the instability of intensity/amplitude.

In order to segment pauses, vowels, consonants, and speaker noise, the phoneme recogniser of the Brno University of Technology for Hungarian was used (Schwarz et al., 2006). The selection of this recogniser was based on its accessibility as well as the extensive inventory of phonemes in Hungarian, which enabled the recognition of a diverse range of productions. Given that the majority of the recorded speakers were non-native speakers of Italian, this choice was deemed more suitable, as an Italian recogniser would not have been able to capture all the non-native cues present. The feature extraction approach adopted here did not account for lexical stress and did not employ actual Italian syllables. Having exploited just the vowel-consonant division allowed to implement a language-independent pipeline that could be readily applied to the data at hand. Hence, language-dependent speech recognisers or pronunciation lexicons were not required. Also, the vowel-consonant segmentation technique used here might be more advantageous than the approaches that employ the voiced/voiceless distinction (Maier et al., 2009). (Pseudo-)syllables were derived from consecutive vowels and trailing consecutive consonants (i.e. leading consonants after pauses were ignored). For each (pseudo-)syllable, 79 micro-structural prosodic descriptors were computed (Batliner et al., 2000). However, it is important to note that the basic prosodic features alone could not be directly employed for classification tasks due to their inherent speaker-dependent characteristics. For this reason, the features were normalised with respect to their mean values across the whole utterance. To obtain a fixed number of features per recording, twelve functionals were computed across the local syllable descriptors that capture statistical and temporal properties: mean, standard deviation, minimum, maximum, median, quantiles 5%, 25%, 50%, 75%, 95%, average absolute local change (similar to Grabe's raw pairwise variability index rPVI (Grabe and Low, 2002), root average squared local change, and the slope of the regression line. Applying this functionals generated 79 * 12 = 948 general-purpose prosodic features.

Based on the segmentation of the phoneme recogniser into vocalic and consonantal intervals, other features were computed: Grabe's raw pairwise variability index rPVI (Grabe and Low, 2002) on consonants and vowels, as well as its (rate-of-speech-) normalised version nPVI. Additionally, variants of Grabe's measures using squared instead of absolute differences (root average squared local change) were obtained. This included 8 features reflecting local variability in durations. Also, 5 features reflecting global variability and proportions were given by Ramus' (2002) %V (percentage of vocalic intervals) and vocalic and con-

sonantal deltas (the global standard deviations of durations), as well as Dellwo's (2010) variation coefficient Varco for vowels and consonants (rate-of-speech normalised standard deviation of durations).

For the first experiment the sequential minimal optimization (SMO) SVM implementation of the WEKA machine learning toolkit[146] was used. Considering that the number of features was quite high, a linear kernel was employed. After the parameter optimization, the penalty parameter *C* was set to 1. The sampling method used for this classification task was leave-one-speaker-out[147] – that appeared the most adequate type of sampling, since the number of speakers was not excessively large. The model converged within an hour. The prosodic SVM system yielded an overall recall/accuracy of 0.36 (chance expectation: 0.14), behind the results achieved by the native speakers of Italian and the performance of the GMM/SVM system. However, the interesting fact about these results is that confusions made by this system seemed closer to the human performance, with results for Russian and Romanian accents strongly affecting the overall score.

The confusion matrix in Table 6.4 displays the raw values for recall/accuracy in terms of correctly classified speakers as belonging to a specific accent; Table 6.5 transforms the data into proportions.

Table 6.4: Confusion matrix for the prosodic SVM system (speakers)

	ENG	FRA	GER	ITA	ROM	RUS	SPA
ENG	**6**	1	3	0	1	1	2
FRA	1	**6**	1	1	1	3	1
GER	3	1	**5**	1	1	2	1
ITA	0	1	1	**8**	3	1	2
ROM	2	2	1	4	**3**	1	1
RUS	2	2	2	1	1	**3**	3
SPA	3	1	1	2	0	1	**6**

146. Available at: https://www.cs.waikato.ac.nz/ml/weka/ (accessed on the 5th of January 2023).

147. Leave-one-out is a cross-validation sampling method, where *p* – the number of observations in the validation set – is 1, while all the other observations represent the training set. Therefore, in this case, the model is trained on every speaker except the one that is left out and then it computes the statistics and evaluations on that speaker (Kneser et al., 1995).

Table 6.5: Confusion matrix for the prosodic SVM system (proportions)

	ENG	FRA	GER	ITA	ROM	RUS	SPA
ENG	**0.43**	0.07	0.21	0.00	0.07	0.07	0.14
FRA	0.07	**0.43**	0.07	0.07	0.07	0.21	0.07
GER	0.21	0.07	**0.36**	0.07	0.07	0.14	0.07
ITA	0.00	0.06	0.06	**0.50**	0.19	0.06	0.12
ROM	0.14	0.14	0.07	0.29	**0.21**	0.07	0.07
RUS	0.14	0.14	0.14	0.07	0.07	**0.21**	0.21
SPA	0.21	0.07	0.07	0.07	0.14	0.00	**0.43**

Although the overall classification rate is low, there were notable improvements in identifying specific accents using this system. For instance, six English-accented speakers were correctly classified, achieving a recall/accuracy rate of 0.43 for this accent, which surpasses the performance of both human raters (0.30) and the GMM/SVM system (0.29). Similarly, the system successfully classified the French accent in six cases, yielding a recall/accuracy rate of 0.43, outperforming human raters (0.30). However, there were challenges in accurately classifying other accents, with frequent confusion between the Romanian and Italian accents, as well as between the English and German accents.

6.4 Prosodic features and other machine learning classifiers

Considering that the overall classification rate of the prosodic SVM system was not satisfactory, additional experiments were conducted using the same set of features but employing different machine learning classifiers. In particular, for these experiments, the data mining Orange package (Demsar et al., 2013)[148] for Python was used. In these experiments, for each class, including the native variety of Italian, 14 speakers were analysed. The classifiers tested were the following:
- k Kearest Neightbour (kNN) which predicts according to the nearest training instances: number of neighbours: 6; metric: Euclidean, weight: uniform.
- Logistic Regression which uses a logistic function to model a categorial dependent variable (the accent); regularization type: Ridge (L2); strength: C = 1.

148. A description of this resource is available at: https://orange.biolab.si (accessed on the 5th of January 2023).

- Random Forest which builds a set of decision trees based on an arbitrary subset of attributes, from which the best attribute for the split is selected; the final model is chosen on the majority vote from individually developed trees in the forest; number of trees: 10 (splits: ≥ 5).
- Naïve Bayes which is based on Bayes' theorem that assumes strong (naïve) independence between the features (McGrayne, 2011).
- Neural Network which uses *sklearn*'s Multi-layer Perceptron (MPA)[149] algorithm with backpropagations that can learn non-linear as well as linear models; neurons per hidden payer: 100, activation: ReLu, solver: Adam, Alpha: 0.0001; maximum iterations: 200.

Regarding the sampling, the machine learning classifiers were tested in two different circumstances: a random sampling (repeat train/test: 10; training set size: 65%, testing set size: 35%) and a leave-one-speaker-out sampling (i.e. this technique used the entire model fit to all the speakers except for one single speakers, and then made a prediction on that speaker that could be compared to the actual value). The results were nearly identical, thus only the outcomes for the leave-one-speaker-out sampling technique will be presented, as it is more suitable for this task, despite being time-consuming. Table 6.6 shows the macro-averaged results across the classes.

Table 6.6: Summary of the results yielded by the 5 classification methods (abbreviations: AUC = area under the curve; CA = classification accuracy, F1 = F-score)

Method	AUC	CA	F1	Precision	Recall
Neural Network	0.93	0.66	0.65	0.66	0.66
Random Forest	0.83	0.51	0.51	0.51	0.51
Naïve Bayes	0.81	0.47	0.47	0.49	0.47
Logistic Regression	0.79	0.45	0.44	0.44	0.45
kNN	0.62	0.25	0.24	0.24	0.25

149. A description of this resource and algorithms are available at: http://scikit-learn.org/stable/modules/neural_networks_supervised.html (accessed on the 1st of February 2023).

Next, Table 6.7 presents the confusion matrix, illustrating the classification of speakers using the prosodic Neural Network system, which yielded the highest performance in terms of classification accuracy/recall (0.66). In Table 6.8 the values are transformed to proportions.

Table 6.7: Confusion matrix for the prosodic Neural Network system (speakers)

	ENG	FRA	GER	ITA	ROM	RUS	SPA
ENG	**10**	0	1	0	0	1	2
FRA	1	**11**	0	0	1	0	1
GER	1	1	**10**	0	1	1	0
ITA	0	0	0	**10**	2	1	1
ROM	0	1	1	3	**8**	0	1
RUS	1	0	2	0	1	**10**	0
SPA	0	1	1	1	1	1	**9**

Table 6.8: Confusion matrix for the prosodic Neural Network system (proportions)

	ENG	FRA	GER	ITA	ROM	RUS	SPA
ENG	**0.71**	0	0.07	0	0	0.07	0.14
FRA	0.07	**0.79**	0	0	0.07	0	0.07
GER	0.07	0.07	**0.71**	0	0.07	0.07	0
ITA	0	0	0	**0.71**	0.14	0.07	0.07
ROM	0	0.07	0.07	0.21	**0.57**	0	0.07
RUS	0.07	0	0.14	0	0.07	**0.71**	0
SPA	0	0.07	0.07	0.07	0.07	0.07	**0.64**

The prosodic Neural Network system exhibits a superior overall performance in accent classification, surpassing all previously discussed methods in this chapter, and displaying an improvement of nearly 0.2 compared to native Italian speakers' performance. In particular, the for the identification of French accent it achieves the highest accuracy (0.79), with 11 out of 14 speakers correctly predicted by the system. Following closely are the English, Russian, German, and Italian accents (0.71), each with 10 accurately identified speakers. The system demonstrates accurate classification for 9 out of 14 Spanish speakers (0.64), while also successfully identifying 8 Romanian speakers (0.57).

Lastly, an investigation was conducted to determine the most informative prosodic features for the accent classification task. Two scoring methods, namely the chi-square statistic and ReliefF, were uses for this purpose. The chi-square statistic measures the dependence between the feature and the class, while ReliefF calculates the strength of a feature in distinguishing between classes on similar data instances. The most highly ranked features for the two measurements are grouped and summarised below:

- 50% and 75% quantiles and mean value for the regression coefficient of energy within the word;
- 50% and 75% quantiles and mean values for the normalised energy within the word;
- maximum values for energy within the word;
- mean square error of the energy curve with respect to the regression curve;
- 5%,25%,50%,75%, 95% quantiles and mean values of f0 within word;
- 25% quantile of root mean squares and standard deviation of jitter;
- 75% quantile and mean values of the absolute word duration;
- Filled pauses before the word.

It appears that energy, f0, word duration, jitter, and filled pauses are the most informative prosodic cues for distinguishing among the accents. These findings are intriguing, given that various studies have demonstrated a connection between segmental length, voice, and energy (Rämö et al., 2004; Horák, 2012; Broussard et al., 2017, etc.). According to these rankings, the variation in the accents can also be partially attributed to features derived from normalised f0 and duration.

The effectiveness of the new features in enhancing the classification of the seven accents was evaluated through the creation of a FreeViz projection[150] as shown in Figure 6.1.

150. Documentation available at: https://orange3.readthedocs.io/projects/orange-visual-programming/en/latest/widgets/visualize/freeviz.html (accessed on the 1st of February 2023).

Figure 6.1: FreeViz map of accents
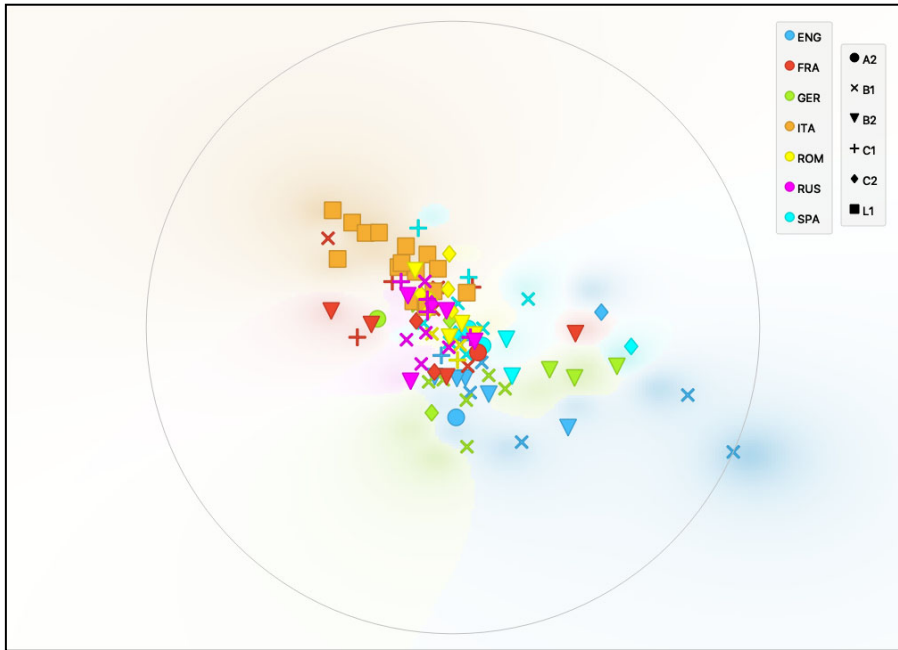
This visualization depicts the speakers, accents, and the speakers' proficiency level in Italian (this element will be discussed in §6.5). Some general trends were observed, although in line with the findings of the previous automated classification experiments. The Italian speakers formed two cohesive clusters, with the central group positioned closer to both Romanian and Russian speakers. Similarly, the Russian speakers clustered tightly and showed proximity to the Romanian speakers. Among the Spanish speakers, half exhibited a scattered distribution while the other half clustered near the Romanian and Italian speakers. In contrast, the French, English, and German speakers were distinctly separated from the Italian group, displaying a scattered distribution across the map. Another classification experiment of the ranked features yielded an overall recognition rate of 0.42 (method: Neural Network; sampling: leave-one-speaker-out); the poorer results likely suggest an overfitting.

6.5 Clustering experiments

In addition to the accent classification experiments, exploring the potential of unsupervised clustering algorithms was undertaken to group speak-

ers based on the comprehensive set of 961 prosodic features described earlier in this chapter. The objective was to examine whether the resulting clusters corresponded to distinct L1/accent groups or if the proximity of speakers could be attributed to their socio-demographic and socio-cultural characteristics.

First, an unsupervised hierarchical clustering analysis was conducted using the Distance and Hierarchical Clustering widgets in Orange. The analysis involved the use of normalised values, with the Euclidean distance metric selected and the average linkage computed using Ward's method[151] (Ward, 1963). The cluster obtained is shown in Figure 6.2.

A preliminary result that emerged from this chart is that, based on their prosodic characteristics, speakers were separated into many clusters with various limbs and branches, and the logic behind this grouping is not necessarily related to their L1. Although the clustering displayed indeed various ramifications, one interesting finding is that 22 speakers out of 100 (i.e. 1 Russian, 3 French, 2 English, 8 Romanian and 8 Spanish) were under the same branch with 8 Italian speakers.

Figure 6.2: Hierarchical clustering
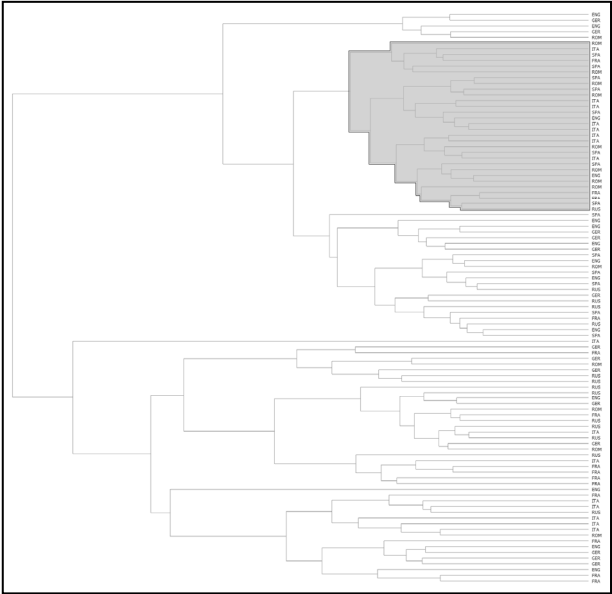
151. Ward's method of hierarchical clustering deals with clusters in terms of an analysis of variance, rather than computing the distance metrics or the measures of association. This method starts the analysis from the leaves that are grouped into branches, the branches into limbs and the limbs into the trunk (Ward, 1963).

On the one hand, the English, the French and the Russian speakers generally had an advanced level in Italian – either C1 or C2 – and their sociolinguistic information suggested an adequate linguistic input for Italian (i.e. pronunciation training; combined learning methods; long stay in Italy, etc.). This was true to some extent also for the Romanian speakers. On the other hand, the Spanish speakers had different levels of proficiency in Italian and their sociolinguistic background was rather varied. Examining the values of these factors for the English, French, and Russian subgroups highlighted in blue in Figure 6.2 and the data in Table 6.9, it appeared that an advanced level in Italian and a rich linguistic input contributed to a more native-like suprasegmental production in Italian. Conversely, for L1 speakers of Spanish, a prosodic performance similar to that of a native speaker of Italian was possible independently of the speaker's sociolinguistic background. With respect to the subgroup of Romanian speakers who were closely associated with the Italian group, it is worth noting that based on the sociolinguistic information provided, the majority had received substantial and high-quality exposure to the Italian language.

Table 6.9 displays the twenty-two L2 speakers who shared the same branch with the eight L1 Italian speakers. However, it is crucial to approach the interpretation of these language groupings from the unsupervised hierarchical clustering method with caution, considering the scoring approach and outcomes of the Neural Network classification discussed in the previous section.

Table 6.9: Information regarding the 22 speakers situated under the same branch as native speakers of Italian

ID	L1	Gender	Italian level	Age of onset	Learning method	Pronunciation training	Length of stay
29	ROM	M	C2	infancy	guided and naturalistic	yes	6-12 months
28	ROM	M	C2	adolescence	guided and naturalistic	yes	> 24 months
26	ROM	F	C2	infancy	guided and naturalistic	yes	> 24 months
25	ROM	F	C2	adolescence	naturalistic	no	> 24 months
24	ROM	F	C2	adulthood	naturalistic	no	> 24 months

11	FRA	F	C2	adolescence	guided and naturalistic	no	> 24 months
2	ENG	M	C2	infancy	guided and naturalistic	yes	> 24 months
30	RUS	F	C2	adulthood	naturalistic	no	> 24 months
22	ROM	F	C1	adolescence	naturalistic	no	> 24 months
7	SPA	F	C1	adulthood	naturalistic	no	> 24 months
5	SPA	F	C1	adulthood	guided and naturalistic	no	> 24 months
1	ENG	F	C1	infancy	guided and naturalistic	no	6-12 months
13	FRA	F	C1	adulthood	guided and naturalistic	no	12-24 months
12	FRA	F	C1	adulthood	naturalistic	no	6-12 months
27	ROM	M	B2	adulthood	naturalistic	no	> 24 months
23	ROM	F	B2	adulthood	guided and naturalistic	no	6-12 months
10	SPA	M	B2	adulthood	naturalistic	no	6-12 months
8	SPA	F	B1	adulthood	naturalistic	no	6-12 months
6	SPA	F	B1	adulthood	naturalistic	no	6-12 months
4	SPA	F	B1	infancy	naturalistic	no	6-12 months
3	SPA	F	B1	adulthood	naturalistic	no	6-12 months
9	SPA	M	B1	adulthood	naturalistic	no	6-12 months

Another unsupervised clustering technique tested in this study was k-Means, a commonly used method in cluster analysis for vector quantization (Hartigan and Wong, 1979). Instead of manually setting a predetermined number of clusters, the initialization function in Orange was used, automatically determining the optimal number of clusters. According to the silhouette scores presented in Table 6.10, it seems that the ideal number of clusters is 3, rather than the total number of L1s/accents (7).

Table 6.10: k-Means silhouette scores

Number of clusters	Silhouette scores
2	0.461
3	0.469
4	0.460
5	0.394
6	0.389
7	0.388
8	0.357

As it will be shown below, this clustering outcome did not seem to depend on the speaker's L1/accent. The data were confirmed by the multidimensional scaling (MDS)[152], where similar speakers are closer to each other and connected by means of visible similarity pairs. Figure 6.3 illustrates the formation of three distinct clusters (C1, C2, C3) with 5% jittering applied for improved visibility of all speakers. In C1 (indicated by circles), there were 12 Italian speakers, along with 9 Romanian, 8 German, 7 French, 7 Spanish, 6 Russian, and 6 English speakers. C2 (indicated by X) consisted of 8 Russian, 6 French, 6 Spanish, 5 German, 4 Romanian, 3 English, and 3 Italian speakers. Lastly, C3 (indicated by triangles) included 5 English speakers, 1 Spanish speaker, 1 French speaker, 1 German speak-

---

152. In this case, MDS provides a representation of the pattern of proximities (similarities or distances) within a set of speakers. For instance, in Figure 6.5, MDS plots the speakers on a map such that the speakers sharing similar values for prosodic features with other speakers are placed together or close, and, on the contrary, the speakers that display different prosodic patterns will be placed further away.

er, and 1 Romanian speaker. A closer examination of the map revealed that most Romanian speakers were situated near the Italian speakers. There were also instances where several Russian, German, and some Spanish speakers appeared in the same cluster, close to the Italian and Romanian speakers, although they did not consistently form cohesive groups. Additionally, many French, English, and German speakers exhibited a scattered distribution, with only few individuals located in proximity to the Italian groups.

Figure 6.3: MDS map
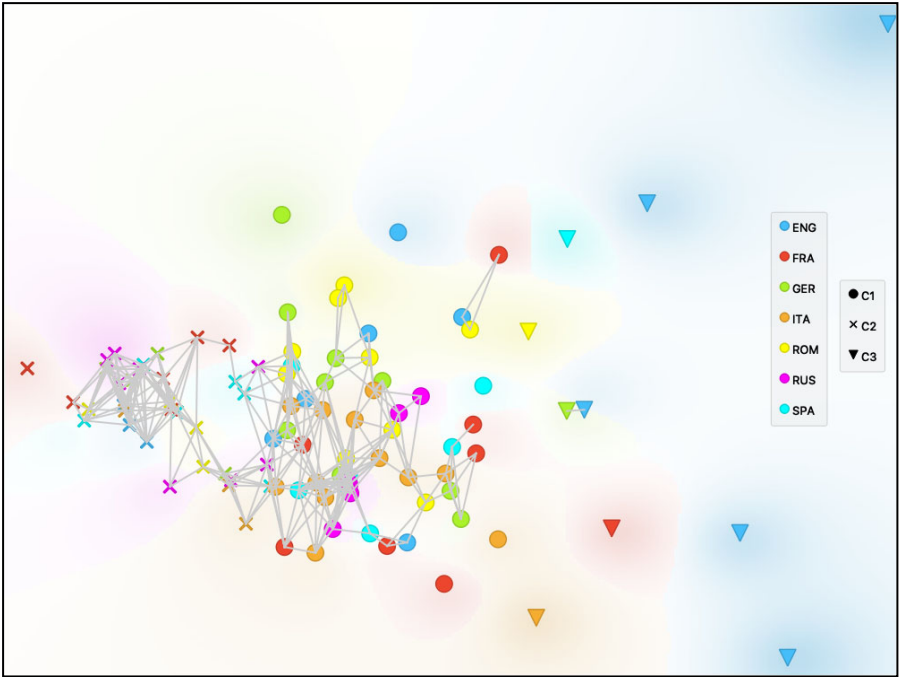
Similar to the Hierarchical unsupervised method, this technique provided insights into the underlying patterns of different groups based on the prosodic features discussed in previous sections. Grey strings are used to represent similar pairs of speakers. In Figure 6.4, the proficiency levels of speakers in Italian are displayed, confirming the observations from Figure 6.3. It is evident that most Italian speakers were closely connected and located in proximity to each other. Additionally, it is noteworthy that several English speakers were significantly distant from the cluster that contained the Italian speakers. Those English speakers who are grouped with Italian

speakers typically had a high level of proficiency in Italian. A similar pattern emerged for Russian and Spanish speakers. Despite the majority of Romanian speakers being grouped with Italian speakers, it is intriguing that two Romanian speakers who self-reported having a C2 proficiency level in Italian were relatively distant from the Italian cluster.

Figure 6.4 – MDS map showing proficiency levels in Italian
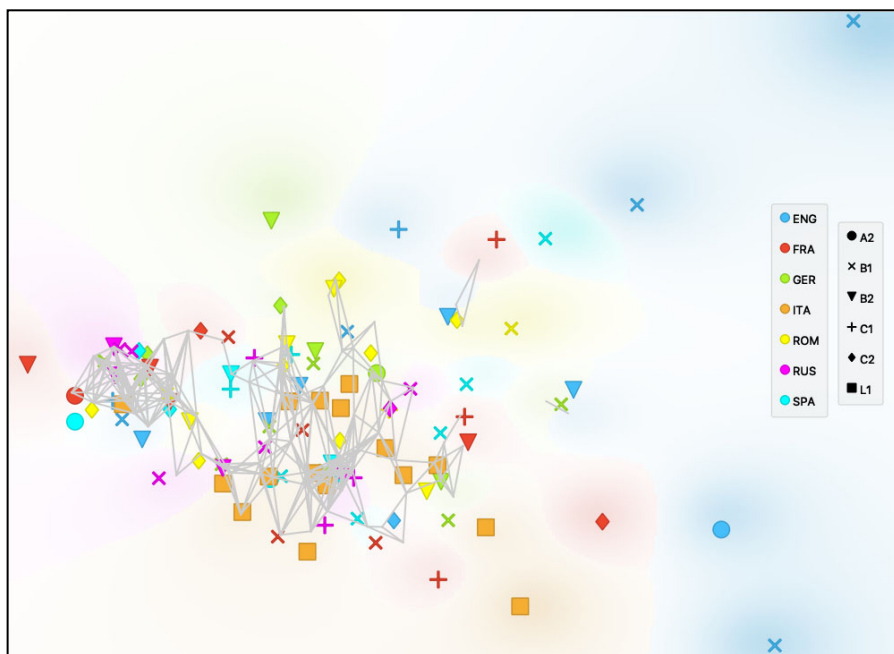
In addition to the variable regarding the proficiency level in Italian, an investigation was conducted into how other socio-demographic and socio-cultural factors were represented on the MDS map. Examining the length of stay in Italy among speakers (Figure 6.5) revealed intriguing patterns, albeit with some inconsistencies. It appeared that the majority of non-native speakers, particularly those having Russian, English, and Spanish as L1, who shared the same cluster with Italian speakers, had resided in Italy for a minimum of 12 months at the time of recording. However, it is important to note that this trend did not hold true for all speakers, as the Italian cluster also included German and Romanian speakers who had lived in Italy for less than 12 months.

With regard to the Italian language learning approach illustrated in Figure 6.6, although no clear-cut trends were readily discernible, a closer analysis of the map uncovered interesting findings. Speakers positioned further away from the native Italian cluster, especially those situated on the right-hand side of Figure 6.6, had acquired the L2 language through naturalistic means, with some individuals having Spanish, English, or Romanian as their L1. However, it is worth noting that the Italian cluster encompassed speakers from various language acquisition paths, suggesting that the learning path alone did not necessarily determine cluster membership.

The finding mentioned above was reinforced by the results of the pronunciation training analysis depicted in Figure 6.7. The majority of speakers who were relatively distant from the Italian cluster had not received any pronunciation training. However, in the proximity of the Italian speakers, there were both speakers who had received pronunciation training and those who had not.

Figure 6.5: MDS map showing the length of stay in Italy
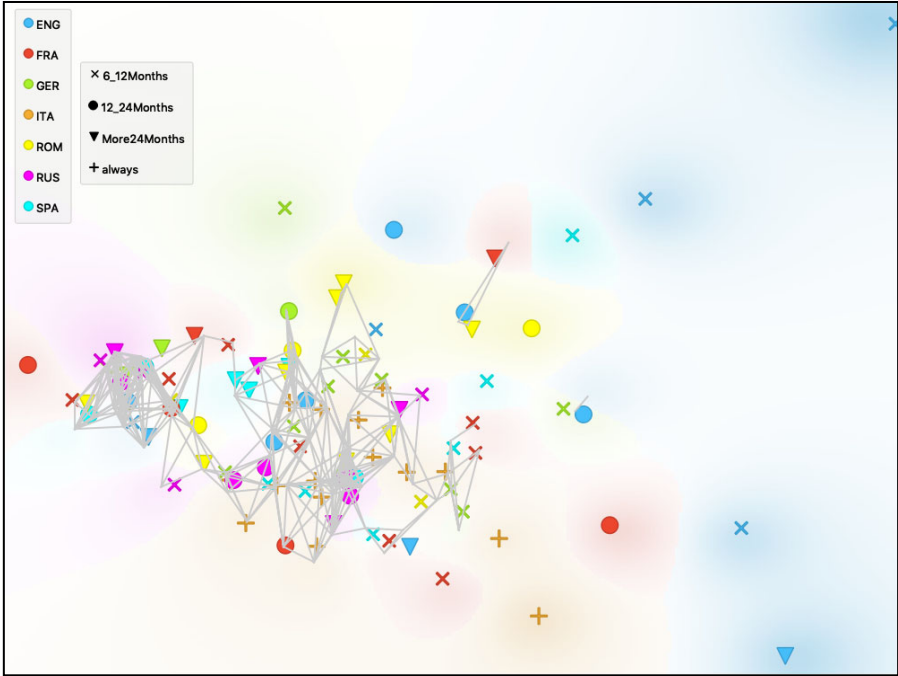
Figure 6.6: MDS map showing the Italian learning method
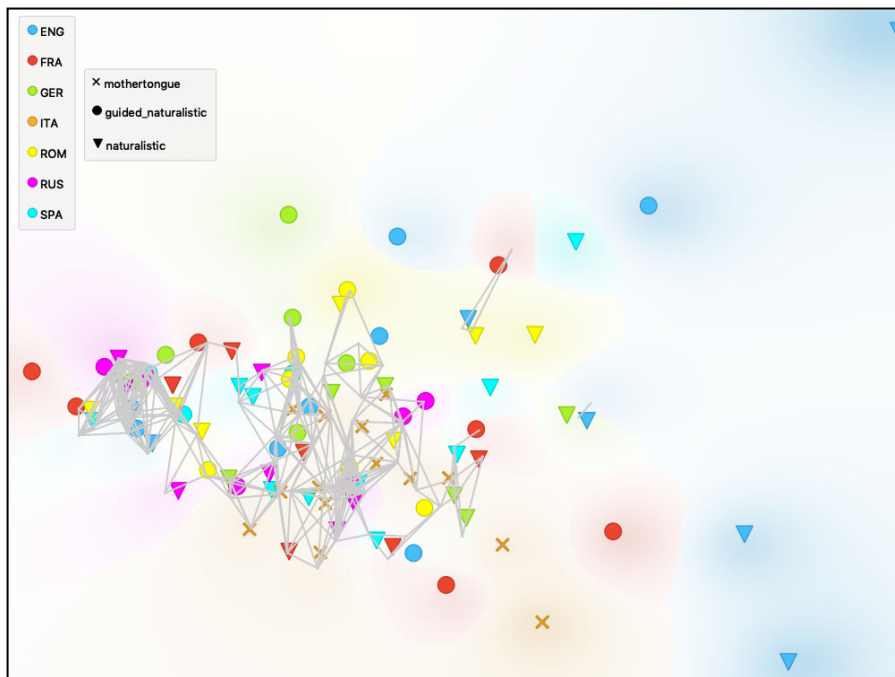
Figure 6.7: MDS map showing the pronunciation training
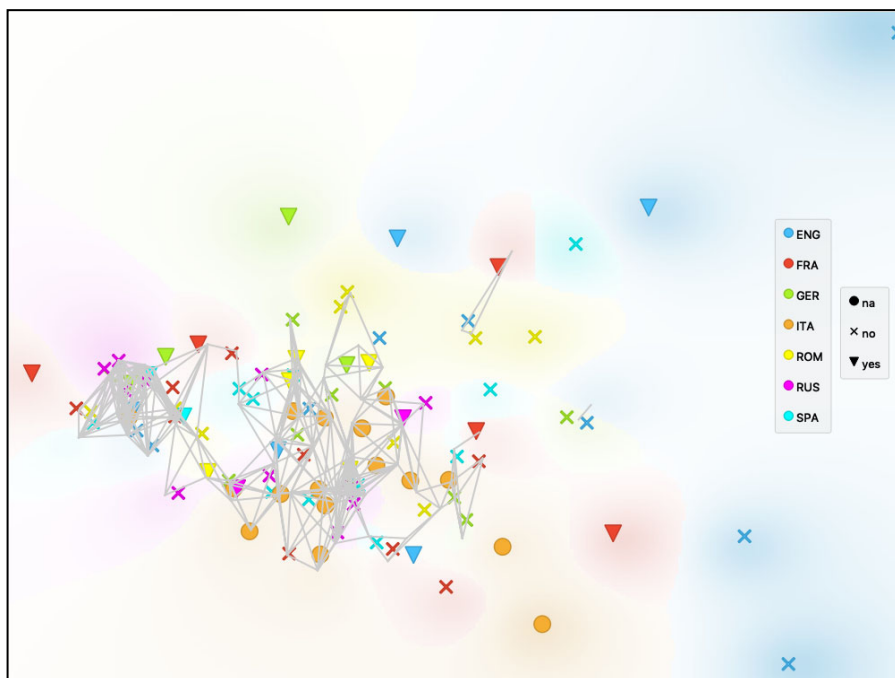
Lastly, Figure 6.8 illustrates the age at which Italian language acquisition began. Within the cluster shared by the Italian speakers, non-native speakers with a range of onset ages could be found. However, outside of this cluster, there were notably more speakers who began learning Italian in adulthood.

Figure 6.8: MDS map showing the age of onset
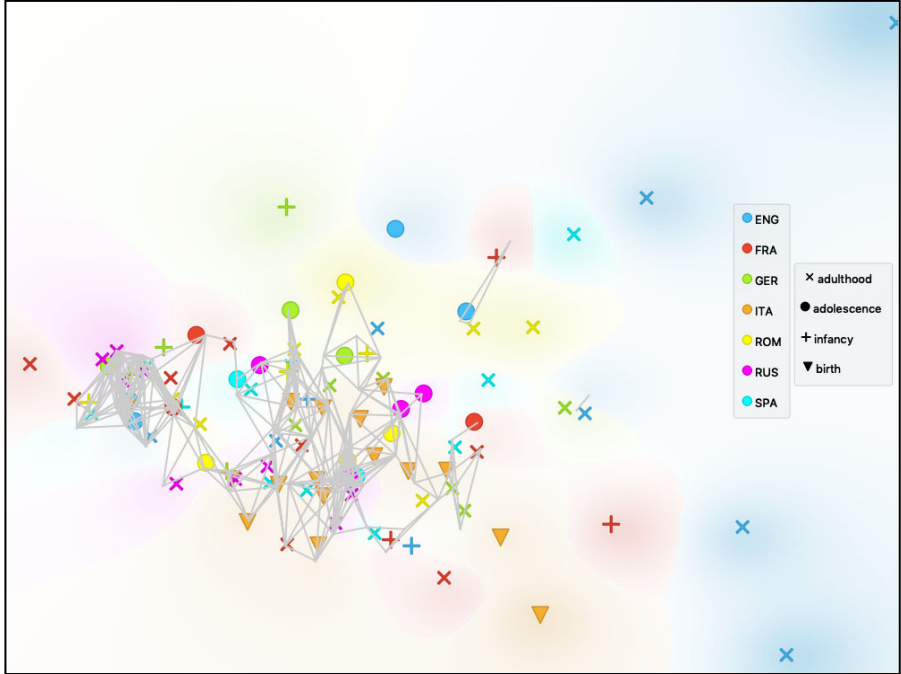
6.6 Discussion

The experiments described in Chapters 5 and 6 provided evidence to validate the hypotheses proposed at the outset of this monograph. In particular, the results of the analyses demonstrated that foreign-accented speech is indeed influenced by a multitude of factors. Moreover, both native speakers and automated classification systems have shown the ability to differentiate between native and non-native Italian speech, albeit with challenges in accurately identifying the exact origin of a speaker.

This accent perception experiment has provided valuable insights into the intricacies of evaluating foreign accents. One interesting finding of the first task is that a high level of familiarity with specific accents or

languages does not necessarily translate into accurate recognition rates. Despite claiming familiarity with the English accent and proficiency in the English language, Italian listeners exhibited low accuracy in identifying English speakers. Surprisingly, correct identification occurred in less than one third of cases. At the same time, individuals with limited exposure to the Romanian language or Romanian-accented Italian exhibited lower recognition rates, reflecting indeed the challenges posed by unfamiliar accents.

An interesting finding emerging from the third task of the accent perception experiment was the tendency of native Italian speakers to generally focus on segmental features when describing the salient cues of foreign-accented Italian. This observation was particularly evident for German and Spanish accents, which were also identified as the most recognisable accents (see Chapter 5). Notably, many of these cues are shared among different accents, while some are specific only to certain accents. Some of the segmental features that were observed to be common across all six accents include the failure to produce consonant gemination and non-native production of the alveolar trill.

Suprasegmental features, on the other hand, were reported less frequently and often with imprecise descriptions, suggesting that listeners may be less familiar with these cues. While it may be tempting to conclude that non-native segmental cues are more salient than suprasegmental cues, it is crucial to take into account the listeners' limited experience with the topic of prosody and the inherent challenges associated with reporting prosodic features. In fact, by testing the discriminative strength of prosodic features on foreign-accented Italian in §6.4 some promising results were obtained.

The results of the automatic classification experiments of accents varied significantly depending on the techniques employed. The fact that a simple acoustic GMM/SVM system showed discrete, yet not quite human, performance was not surprising. The performance of the automated systems using prosodic features varied, with the SVM system showing the lowest performance (0.36). However, a change in the classification method while retaining the same prosodic features resulted in a significant improvement in performance. Particularly, the prosodic Neural Network system achieved an overall classification score of 0.66, surpassing both the GMM/SVM system (0.42) and human raters (0.47) in terms of accuracy. The rather poor performance on the Romanian accent (i.e. 8 out of 14 speakers correctly identified) was predictable, considering that, from a prosodic point of view, Italian and Romanian are closely related, and in fact, in one fifth of the cases, Romanian speakers were mistaken for Italian; additionally, this matches somehow the trend for the overall human performance, even if in that case, the Italian listeners reached an even poorer classification score for the Ro-

manian accent (0.22). The English, the German, and the French accents were best classified by the prosodic Neural Network method which yielded classification rates above 0.70, compared to the scores achieved by Italian listeners (around 0.30). Finally, the Spanish and the Italian accents, which were easily identified by humans (0.78 and 0.95, respectively), were also well detected, even if the classification scores were slightly lower for the automatic systems (respectively 0.64 and 0.71 speakers correctly identified).

Furthermore, the unsupervised clustering methods revealed that speakers were generally placed into clusters that did not necessarily trace back to their L1s. In fact, sociolinguistic information and idiosyncratic prosodic characteristics contributed to the creation of 2-4 large clusters. Another interesting trend that emerged from the experiments with the unsupervised methods was that in both the hierarchical and in the k-Means/MDS clustering, most Romanian speakers and some Spanish speakers were found in the vicinity of the Italian speakers. English, French, and Russian speakers that had an advanced level in Italian – either C1 or C2 – and whose sociolinguistic information suggested an adequate linguistic input for Italian (i.e. pronunciation training; combined learning methods; long stay in Italy, etc.) were also in the same clusters or under the same branches with the Italian speakers. Finally, to determine the suprasegmental cues that played a significant role in explaining the variation among the seven accent groups, two scoring methods were employed: the chi-square statistic and ReliefF. From these data, it seems that the most informative features were energy, fundamental frequency (f0), word duration, and the duration of filled pauses. The inclusion of these features offered valuable insights into their effectiveness in characterising foreign accents. However, replicating the automated accent classification experiment only with these new set of features did not lead to improved accuracy, possibly due to overfitting.

Despite the relatively simple nature of the computational methods proposed in this study for exploring, classifying, and characterising foreign-accented Italian, the results obtained were promising and thought-provoking. These findings have the potential to pave the way for more sophisticated analyses in the future, while also building upon existing research conducted in other languages (De Marco and Cox, 2013; Verma and Das, 2015; Nejjari et al., 2020; Gosselin et al., 2021; etc.).

Conclusions

This monograph investigated the phenomenon of foreign-accented Italian, by examining in detail a corpus of Russian, English, German, French, Romanian, and Spanish accents. The main aim of the study revolved around exploring the various factors that influence the perception of foreign accent, including elements associated with the speaker, the listener, and the style. In particular, the study hypothesised that certain non-native accents would exhibit greater recognisability than others, even when other variables remained constant. This distinction was hypothesised to arise from highly salient segmental and suprasegmental features of foreign-accentedness. These features were assumed to have a relevant role in facilitating the identification of non-native Italian speech and distinguishing between various foreign accents, both for automated classification systems and native listeners. To test these hypotheses, an empirical investigation was undertaken, deploying a mixed-method approach. The study combined quantitative and qualitative analyses of accent perception experiments, along with the use of automated accent classification techniques. This chapter will provide the conclusions derived from this research, starting with the theoretical chapters and then summarising the findings obtained from the experimental study.

Chapter 1 explored the differences between the acquisition of native and non-native speech, an area of substantial interest within the linguistic community. Extensive research efforts in SLA have been dedicated to characterising, analysing, and elucidating the observable phenomena in L2 speech perception and production. In this context, Chapter 1 provided a comprehensive overview of prominent models that have been widely referenced to shed light on the perception of non-native speech and the acquisition of L2 phonology. These theoretical frameworks were classified by differentiating between models that interpreted L2 speech considering L1 influence and those built around universal principles. Considering the strengths and limitations of these models, it was shown that certain models, such as Flege's (1995) SLM, could offer a suitable theoretical framework for understanding

foreign-accented Italian, despite their primary focus and application being the acoustic dimension of non-native speech.

The scholars' interest in foreign-accented speech has been testified by the noticeable amount of research on this topic. The first part of Chapter 2 examined the multifaceted aspects of foreign accent, drawing upon existing literature to offer a definition of this phenomenon. The focus was directed towards the perception of non-native speech by native speakers of the target language. The chapter further explored the intricate methodological challenges associated with analysing non-native speech, outlining the factors that previous studies identified as being correlated with perceived foreign accent (e.g. age of L2 onset, input, gender, motivation, attitude, etc.). Additionally, the state-of-art survey on the perception of foreign-accented Italian revealed that the topics and the accents investigated in this work were not fully covered by previous research.

While this study did not directly investigate the social dynamics associated with non-native speech, Chapter 2 provided a glimpse into how foreign accent is reflected in the Italian media, employing a corpus-based approach. Keeping in mind that attitudes towards foreign accents are largely subjective, since they are shaped by individual beliefs as well as socio-cultural, socio-demographic, and linguistic backgrounds, Chapter 2 discussed how accent bias may be a form of linguistic discrimination. The presence of accent bias in our society may be linked to uneven treatment of people based on their speech productions. The discussion centred on the detrimental impact of accent bias on the personal and professional opportunities of L2 speakers, as it creates barriers in various contexts such as the workplace and social interactions. The brief corpus analysis pointed out that certain foreign accents may be stigmatised or stereotyped. The discussion underscored a circular relationship, wherein cultural prejudice or negative attitudes towards people of certain origins can influence the listener's perception of foreign accents, and vice versa. As a result, foreign accent bias has the potential to serve as a catalyst for the (re)activation of prejudice or hatred.

Cultural prejudice and/or historical circumstances might have influenced the choice of voicing with foreign accents the villains in films or animations. That was the case of some villains that in US films speak with Russian or German accents. The qualitative analysis of over 40 films and cartoons revealed that, in most cases, and especially in the Disney productions, standard or, very rarely, regional varieties of Italian were used to voice villains or other originally foreign-accented characters. This inapplicability of foreign accents in the Italian dubbed versions of these films was attributed to cultural considerations, making foreign accents unsuitable for the Italian audience. Additionally, taking into account that until the 1950's – the dec-

ade corresponding to the introduction of the television in the Italian homes – Italy had displayed a pronounced diatopic variation, the choice of voicing film characters with a 'standard' accent would have provided the target public – often children and young people – with a proper pronunciation model.

Starting from authentic testimonials delivered by L2 speakers of Italian, collected from Italian newspapers and forums, an exploration was conducted into the potential impact of foreign accents on the perception of non-native performance by native Italian speakers. It was shown that speaking with a foreign accent may have implications at personal, social, communicative, and professional levels. These implications encompass concerns about not mastering the language, leading to hesitancy in interacting with others, incomplete integration to the community of native speakers, as well as a more challenging road to better job opportunities and university degree programmes.

The chapter also presented a review of studies that investigated the effect of pronunciation training on foreign accented Italian. The reviewed works indicated that teaching pronunciation is generally useful in raising the learners' awareness of segmental and suprasegmental features of L2. However, due to the multitude of methodological differences between the studies reviewed, it is still particularly difficult to quantify objectively how beneficial really is receiving pronunciation training.

Chapter 3 emphasised the significant transformation, over the last decades, in the way we communicate and interact with the world. The use of NLU and NLG has become prevalent in Artificial Intelligence technologies, facilitating human-machine interactions. The chapter discussed about the need to prevent 'electronic imperialism' and emphasised the importance for voice assistants and similar applications to be able to handle diverse speech inputs, including those marked by background noise, varying ages, pathological conditions, and accents. These tools should be engineered well enough to decipher the message and fulfil the speaker's requests in all circumstances. As far as foreign accent is concerned, different techniques, such as acoustic model adaptation and pronunciation adaptation have been proposed to overcome this issue. The most relevant types of accent identification systems – generally classifiable into phonotactic and spectral approaches – were presented in Chapter 3. Within this chapter, most studies proposed highly engineered solutions, either for reducing the effect of foreign accents in ASR or for classifying foreign-accented speech. However, it is worth mentioning that these studies tend to offer limited linguistically meaningful insights, primarily due to the highly abstract nature of the features and techniques employed to model the proposed systems. Another

section of Chapter 3 was dedicated to corpora of non-native speech, highlighting the scarcity of such resources for the Italian language.

In order to fill this gap, Chapter 4 presented the non-native speech database built specifically for this work. The corpus consists of over 8 hours of read and spontaneous speech uttered by 122 young adult speakers whose L1s are either Russian, English, German, French, Romanian, Spanish, and Italian. Considering the scarcity of freely available corpora of spoken L2 Italian, one contribution of this work is that of collecting, compiling and storing these data in a repository. The application can extrapolate and classify the audio files from the corpus, according to specific search criteria.

Chapter 5 marks the beginning of the analyses on accent perception, featuring a large scale three-task experiment aimed at measuring foreign accent through the ears of 288 native speakers of Italian, an approach that is common in perceptual dialectology. The novelties of this study lie in the varieties of the accents considered, the number of speakers and listeners enrolled, as well as in the application of statistical analyses used to explain the identification accuracy and the degree of perceived foreign accent. Interpreting the results allowed to provide a detailed data-driven representation of Russian-, English-, German-, French-, Romanian-, and Spanish-accented Italian, as compared to the native varieties of Italian.

The Italian listeners achieved an overall identification score of 0.47 in the task of accent identification, surpassing the chance expectation of 0.14. They understood when the voice they heard was not Italian and belonged to a non-native speaker (0.89). Apart from the expected proficiency in recognising native Italian varieties (0.95), the listeners exhibited limited accuracy in identifying specific foreign accents. There was only a notable exception, namely the Spanish-accented Italian, where Italian respondents displayed a recognition score of 0.78. Their performance was comparatively poor for accents such as Russian, German, English, French, and Romanian. The results of the regression models suggested that certain variables, namely the listener's education level, region of origin, and experience with linguistics classes, had significant effects on their ability to recognise accents. It is, however, important to note that these variables accounted for only a portion of the observed variation in the data. Interestingly, being familiar with a specific accent and speaking the corresponding language did not necessarily ensure high accent identification rates. On the speaker's side, her/his L1 and proficiency level in Italian had a significant role in predicting the accent identification scores, although these factors alone were not sufficient to account for all the variation observed in the model. It is worth noting that certain accents (e.g. Spanish) were more easily recognisable than others.

The second part of Chapter 5 examined how several variables related to listeners, speakers, and style influenced the accentedness rating scores. First of all, an inter-rater reliability test was performed on the listeners' judgements, yielding a positive outcome, suggesting that there was often agreement between the listeners' ratings. All other listener and speaker variables being comparable, the accentedness rating revealed that some accents were perceived as being more marked than others. Listeners agreed that German, English, and Spanish accents were perceived as being the strongest, followed by French, Russian, and Romanian accents. Additionally, spontaneous speech was rated as less accented than read speech. This was in part explained by the fact that read speech could mirror reading errors arisen due to conflicting orthographic norms between the speakers' L1 and Italian. Also, some L2 speakers might have had an inadequate or insufficient exposure to written Italian (especially exchange students that had lived in Italy less than 12 months).

As far as the global accentedness rating is concerned, the descriptive statistics revealed that the listener-dependent variables slightly influencing the distribution of accentedness scores were the listener's education and her/his attendance in linguistics classes. Nevertheless, the contribution of these factors in explaining the variation was not substantial. There was only a slight inclination for educated listeners to evaluate the speakers' degree of accentedness more critically. Conversely, listeners with a background in linguistics tended to be more tolerant, although this variable did not consistently result in statistically significant differences. On the speakers' side, based on the regression model, the degree of foreign-accented Italian appeared to be predicted especially by the following factors: the speaker's L1 (i.e. some accents, such as German or Spanish, were judges as more marked than others), the presence or absence of specific pronunciation training during the learning path (i.e. having received pronunciation training during Italian language classes generally led to a more native-like accent), and the style (i.e. spontaneous speech was perceived as less accented than read speech).

Throughout this monograph, the discussion touched upon the contribution of different segmental and suprasegmental features to the perception of foreign-accented Italian. While previous studies have indeed provided descriptions of non-native Italian speech, the novelty of this work stems from the adoption of an empirical multi-method approach to the subject matter. Specifically, in Chapter 6, authentic speech samples produced by a wide variety of speakers were described in terms of their foreign-accent traits by a large pool of listeners. Although the cues that emerged from the listeners' comments may not have the same level of objectivity as those obtained

through instrumental analyses, they still hold value in providing an initial description of accent-specific feature patterns. An interesting finding emerged when examining the comments of native Italian speakers regarding the salient features of foreign-accentedness. It was observed that segmental features were mentioned more frequently than suprasegmental features, particularly in the case of the German and Spanish accents. The salience of segmental features mentioned by native Italian speakers when describing foreign-accentedness does not imply that these cues inherently play a more relevant role in accent perception. As a matter of fact, this finding could have been influenced by the listeners' limited experience in reporting and describing prosodic features. More generally, it is evident that some of these segmental features are accent-dependent (e.g. the Spanish speakers' deviant production of the voiced labiodental fricative; the English speakers' lengthening of the open-mid back rounded vowel; etc.), while others are common to all or most of the six accents (e.g. improper use of consonant gemination). In most cases these phenomena regard non-native contrasts or sounds that are similar but not identical to L1 sounds, therefore the failure to produce them properly could be explained in terms of an equivalence classification as proposed by Flege (1995).

As concerns prosody, the listeners' comments were rather similar among the six foreign accents, but in case of some accents (e.g. English) they were more frequent. Italian listeners mentioned frequently word stress misplacement and non-native rhythmic patterns. Regarding stress, some non-native speakers overgeneralised the fact that in Italian the stress falls frequently on the penultimate syllable, and therefore they tended to apply this rule for various other instances, in oxytone and proparoxytone words. An interesting finding emerged from the analysis of the features reported for the Romanian speakers. Due to a comparatively lower number of segmental cues mentioned for this accent, Italian listeners placed greater emphasis on the prosody (e.g. rhythm and intonation).

Even if, undeniably, L1 transfer explained most of the phenomena described in Chapter 6, in other cases non-native productions depended on overgeneralizations, as for instance the unnecessary use of consonant gemination or the tendency of using the paraoxytone – the most common type of stress for Italian – even for oxytone- or proparoxytone-stressed words. Also, the speaker's inexperience with a reading task in Italian or anxiety might have led to non-native prosody.

Given the relatively limited number of empirical studies examining specific segmental and suprasegmental features of foreign-accented Italian, Chapter 6 employed machine learning classification techniques to investigate the role of temporal and spectral features (derived from MFCC and

energy analysis) and prosodic features in the automatic classification of foreign-accented Italian. The first experiment was conducted on a GMM-UBM-SVM system based on spectral features. It achieved an overall recognition rate of 0.42 (chance expectation: 0.14), which rendered it quite comparable with the performance achieved by native speakers of Italian (0.47) and reflected similar (mis)-classification trends. However, a good performance was observed in recognising the Russian accent, where native speakers did not perform as effectively.

Considering that the Italian listeners that participated in the accent perception experiment found it challenging to provide accurate judgements on the prosody of non-native speakers, a machine learning accent classification system based on suprasegmental features was developed. Unlike the GMM-SVM system that was based on the rather opaque MFCC-derived features, a system built on prosodic features appeared more linguistically meaningful. Therefore, a set of 961 prosodic features were extracted, encompassing f0, energy, pauses, duration and rhythmic measurements. Various classification methods were tested, but the Neural Network approach emerged as the most successful, achieving an overall classification accuracy of 0.66 (macro-averaged for classes). This performance significantly surpassed that of native Italian speakers. Among the various accents, the French accent was classified with the highest accuracy, followed by English, Russian, German, Italian, Spanish, and Romanian accents, respectively, with more than half speakers correctly identified as native speakers of their respective L1s.

Next, the unsupervised clustering methods presented in Chapter 6 revealed that, based on their prosodic characteristics, speakers were generally placed into 2-4 large clusters that did not necessarily reflect their L1s. In both the hierarchical and in the k-Means/MDS clustering, most Romanian speakers and some Spanish speakers were found in the vicinity of the Italian speakers. English, French, and Russian speakers that had an advanced level in Italian – either C1 or C2 – and whose background suggested an adequate linguistic input for Italian were also found in the same clusters or under the same branches with the Italian speakers. This finding holds particular interest in the context of teaching L2 Italian and merits further investigation.

Chapter 6 also measured which suprasegmental features were the most informative for the accent classification task. The best ranked features according to the chi-square and ReliefF scoring methods were in this order: energy, f0, word duration, and the duration of filled pauses. The rankings suggested that energy-derived features played a particularly informative role in automatic classification of foreign-accented Italian. Features deriv-

ing from the normalised f0 as well as durational features explained some of the variation among the accents considered. This order of significance of the features is intriguing considering the tendency of Italian listeners to report segmental features more frequently. The correlation between segmental length and voice with energy adds to this observation. These results serve as a motivation to further investigate segmental and suprasegmental features, also by means of other instrumental techniques, so as to improve the description and comprehension of the foreign-accented Italian.

All in all, this study proposed an empirical multi-method approach for the analysis of non-native Italian speech, focusing on six foreign accents. The findings of this work display promising results that could lay the foundations for more complex analyses, including fine-grained acoustic measurements of specific features of non-native speech. As a matter of fact, one possible future direction for this research could be to expand the sample of speakers, encompassing a wider range of ages and diverse socio-cultural and socio-demographic backgrounds. This would allow to uncover other patterns of non-native speech and to test the contribution of sociopsychological factors, such as attitude, motivation, and personality on the pronunciation skills in L2 Italian.

Appendices

Appendix A: Informed consent

Consenso informato

<u>**Titolo del progetto**</u>:
A non-native Speech Database for Italian (CorAIt)
Corpus audio di italiano L2 (CorAIt)
Responsabile del progetto: Claudia Roberta Combei

Io sottoscritto/a …... dichiaro:
Di aver letto il foglio informativo dei/delle partecipanti (oppure che questo è stato letto e tradotto in lingua inglese per me) e di aver ricevuto l'opportunità di discutere con la responsabile del progetto delle informazioni in esso contenute e del mio coinvolgimento nel progetto.
Che le questioni relative alle procedure richieste dal progetto e al tempo impiegato per il suo svolgimento sono state chiarite, e che qualunque domanda io abbia avuto riguardo al progetto sia stata trattata in maniera soddisfacente.
Di acconsentire a completare il questionario e l'intervista di questo studio e a essere registrato/a via audio con strumenti elettronici, mentre produco suoni, parole, narrazioni, e di autorizzare la raccolta e l'utilizzo di tutto il materiale per finalità di ricerca.
Di essere consapevole che il mio coinvolgimento è di tipo confidenziale e che i dati raccolti durante quest'incontro potranno essere pubblicati tramite diversi mezzi, ma che nessuna informazione su di me sarà utilizzata in modo da rivelare la mia identità.
Di essere consapevole di potermi ritirare dal progetto in qualsiasi momento, senza che questo influisca in alcun modo sul mio rapporto con la responsabile, ora o in futuro.
Nome e cognome:
Indirizzo e-mail:
Luogo e data:
Firma:

Appendix B: Information sheet

Informazioni per i/le partecipanti

<u>**Titolo del progetto**</u>:
A non-native Speech Database for Italian (CorAIt)
Corpus audio di italiano L2 (CorAIt)
Responsabile del progetto: Claudia Roberta Combei

Quale è lo scopo del progetto?
Lo scopo di questo progetto è quello di raccogliere dati audio e informazioni di tipo sociolinguistico relativi all'accento straniero nella lingua italiana, registrando e analizzando parlanti madrelingua e non madrelingua.

Che cosa implica?
L'indagine prevede un incontro con Claudia Roberta Combei per ottenere informazioni sociolinguistiche e registrazioni di parlato letto e spontaneo.

Quanto tempo durerà l'esperimento?
La sessione durerà circa 60 minuti.

Che beneficio ne traggo?
Lo studio non Le darà alcun beneficio diretto, ma avrà l'opportunità di contribuire con informazioni utili alla ricerca sull'accento straniero nella lingua italiana.

Lo studio comporterà qualche danno o disagio alla mia persona?
Lo studio non Le procurerà alcun danno o disagio.

Qualcun altro verrà a conoscenza degli esiti della ricerca? Come saranno divulgati i risultati?
Solo il/la responsabile del progetto avrà accesso alle informazioni personali dei partecipanti. I dati saranno trattati in modalità confidenziale e saranno immagazzinati su supporti fisici e su servizi cloud. I risultati dello studio, il materiale audio registrato e le informazioni sociolinguistiche anonimizzate raccolte potranno essere divulgati al pubblico tramite corpora, siti web, presentazioni professionali in conferenze e pubblicazioni in riviste scientifiche e libri.

Posso ritirarmi dall'indagine?
La partecipazione è interamente volontaria. Lei non è obbligato/a a partecipare e se sceglie di aderire può ritirarsi in qualsiasi momento senza fornire alcuna giustificazione e senza che ci siano conseguenze.

Posso parlare ad altre persone di questa ricerca?
Sì. Può parlarne con altre persone fornendo loro il contatto della responsabile dell'indagine. Esse potranno contattarla per discutere della loro eventuale partecipazione al progetto di ricerca e per ricevere il foglio informativo.

Come posso chiedere ulteriori informazioni?
Per ulteriori informazioni può contattare Claudia Roberta Combei
Qualsiasi questione relativa alla ricerca sarà trattata in maniera confidenziale e investigata a fondo e Lei sarà informato/a degli esiti. Se accetta di partecipare a questo studio, Le verrà richiesto di firmare il Consenso informato.

Appendix C: Sociolinguistic survey

Questionario sociolinguistico

<u>**Titolo del progetto**</u>:
A non-native Speech Database for Italian (CorAIt)
Corpus audio di italiano L2 (CorAIt)
Responsabile del progetto: Claudia Roberta Combei

Data: Partecipante:

1. Età:

2. Che mano usa per scrivere? (Destra/Sinistra/Entrambe):

3. Genere (Maschio/Femmina/Altro/Preferisco non rispondere):

4. Ha avuto recentemente problemi di udito? (Sì/No):
Se sì, descriva il disturbo.

5. Ha dei problemi visivi che non sono corretti con occhiali o lenti a contatto? (Sì/No):
Se sì, descriva il disturbo.

6. Lei o i suoi famigliari più stretti avete avuto/avete problemi particolari nello sviluppo linguistico (es: *dimenticanza del linguaggio imparato, serie difficoltà ad imparare parole nuove, o a ricordare il nome degli oggetti*)? (Sì/No):
Se sì, chi ha/aveva il problema e di che natura era? A che età? È stato necessario l'intervento di un logopedista?

7. Lei o i suoi famigliari più stretti avete avuto/avete problemi con il linguaggio parlato? *(es: balbuzie, "lisca", ecc.)*? (Sì/No):
Se sì, chi? Che tipo di problema? A che età? È stato necessario l'intervento di un logopedista?

8. Lei o i suoi famigliari più stretti avete avuto/avete disturbi specifici dell'apprendimento (*es: dislessia, disgrafia, discalculia, ecc.)*? (Sì/No):
Se sì chi? Che tipo di problema? A che età? È stato necessario l'intervento di un educatore alla lettura?

9. Indichi la sua lingua madre (in caso di bilinguismo inserire entrambe le lingue):

10. Venivano parlate altre lingue, in casa, oltre la sua madrelingua mentre cresceva?
Se sì, elenchi ogni lingua

11. Quanti anni aveva quando ha iniziato a sentire parlare l'italiano *(es: dalla nascita; 2 anni; 15 anni; ecc.)*?

12. Elenchi tutte le lingue (compreso l'italiano) che ha appreso al di fuori della famiglia *(es: studiate a scuola, studiate in scuole di lingue nel suo paese o all'estero, apprese spontaneamente durante permanenze all'estero, ecc.)*, rispondendo alle seguenti domande:
- quanti anni aveva quando ha iniziato a studiare la lingua?
- come l'ha studiata?
- se imparata tramite insegnamento guidato, l'insegnante dedicava del tempo per insegnare la pronuncia?
- quanto bene parla, comprende, legge, e scrive in questa lingua?

Lingua e età di inizio dello studio/ dell'apprendimento della lingua	Informazioni sul modo in cui ha imparato/ appreso la lingua (v. sopra)	Indichi il suo livello di competenze nella lingua inserendo un numero da 1 (*molto basso*) a 5 (*molto alto*)				Presenza di lezioni dedicate alla pronuncia (Sì/No)
		Parlato	Comprensione	Lettura	Scrittura	

13. Indichi i posti in cui ha vissuto, menzionando per ognuno di essi il periodo trascorso lì.
Città, Regione, Stato, per quanto tempo ha vissuto lì?
Sono nato/a a:
Ho vissuto a:

14. Dove sono cresciuti Sua madre e Suo padre?

Madre (Città, Regione, Stato):
Padre (Città, Regione, Stato):

15. Qual è il suo ultimo titolo di studio conseguito (Diploma scuola secondaria di 1º grado, Diploma di istruzione professionale, Diploma di istruzione tecnica, Diploma di istruzione liceale, Laurea triennale, Laurea specialistica o magistrale, Dottorato di ricerca):

16. La sua occupazione (studente, impiegato, insegnante, libero professionista, imprenditore, disoccupato, pensionato):

17. L'occupazione di sua madre (studente, impiegato, insegnante, libero professionista, imprenditore, disoccupato, pensionato):

18. L'occupazione di suo padre (studente, impiegato, insegnante, libero professionista, imprenditore, disoccupato, pensionato):

GRAZIE PER LA SUA DISPONIBILITÀ!

Appendix D: Recording session

<u>Titolo del progetto</u>:
A non-native Speech Database for Italian (CorAIt)
Corpus audio di italiano L2 (CorAIt)
Responsabile del progetto: Claudia Roberta Combei

<u>ESPERIMENTO DI REGISTRAZIONE</u>

L'esperimento di oggi ha lo scopo di raccogliere materiale audio utile per studiare l'accento straniero nella lingua italiana. I dati raccolti saranno utilizzati esclusivamente per scopi di ricerca.

Come menzionato in precedenza, i dati raccolti durante l'esperimento di oggi saranno anonimizzati, immagazzinati su supporti fisici e servizi cloud e usati per finalità di ricerca. Nello specifico, le sue informazioni sociolinguistiche e le registrazioni audio della sua voce potrebbero essere inserite in forma del tutto anonima su corpora o su piattaforme online protette da password. Alla piattaforma si potrà accedere solo ed esclusivamente per scopi di ricerca previa autenticazione e approvazione da parte della responsabile del progetto (Claudia Roberta Combei).

L'esperimento di oggi è diviso in tre parti:
- Nella prima parte dovrà compilare una scheda sociolinguistica in cui le sarà chiesto di rispondere ad alcune domande riguardanti l'età, le competenze con le lingue straniere, ecc.
- Nella seconda parte parlerà liberamente di un argomento presentato sotto forma di domanda.
- Nella terza parte leggerà un brano tratto da un articolo di giornale. Le verrà chiesto di leggerlo due volte, facendo una pausa fra le due letture.

Girando il foglio troverà le istruzioni dettagliate relative alla seconda e alla terza parte dell'esperimento.

Prima di iniziare la registrazione La invito a compilare la scheda sociolinguistica che può trovare sulla scrivania.

Grazie per la Sua disponibilità!

Parte 1
Risponda alla seguente domanda parlando nel modo più spontaneo possibile:
1. Si presenti e racconti dove e come ha trascorso le ultime vacanze d'inverno.

Parte 2
Legga il brano nel modo più naturale possibile. È necessario leggere il brano due volte, facendo una breve pausa tra le due letture.

Don Geretti è un grande affabulatore. Siccome è un prete, è più giusto dire che è un bravissimo predicatore. Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi. Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. «La fede, sostiene, consiste in uno sguardo di meraviglia. Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà». Parroco di Illegio, minuscolo paese nel cuore della Carnia, don Geretti ha fondato il Comitato di San Floriano, con cui da una decina di anni porta avanti questa missione di coniugare arte e fede. La sua idea di creare un percorso valido dal punto di vista scientifico e al tempo stesso significativo dal punto di vista cristiano ha convinto gli studiosi, i direttori di musei che non esitano a prestare capolavori e i visitatori che accorrono numerosi ad ammirare le rassegne allestite nella canonica del villaggio in mezzo alle montagne. Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo. Ma la scelta del luogo non deriva dal fatto che rappresenta una vetrina internazionale. Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.

Colonnelli Lauretta
Pagine 34-35
(06 febbraio 2013) - Corriere della Sera
Available at:
http://cinquantamila.corriere.it/storyTellerArticolo.php?storyId=0000002228555
(accessed on the 3rd of February 2023).

Appendix E: Stimuli for the accent perception experiment

Accent identification task

L1	*Stimulus*
Russian	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi.
Russian	Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. La fede, sostiene, consiste in uno sguardo di meraviglia.
Russian	Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
Russian	Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà.
Russian	Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo.
Russian	Parroco di Illegio, minuscolo paese nel cuore della Carnia, don Geretti ha fondato il Comitato di San Floriano, con cui da una decina di anni porta avanti questa missione di coniugare arte e fede.
Russian	Don Geretti è un grande affabulatore. Siccome è un prete, è più giusto dire che è un bravissimo predicatore.
English	Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà.
English	Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. La fede, sostiene, consiste in uno sguardo di meraviglia.
English	Parroco di Illegio, minuscolo paese nel cuore della Carnia, don Geretti ha fondato il Comitato di San Floriano, con cui da una decina di anni porta avanti questa missione di coniugare arte e fede.
English	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi.
English	La sua idea di creare un percorso valido dal punto di vista scientifico e al tempo stesso significativo dal punto di vista cristiano ha convinto gli studiosi.
English	Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo.
English	Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. La fede, sostiene, consiste in uno sguardo di meraviglia.
German	Don Geretti è un grande affabulatore. Siccome è un prete, è più giusto dire che è un bravissimo predicatore.
German	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi.
German	Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
German	Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo.
German	Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. La fede, sostiene, consiste in uno sguardo di meraviglia.

German	La sua idea di creare un percorso valido dal punto di vista scientifico e al tempo stesso significativo dal punto di vista cristiano ha convinto gli studiosi.
German	Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà.
French	Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo.
French	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi.
French	Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. La fede, sostiene, consiste in uno sguardo di meraviglia.
French	Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
French	Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà.
French	Don Geretti è un grande affabulatore. Siccome è un prete, è più giusto dire che è un bravissimo predicatore.
French	Parroco di Illegio, minuscolo paese nel cuore della Carnia, don Geretti ha fondato il Comitato di San Floriano, con cui da una decina di anni porta avanti questa missione di coniugare arte e fede.
Romanian	Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo.
Romanian	Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
Romanian	La sua idea di creare un percorso valido dal punto di vista scientifico e al tempo stesso significativo dal punto di vista cristiano ha convinto gli studiosi.
Romanian	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi.
Romanian	Ma la scelta del luogo non deriva dal fatto che rappresenta una vetrina internazionale.
Romanian	Don Geretti è un grande affabulatore. Siccome è un prete, è più giusto dire che è un bravissimo predicatore.
Romanian	Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. La fede, sostiene, consiste in uno sguardo di meraviglia.
Spanish	Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo.
Spanish	Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
Spanish	La sua idea di creare un percorso valido dal punto di vista scientifico e al tempo stesso significativo dal punto di vista cristiano ha convinto gli studiosi.
Spanish	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi.
Spanish	Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà.
Spanish	Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. La fede, sostiene, consiste in uno sguardo di meraviglia.

Spanish	Parroco di Illegio, minuscolo paese nel cuore della Carnia, don Geretti ha fondato il Comitato di San Floriano, con cui da una decina di anni porta avanti questa missione di coniugare arte e fede.
Italian	Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. La fede, sostiene, consiste in uno sguardo di meraviglia.
Italian	Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
Italian	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi.
Italian	Ma la scelta del luogo non deriva dal fatto che rappresenta una vetrina internazionale.
Italian	Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo.
Italian	Don Geretti è un grande affabulatore. Siccome è un prete, è più giusto dire che è un bravissimo predicatore.
Italian	La sua idea di creare un percorso valido dal punto di vista scientifico e al tempo stesso significativo dal punto di vista cristiano ha convinto gli studiosi.

Accentedness rating task

L1	Type of speech	Stimulus
Russian	Read	Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo.
Russian	Read	Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà.
Russian	Read	Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
Russian	Spontaneous	Le mie vacanze invernali non erano tanto interessante, diciamo. Nella maggior parte del tempo ho studiato perché ho esami.
Russian	Spontaneous	Ho passato tre giorni a casa, abbiamo mangiato, abbiamo bevuto, abbiamo festeggiato la Pasqua, ho visto la mia famiglia.
Russian	Spontaneous	Ahm, comunque per me ferie d'inverno è sempre meglio stare a casa mia perché c'è molto più attività bella da fare.
English	Read	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi.
English	Read	Don Geretti è un grande affabulatore. Siccome è un prete, è più giusto dire che è un bravissimo predicatore.
English	Read	Parroco di Illegio, minuscolo paese nel cuore della Carnia, don Geretti ha fondato il Comitato di San Floriano, con cui da una decina di anni porta avanti questa missione di coniugare arte e fede.

English	Spontaneous	Pensavo che sarebbe più come Italia o come Spagna, molto mediterraneo, ma era una sorpresa per me perché era più influenza arabo, araba, dico.
English	Spontaneous	La mia ultima vacanza sono andata in Sicilia con la mia coinquilina, e abbiamo passato tipo quattro giorni, era un'esperienza molto bellissima.
English	Spontaneous	E poi mi pare sono anche andato dopo a visitare un mio amico sulla costa ovest, abbiamo fatto in macchina da Los Angeles fino a San Francisco.
German	Read	Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
German	Read	La sua idea di creare un percorso valido dal punto di vista scientifico e al tempo stesso significativo dal punto di vista cristiano ha convinto gli studiosi.
German	Read	Parroco di Illegio, minuscolo paese nel cuore della Carnia, don Geretti ha fondato il Comitato di San Floriano, con cui da una decina di anni porta avanti questa missione di coniugare arte e fede.
German	Spontaneous	Per questo non ho fatto quest'anno un viaggio di, d'inverno, però in febbraio durante le vacanze d'università, durante le vacanze, sono stato in Italia.
German	Spontaneous	Ci ho f... ci ho fatto un tirocinio per l'università all'ospedale, in pediatra, era molto interessante, e nel tempo libero ho scoperto un po' isola.
German	Spontaneous	E sono andata a trovarli perché mio papà non poteva proprio fare sport perché si era fatto male al gionocchio
French	Read	Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo.
French	Read	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi.
French	Read	Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. La fede, sostiene, consiste in uno sguardo di meraviglia.
French	Spontaneous	Hanno cucinato una torta deliciosa e ci hanno fatto regali, e sabato sono andati a goderci il sole e il caldo durante un picnic.
French	Spontaneous	E allora ho cominciato a scrivere mia tesina per la fine del master e fortunamente ho potuto scrivere molte cose.
French	Spontaneous	Quindi sono tornata a casa, sono stata un po' com i miei, ho visto i miei amici, ne ho incontrato altri, perché facendo serate su serate incontri gente, quindi ho incontrato amici che ancora vedo
Romanian	Read	Ma la scelta del luogo non deriva dal fatto che rappresenta una vetrina internazionale.
Romanian	Read	Parroco di Illegio, minuscolo paese nel cuore della Carnia, don Geretti ha fondato il Comitato di San Floriano, con cui da una decina di anni porta avanti questa missione di coniugare arte e fede.

Romanian	Read	Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
Romanian	Spontaneous	E qui sono stata con la famiglia, e insiem, cioè, siam, sono stata con la famiglia e insieme anche alla mia sorella che è venuta a trovarci e abbiamo passato insieme il resto della vacanza.
Romanian	Spontaneous	E poi affittiamo una casa lì e di solito rimaniamo pel Capodanno almeno tre quattro giorni con tutti i miei amici soprattutto quelli che adesso siamo tanti tanto tempo che non ci vediamo.
Romanian	Spontaneous	E poi per Capodanno sono tornata a Bologna e ho festeggiato con i miei amici, molto bello però un po' freschino.
Spanish	Read	Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo.
Spanish	Read	Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà.
Spanish	Read	Don Geretti è un grande affabulatore. Siccome è un prete, è più giusto dire che è un bravissimo predicatore.
Spanish	Spontaneous	E la mia nonna ha cucinato per tutti e dopo ehm, abbiamo fatto un ehm gioco che si chiama amici invisibile.
Spanish	Spontaneous	Questo estate sono arrivata all'Italia e a luglio sono stata a Porto Recanati, che è un posto vicino a Ancona e vicino a Recanati che dove è nato questo famoso poeta scrittore.
Spanish	Spontaneous	È stata un'esperienza bellissima perché abbiamo vabbè mi hanno fatto anche assaggiare un sacco di robe un sacco di cose, mi hanno fatto vedere il mare, i musei, la città, i piccoli paesi che sono intorno di questa città e così così via.
Italian	Read	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi.
Italian	Read	Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. La fede, sostiene, consiste in uno sguardo di meraviglia.
Italian	Read	Ma la scelta del luogo non deriva dal fatto che rappresenta una vetrina internazionale.
Italian	Spontaneous	E quindi siamo rimasti bloccati parecchio tempo, uhm, quasi una settimana e mezzo.
Italian	Spontaneous	E abbiamo visto le varie opere che ci sono in città e siamo stati con i nostri amici, è stata una bellissima vacanza.
Italian	Spontaneous	Abbiamo vabbè, mi sono divertito, mi sono riposato, ma anche un po' di dovere di lavoro da portare c'è sempre.

Salient features task

L1	Stimulus
Russian	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi. Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. La fede, sostiene, consiste in uno sguardo di meraviglia.
Russian	Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà. I direttori di musei che non esitano a prestare capolavori e i visitatori che accorrono numerosi ad ammirare le rassegne allestite nella canonica del villaggio in mezzo alle montagne.
Russian	Ma la scelta del luogo non deriva dal fatto che rappresenta una vetrina internazionale. Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
English	Don Geretti è un grande affabulatore. Siccome è un prete, è più giusto dire che è un bravissimo predicatore. Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi.
English	La fede, sostiene, consiste in uno sguardo di meraviglia. Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà.
English	Parroco di Illegio, minuscolo paese nel cuore della Carnia, don Geretti ha fondato il Comitato di San Floriano, con cui da una decina di anni porta avanti questa missione di coniugare arte e fede.
German	Don Geretti è un grande affabulatore. Siccome è un prete, è più giusto dire che è un bravissimo predicatore. Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi. Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza.
German	Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà. Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
German	Ma la scelta del luogo non deriva dal fatto che rappresenta una vetrina internazionale. Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo.
French	I direttori di musei che non esitano a prestare capolavori e i visitatori che accorrono numerosi ad ammirare le rassegne allestite nella canonica del villaggio in mezzo alle montagne. Passo dopo passo, il parroco è arrivato a Castel Sant'Angelo, il monumento romano più conosciuto al mondo insieme al Colosseo.
French	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi. Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza.
French	Ma la scelta del luogo non deriva dal fatto che rappresenta una vetrina internazionale. Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
Romanian	Ma la scelta del luogo non deriva dal fatto che rappresenta una vetrina internazionale. Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.

Romanian	Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza. La sua idea di creare un percorso valido dal punto di vista scientifico e al tempo stesso significativo dal punto di vista cristiano ha convinto gli studiosi, i direttori di musei che non esitano a prestare capolavori e i visitatori che accorrono numerosi ad ammirare le rassegne allestite nella canonica del villaggio in mezzo alle montagne.
Romanian	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi. Parroco di Illegio, minuscolo paese nel cuore della Carnia, don Geretti ha fondato il Comitato di San Floriano, con cui da una decina di anni porta avanti questa missione di coniugare arte e fede.
Spanish	I direttori di musei che non esitano a prestare capolavori e i visitatori che accorrono numerosi ad ammirare le rassegne allestite nella canonica del villaggio in mezzo alle montagne. Ma la scelta del luogo non deriva dal fatto che rappresenta una vetrina internazionale. Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
Spanish	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi. Parroco di Illegio, minuscolo paese nel cuore della Carnia, don Geretti ha fondato il Comitato di San Floriano, con cui da una decina di anni porta avanti questa missione di coniugare arte e fede.
Spanish	La fede, sostiene, consiste in uno sguardo di meraviglia. Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà.
Italian	Non cerca di condurre le persone alla fede con la minaccia delle fiamme dell'inferno, come in passato hanno fatto molti suoi celebri colleghi. Don Geretti vuole arrivare al cuore attraverso lo stupore della bellezza.
Italian	Ma la scelta del luogo non deriva dal fatto che rappresenta una vetrina internazionale. Dagli spalti del castello si vede, vicinissima, la basilica di San Pietro, il punto dove Pietro terminò il suo cammino terreno e quello, tormentatissimo, verso la fede.
Italian	Perciò le si accompagnano da sempre la filosofia, ovvero il pensiero scaturito dalla meraviglia per il fatto che la realtà c'è e l'arte, ovvero le forme scaturite dalla meraviglia per la bellezza della realtà.

Acknowledgements

I would like to express my gratitude to everyone who contributed to the realisation of this book.

Firstly, I would like to thank Dipartimento di Studi Umanistici (Università di Pavia) for funding the publication of my monograph through *Fondo Ricerca e Giovani (FRG) – Pubblicazioni Giovani Studiosi*.

Then, I am grateful to my PhD advisors, Giovanna Marotta and Fabio Tamburini. Their expertise was of immense value in shaping the content and direction of this work. I extend my appreciation to the reviewers and committee members who evaluated my PhD thesis, as well as the reviewers of this book, for their insightful feedback. I also wish to acknowledge all speakers and listeners that participated in the experiments.

Furthermore, I would like to thank the exceptional colleagues and dear friends I have had the pleasure of meeting along the way in Pisa, Erlangen, Nürnberg, Birmingham, Bologna, and Pavia. The knowledge they have shared and the fruitful discussions we have engaged in have been an endless source of inspiration and intellectual growth for me. In particular, I am grateful to my linguistics colleagues in Pavia who made me feel welcome from the moment I arrived and encouraged me to publish this book.

Finally, a special thanks goes to my family in Romania and Italy for their encouragement and understanding throughout this journey.

References

Adjemian, C. (1976). On the nature of interlanguage systems. *Language Learning*, 26, 297–320.

Ahad, M. T., Ahsan, M. M., Jahan, I., Nazim, R., Yazdan, M. M. S., Huebner, P., and Siddique, Z. (2021). Behavioral Pattern Analysis between Bilingual and Monolingual Listeners' Natural Speech Perception on Foreign–Accented English Language Using Different Machine Learning Approaches. *Technologies*, 9(3), 51, 1–11.

Ahamad, A., Anand, A., and Bhargava, P. (2020). AccentDB: A Database of Non–Native English Accents to Assist Neural Speech Recognition. In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)* (pp. 5351–5358). Marseille, 11–16 May 2020. European Language Resources Association (ELRA).

Algethami, G. (2017). The effects of explicit pronunciation instruction on the degree of perceived foreign accent in the speech of EFL learners. *Research in Language,* 15(3), 253–263.

Archibald, J. (1998). *Second language phonology*. Amsterdam: John Benjamins.

Alsteris, L. D. and Paliwal, K. K. (2007). Short–time phase spectrum in speech processing: A review and some experimental results. *Digital Signal Processing*, 17, 578–616.

Atagi, E. and Bent, T. (2016). Auditory free classification of native and nonnative speech by nonnative listeners. *Applied Psycholinguistics*, 37(2), 241–263.

Backman, N. 1979. Intonation Errors in Second–language Pronunciation of Eight Spanish–speaking Adults Learning English. *Interlanguage Studies Bulletin*, 4, 239–265.

Ballem K. D. and Plunkett K. (2005). Phonological specificity in children at 1; 2. *Journal of Child Language*, 32(1), 159–173.

Ballerini, M. and Rosati, R. (1990). *Laboratorio linguistico. Esercizi di fonetica per stranieri (con due cassette)*. Scandicci: Centro Editoriale Toscano.

Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G., and Mazzoleni, M. (2004). Introducing the La Repubblica Corpus: A Large, Annotated, TEI(XML)–compliant Corpus of Newspaper Italian. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)* (pp. 1771–1774). Lisbon, 26–28 May 2004. European Language Resources Association (ELRA).

Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009). The WaCky wide web: A collection of very large linguistically processed web–crawled corpora. *Language Resources and Evaluation*, 43(3), 209–226.

Bartelds, M., Richter, C., Liberman, M., and Wieling, M. (2020). A new acoustic–based pronunciation distance measure. *Frontiers in Artificial Intelligence*, 3(39), 1–20.

Bartelds, M., Vries, de, W., Richter, C., Liberman, M., and Wieling, M. (2021). Measuring foreign accent strength using an acoustic distance measure. In M. Tiede et al. (Eds.), *Proceedings of the 12th International Seminar on Speech Production* (pp. 17–20). New Haven: Haskins Press.

Batliner, A., Buckow, J., Niemann, H., Nöth, E., and Warnke, V. (2000). The Prosody Module. In W. Wahlster (Ed.), *Verbmobil: Foundations of Speech–to–Speech Translation*, 106–121. Berlin: Springer.

Bazzi, L., Brouwer, S., and Foucart, A. (2022). The impact of foreign accent on irony and its consequences on social interaction. *Journal of Multilingual and Multicultural Development*, article 2126487, 1–13.

Behravan, H. (2016). *Advances in Automatic Foreign Accent Recognition*. University of Eastern Finland. PhD Dissertation.

Behravan, H., Hautamäki, V., and Kinnunen, T. (2015). Factors affecting i–vector based foreign accent recognition: A case study in spoken Finnish. *Speech Communication*, 66, 118–129.

Behravan, H., Hautamäki, V., Siniscalchi, S. M., Kinnunen, T., and Lee, C. H. (2014). Introducing attribute features to foreign accent recognition. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)* (pp. 5332–5336). Florence, 4–9 May 2014.

Behravan, H., Hautamäki, V., Kinnunen T., Siniscalchi, S. M., and Lee, C. H. (2016). i–Vector modeling of speech attributes for automatic foreign accent recognition. *IEEE/ACM Trans. Audio, Speech and Lang*, 24(1), 29–41.

Bernini, G. (1988). Questioni di fonologia nell'italiano lingua seconda. In A. Giacalone Ramat (Ed.) *L'italiano tra le altre lingue. Strategie di acquisizion*e (pp. 77–90). Bologna: Il Mulino.

Bernini, G. (2015). La linguistica acquisizionale. In M. Daloiso (Ed.), *Scienze del linguaggio e educazione linguistica* (pp. 119–128). Torino: Loescher.

Bernini, G. (2018). La fonetica delle varietà iniziali di L2: regolarità e instabilità. In R. Bombi and F. Costantini (Eds.), *Percorsi linguistici e interlinguistici: Studi in onore di Vincenzo Orioles* (pp. 81–92). Udine: Forum.

Bertinetto, P. M. and Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association,* 35(2), 131–151.

Best, C. (1995). A direct realistic view of Cross–Language Speech Perception. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in Cross–Language Research* (pp. 171–204). Baltimore: York Press.

Best, C. and Strange, W. (1992). Effects of Phonological and Phonetic Factors on Cross–Language Perception of Approximants. *Haskins Laboratories Status Report on Speech Research*, SR–109/110, 89–108.

Best, C. and McRoberts, G. (2003). Infant Perception of Non–Native Consonant Contrasts that Adults Assimilate in Different Ways. *Language and Speech*, 46(2–3), 183–216.

Best, C. and Tyler, M. (2007). Nonnative and second–language speech perception. In O. S. Bohn and. M. J. Munro (Eds.), *Language Experience in Second language Speech Learning. In Honor of James Emil Flege* (pp. 13–45). Amsterdam: John Benjamins.

Best, C., Goldstein, L., Tyler, M., and Nam, H. (2009). Articulating the Perceptual Assimilation Model (PAM): Perceptual assimilation in relation to articulatory organs and their constriction gestures. *The Journal of the Acoustical Society of America*, 125(4), 2758–2758.

Biadsy, F., Hirschberg, J., and Ellis, D. P. W. (2011): Dialect and accent recognition using phonetic–segmentation supervectors. In *Interspeech 2011* (pp. 745–748). Florence, 27–31 August 2011. ISCA.

Bianchi, F. and Calamai, S. (2012). Voci italiane e straniere a confronto. Indagine sugli stereotipi associati agli accenti stranieri. *Cambio*, 2(4), 149–164.

Bickel, B. (2010), Absolute and statistical universals. In P. C. Hogan (Ed.) *The Cambridge Encyclopedia of the Language Sciences* (pp. 380–384), Cambridge: Cambridge University Press.

Bongaerts, T., Summeren, C., Planken, B., and Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, 19, 447–466.

Boula De Mareuil, P, Marotta, G., and Adda–Decker, M. (2004). Contribution of Prosody to the Perception of Spanish/Italian Accents In *Speech Prosody 2* (pp. 681–684). Nara, 23–26 March 2004.

Boula De Mareüil, P., Vieru, B., Woehrling, C., and Adda–Decker, M. (2008). Foreign and regional accents in French. Characterisation and identification. *TAL, 49*, 135–163.

Bridle, J. S. and Brown, M. D. (1974). *An Experimental Automatic Word–Recognition System*. Report No. 1003. Ruislip: Joint Speech Research Unit.

Brown, G. (2016). Automatic Accent Recognition Systems and the Effects of Data on Performance. In *Odyssey 2016* (pp. 94–100). Bilbao, 21–24 June 2016.

Brown, H. D. (1994). *Principles of Language Learning and Teaching*. 3rd ed. New Jersey: Prentice Hall Regents.

Broussard, S., Hickok, G., and Saberi. K. (2017). Robustness of speech intelligibility at moderate levels of spectral degradation. *PLoS ONE*, 12(7), article e0180734, 1–14.

Burgess, J. and Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System*, 28, 191–215.

Byrne, W., Knodt, E., Khudanpur, S., and Bernstein, J. (1998). Is automatic speech recognition ready for non–native speech? A data collection effort and initial experiments in modelling conversational Hispanic English. In *Speech Technology in Language Learning (STiLL)* (pp. 37–40), Marholmen, 25–27 May 1998.

Calabrò, L. (2010). *E tu… come pronunci? Eserciziario di fonetica italiana per stranieri. Con CD Audio.* Alessandria: Edizioni dell'Orso.

Calabrò, L. (2015). Il workshop di fonetica in italiano L2/LS. *Italiano LinguaDue*, 7(1), 40–49.

Calabrò, L. (2016). Il workshop di fonetica in italiano L2/LS: accento di parola e durata sillabica. *Italiano LinguaDue*, 8(1), 322–326.

Calabrò, L and Mairano, P. (2017). Insegnare la pronuncia a studenti sinofoni di italiano L2 con il Minimal Pair Finder. *Italiano LinguaDue*, 9(1), 83–97.

Calamai, S. (2015). Between linguistics and social psychology of language: the perception of non–native accents. *Studi e Saggi Linguistici*, 53(2), 289–308.

Calamai, S. and Ardolino, F. (2020). Italian with an accent: The case of "Chinese Italian" in Tuscan high schools. *Journal of Language and Social Psychology*, 39(1), 132–147.

Calamai, S., Nodari, R., and Galatà, V. (2020). "Fregati dall'accento!". Lo stereotipo etnico e linguistico nei contesti scolastici. *Italiano LinguaDue*, 12(1), 430–458.

Campbell, W., Sturim, D., and Reynolds, D. (2006). Support Vector Machines Using GMM Supervectors for Speaker Verification. *Signal Processing Letters IEEE*, 13, 308–311.

Canepari, L. (2007). *Pronunce straniere dell'italiano (ProSIt)*. München: LINCOM GmbH.

Carboni, G. and Sorianello, P. (2011). *Manuale professionale di dizione e pronuncia (con CD-ROM)*. Milano: Hoepli.

Carlisle, R. S. (1997). The Modification of Onsets in a Markedness Relationship: Testing the Interlanguage Structural Conformity Hypothesis. *Language Learning*, 47, 327–361.

Carlisle, R. S. (1998). The acquisition of onsets in a markedness relationship: a longitudinal study. *Studies in Second Language Acquisition*, 20, 245–60.

Carlisle, R. S. (1999). The Modification of Onsets in a Markedness Relationship: Testing the Interlanguage Structural Conformity Hypothesis. *Language Learning*, 49, 59–93.

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

Chakraborty, R., Schwarz, A., and Chakraborty, P. (2017). Perception of Nonnative Accent: A Cross–Sectional Perspective Pilot Survey. *International Journal of Society, Culture and Language*, 5(2), 26–36.

Chang, C. C. and Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM Trans. Intelligent Systems and Technology*, 2(3), 1–27.

Chao, I. H., Tsai, W. H., and Wang, H. M. (2009). Improving GMM–UBM speaker verification using discriminative feedback adaptation. *Computer Speech and Language*, 23(3), 376–388.

Chen, J., Best, C. T., and Antoniou, M. (2020). Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners. *Journal of Phonetics*, 83, article 101013.

Chen, J., Antoniou, M., and Best, C. T. (2023). Phonological and phonetic contributions to perception of non–native lexical tones by tone language listeners:

Effects of memory load and stimulus variability. *Journal of Phonetics*, 96, article 101199.

Chen, N. F. and Li, H. (2016). Computer–assisted pronunciation training: From pronunciation scoring towards spoken language learning. In *2016 Asia–Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)* (pp. 1–7). Jeju, 13–15 December 2016.

Chini, M. (2005). *Che cos'è la linguistica acquisizionale*. Roma: Carocci.

Chini, M. (Ed.). (2005). *Il parlato in (italiano) L2*. Milano: Franco Angeli.

Chițoran, I. (2002a). *The Phonology of Romanian: A Constraint–based Approach*. Berlin and New York: Mouton de Gruyter.

Chițoran, I. (2002b). A perception–production study of Romanian diphthongs and glide–vowel sequences. *Journal of the International Phonetic Association*, 32(2), 203–222.

Chițoran, I. (2002c). The phonology and morphology of Romanian diphthongization. *Probus*, 14(2), 205–246.

Chițoran, I. and Hualde, J. I. (2007). From Hiatus to Diphthong: The Evolution of Vowel Sequences in Romance. *Phonology* 24(1), 37–75.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290.

Colantoni, L., Steele, J., and Escudero, P. (2015). *Second language speech. Theory and practice*. Cambridge: Cambridge University Press.

Collins, B. and Mees, I. (2013). *Practical Phonetics and Phonology*. London and New York: Routledge.

Combei, C. R. (2017). CorAIt – A non–native speech database for Italian. In R. Basili et al. (Eds.), *Proceedings of the 4th Italian Conference on Computational Linguistics* (pp. 113–118). Torino: Accademia University Press.

Combei, C. R. and Marotta, G. (2019). Where Do Accents Come from? Factors Affecting the Degree of Foreign–accented Italian. In D. Piccardi et al. (Eds.), *Audio archives at the crossroads of speech sciences, digital humanities and digital heritage* (pp. 233–248). Milano: Officinaventuno.

Combei, C. R., Tordini, O., and Ardolino, F. (2020). L'acquisizione del sistema vocalico italiano: uno studio acustico su parlanti germanofoni e romenofoni. *Italiano LinguaDue*, 12(2), 175–188.

Conceição Silva, C. and Almeida Barbosa, P. (2017). The contribution of prosody to foreign accent: A study of Spanish as a foreign language. *Loquens*, 4(2), article e041.

Conklin, J. T., Kentner, A., Law, W. L., Lin, M., Wang, Y., and Dmitrieva, O. (2015). Audience design in non–native speech. *The Journal of the Acoustical Society of America*, 137(4_Supplement), article 2384.

Corder, S. (1967). The Significance of Learner's Errors. *International Review of Applied Linguistics in Language Teaching*, 5(4), 161–170.

Cosi, P., Cohen, M. M., and Massaro, D. W. (2002). Baldini: Baldi Speaks Italian! In *Proceedings of the 7th International Conference on Spoken Language Processing, ICSLP2002* (pp. 16–20). Denver, 16–20 September 2002.

Cosi, P., Fusaro, A., and Tisato, G. (2003). LUCIA: a New Italian Talking–Head Based on a Modified Cohen–Massaro's Labial Coarticulation Model. In *Eurospeech 2003* (pp. 2269–2272). Geneva, 1–4 September 2003.

Costamagna, L. (1990). *Cantare l'italiano – materiali per l'apprendimento dell'italiano attraverso le canzoni*. Perugia: Guerra.

Costamagna, L. (1996). *Pronunciare l'italiano – Manuale di pronuncia italiana per stranieri: livello intermedio e avanzato*. Perugia: Guerra.

Costamagna, L. (2002). *Pronunciare l'italiano. CD–ROM*. Perugia: Guerra Edizioni.

Costamagna, L. (2010). *Insegnare e imparare la fonetica – Testo per la Formazione degli Insegnanti. Torino*: Paravia Scriptorium.

Costamagna, L. and Giannini, S. (2003). *La fonologia dell'interlingua.* Milano: Franco Angeli.

Costamagna, L. and Marotta, G. (2008). *Processi fonetici e categorie fonologiche nell'acquisizione dell'italiano.* Pisa: Pacini Editore.

Costamagna, L., Marasco, M. V., and Santeusanio, N., (2010). *L'italiano con le canzoni*. Perugia: Guerra.

Crystal, D. (1987). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.

Csikszentmihalyi, M. and Larson, R. (1987). Validity and reliability of the experience–sampling method. *Journal of Nervous and Mental Disease*, 175(9), 526–536.

Cunningham, U. 2008. Acoustic variability in the production of English Vowels by Native and Non–Native Speakers. In E. Waniek–Klimczak (Ed.), *Variability in Accents of English: A non– native perspective*. Cambridge: Cambridge Scholar Press.

Dall'Armellina, R., Gori, G., and Turolla, M. L., (2005). *Giocare con la fonetica.* Firenze: Alma Edizioni.

Darcy, I., Ewert, D., and Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation toolbox. In J. Levis and K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 93–108). Ames, 16–17 September 2021.

Davies, B. (2004). The gender gap in modern languages: A comparison of attitude and performance in year 7 and year 10. *Language Learning Journal*, 29, 53–58.

De Marco, A. and Cox, S. J. (2013). Native Accent Classification via I–Vectors and Speaker Compensation Fusion. In *Interspeech 2013* (pp. 1472–1476). Lyon, 25–29 August 2013. ISCA.

De Meo, A., Vitale, M., Pettorino, M., and Philippe, M. (2011). Acoustic–Perceptual Credibility Correlates of News Reading by Native and Chinese Speakers of Italian. In W. S. Lee and E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1366–1369). Hong Kong, 17–21 August 2011.

De Meo, A., Vitale, M., Pettorino, M., Cutugno, F., and Origlia, A. (2013). Imitation/self–imitation in computer– assisted prosody training for Chinese learners

of L2 Italian. In. J. Levis and K. LeVelle (Eds.), *Proceedings of the 4ᵗʰ Pronunciation in Second Language Learning and Teaching Conference* (pp. 90–100). Vancouver, 24–25 August 2012.

De Meo, A., Vitale, M., and Pettorino, M. (2015). Parli nativo? I correlati acustici dell'accento "nativo" in italiano: uno studio acustico–percettivo. In C. Bruno et al. (Eds.), *Plurilinguismo/Sintassi* (pp. 191–206). Roma: Bulzoni.

De Meo, A., Vitale, M., and Pellegrino, E. (2016). Tecnologia della voce e miglioramento della pronuncia in una L2: imitazione e autoimitazione a confronto. Uno studio su sinofoni apprendenti di italiano L2. In F. Bianchi and P. Leone (Eds.), *Studi AItLA 4: Linguaggio e apprendimento linguistico* (pp. 13–25). Milano: AItLA.

Del Monte, R. (2011). Exploring Speech Technologies for Language Learning. In Ipsic, I. (Ed.), *Speech and Language Technologies* (pp. 71–104). London: IntechOpen.

Dellwo, V. (2010). *Influences of speech rate on the acoustic correlates of speech rhythm. An experimental phonetic study based on acoustic and perceptual evidence.* Rheinische Friedrich Wilhelms–Universität Bonn. PhD Dissertation.

Demsar, J., Curk, T., Erjavec, A., Gorup, C., Hocevar, T., Milutinovic, M., Mozina, M., Polajnar, M., Toplak, M., Staric, A., Stajdohar, M., Umek, L., Zagar, L., Zbontar, J., Zitnik, M., and Zupan, B. (2013). Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research*, 14la, 2349−2353.

Derwing, T. M., Munro, M. J., and Wiebe, G. E. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48, 393–410.

Derwing, T. M. and Rossiter, M. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning*, 13, 1–17.

Derwing, T. M., Rossiter, M. J., Munro, M. J., and Thomson, R. I., (2004). L2 fluency: Judgements on different tasks. *Language Learning* 54(4), 655–679.

Derwing, T. M. and Munro, M. J. (2005). Second Language Accent and Pronunciation Teaching: A Research–Based Approach. *TESOL Quarterly*, 39(3), 379–397.

Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., and Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning*, 64(3), 526–48.

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

Devis, A. (2005). Percezione dell'accento straniero. In *Comunicazioni presentate all'VII Convegno AISC* (pp. 1–22). Napoli, 2005

D'Imperio, M. P. and Rosenthall, S. (1999). Phonetics and phonology of main stress in Italian. *Phonology*, 16(1), 1–28.

Dmitrieva, O. (2017). Production of geminate consonants in Russian: Implications for typology. In H. Kubozono (Ed.), *The Phonetics and Phonology of Geminate Consonants* (pp. 34–65). Oxford: Oxford Universiy Press.

Dobrow, J. R. and Gidney, C. L. (1998). The Good, the Bad, and the Foreign: The Use of Dialect in Children's Animated Television. *The ANNALS of the American Academy of Political and Social Science, 557*(1), 105–119.

Doughty, C. and Long, M. (2003). *The handbook of second language acquisition*. Malden: Blackwell Publishers.

Durand, J., Gut, U., and Kristoffersen, G. (Eds.). (2014). *The Oxford Handbook of Corpus Phonology*. Oxford: Oxford University Press.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, 27, 315–330.

Eckman, F. (1984). Universals, typologies and interlanguage. In William Rutherford (ed.), *Language universals and second language acquisition* (pp. 79–105). Amsterdam: John Benjamins.

Eckman, F. (1985). Some theoretical and pedagogical implications of the Markedness Differential Hypothesis. *Studies in Second Language Acquisition*, 7, 289–307.

Eckman, F. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition*, 13, 23–41.

Eckman, F. and Iverson, G. K. (1994). Pronunciation difficulties in ESL: coda consonants in English interlanguage. In H. Yavas (Ed.), *First and second language phonology* (pp. 251–265). San Diego: Singular Publishing Group, Inc.

Edwards, J., Zampini, M., and Hansen, J. G. (2008). *Phonology and second language acquisition*. Amsterdam: John Benjamins.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech Perception in Infants. *Science,* 171, 3968, 303–306.

Elliott, A. R. (1995). Field dependence/independence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *The Modern Language Journal*, 79, 356–371.

Elliott, A. R. (1997). On the Teaching and Acquisition of Pronunciation within a Communicative Approach. *Hispania,* 80(1), 95–108.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (1997). *Second Language Acquisition.* Oxford: Oxford University Press.

Escudero, P. (2005). *Linguistic perception and second language acquisition. Explaining the attainment of optimal phonological categorization*. Utrecht: LOT (Netherlands Graduate School of Linguistics). PhD Dissertation.

Escudero, P. (2007). Second–language phonology: The role of perception. In M. Pennington (Ed.), *Phonology in Context* (pp. 109–134). London: Palgrave Macmillan.

Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning and Technology*, 2(2), 62–76.

Fathman, A. (1975). The Relationship Between Age and Second Language Productive Ability. *Language Learning*, 25(2), 245–253.

Flege, J. E. (1981). The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly*, 15, 443–455.

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America*, 76, 692–707.

Flege, J. E. (1987). The instrumental study of L2 speech production: Some methodological considerations. *Language Learning*, 37, 285–296.

Flege, J. E. (1988a). Effects of equivalence classification on the production of foreign language speech sounds, In A. James and B. Leather (Eds.), *Sound Patterns in Second Language Acquisition* (pp. 9–39). Dordrecht: Foris.

Flege, J. E. (1988b). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, 84, 70–79.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in Cross–Language Research* (pp. 233–277). Baltimore: York Press.

Flege, J. E. (1999). Age Constraints on Second–Language Acquisition. *Journal of Memory and Language*, 41, 78–104.

Flege, J. E. (2003). A method for assessing the perception of vowels in a second language. In E. Fava and A. Mioni (Eds.), *Issues in Clinical Linguistics* (pp. 19–45). Padova: Unipress.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole, and J. Hualde (Eds.), *Laboratory Phonology 9*, (pp. 353–380). Berlin: Mouton de Gruyter.

Flege, J. E. (2008). Give input a chance! In T. Piske and M. Young–Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Bristol: Multilingual Matters.

Flege, J. E. (2018) It's input that matters most, not age. *Bilingualism: Language and Cognition*, 21(5), 919–920.

Flege, J. E., Munro, M. J., and MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125–3134.

Flege, J. E., Bohn O. S., and Jang, S., 1997. Effects of experience on non–native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.

Flege, J. E., Frieda, E., and Nozawa, T. (1997). Amount of native–language (L1) use affects the pronunciation of an L2. *Journal of Phonetics,* 25(2), 169–186.

Flege, J. E., MacKay, I. R. A., and Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of Acoustical Society of America*, 106(5), 2973–2987.

Flege, J. E. and Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, 23, 527–552.

Flege, J. E., Schirru, C., and MacKay, I. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, 40, 467–491.

Flege, J. E. and MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26, 1–34.

Flege, J. E. and Wayland, R. (2019). The role of input in native Spanish Late learners' production and perception of English phonetic segments. *Journal of Second Language Studies*, 2(1), 1–44.

Flege, J. and Bohn, O. (2021). The Revised Speech Learning Model (SLM–r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge: Cambridge University Press.

Fletcher, P. (2005). *Classically Speaking: Dialects for Actors: Neutral American, Classical American, Standard British (RP)*. Bloomington: Trafford.

Formiga, L. and Alias, F. (2009). GMT user modelling for aIGA weight tunning in TTS synthesis. In J. Rabunal Dopico et al. (Eds.), *Encyclopedia of artificial intelligence* (pp. 788–795). Hershey: Information Science Reference.

Foucart, A., Costa, A., Morís–Fernández, L., and Hartsuiker, R. J. (2020). Foreignness or Processing Fluency? On Understanding the Negative Bias Toward Foreign–Accented Speakers. *Language Learning*, 70(4), 974–1016.

Fougeron, C. and Smith, C. L. (1993), Illustrations of the IPA: French. *Journal of the International Phonetic Association*, 23(2), 73–76.

Fouz–González, J. (2015). Trends and Directions in Computer–Assisted Pronunciation Training. In J. Mompean and J. Fouz–González (Eds.) *Investigating English Pronunciation* (pp. 314–342). New York: Palgrave MacMillan.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics*, 14, 3–28.

Fox, C. and Combley, R. (2014). *Longman Dictionary of Contemporary English:* (6th ed.). Harlow: Pearson Education Limited.

Frontera, M. and Paone, E. (2018). Prosodic accuracy and foreign accent in cultural migrants. In *Proceedings of the 9th Tutorial and Research Workshop on Experimental Linguistics – ExLing 2018* (pp. 37–40). Athens, 28–20 August 2018.

Fukuda, T., Fernandez, R., Rosenberg, A., Thomas, S., Ramabhadran, B., Sorin, A., Kurata, G. (2018). Data Augmentation Improves Recognition of Foreign Accented Speech. In *Interspeech 2018* (pp. 2409–2413). Hyderabad, 2–6 September 2018. ISCA.

Furui, S. (1994). An Overview of speaker recognition technology. In *Proceedings of the ESCA Workshop on Automatic Speaker Recognition Identification and Verification* (pp. 31–56). Martigny, 5–7 April 1994.

Gabriel, J. (1994). *Racism, Culture, Markets.* London and New York: Routledge.

Gass, S. and Selinker, L. (2008). *Second language acquisition. An introductory course*. London and New York: Routledge.

Giacalone Ramat, A. (Ed.). (2003). *Verso l'italiano: Percorsi e strategie di acquisizione*. Roma: Carocci.

Giacalone Ramat, A. (2009). Typological universals and second language acquisition. In S. Scalise et al. (Eds.), *Universals of language today* (pp. 253–272). Berlin: Springer.

Gibbon, D., Moore, R., and Winski, R. (1997). *Handbook of Standards and Resources for Spoken Language Systems*, Berlin: Mouton de Gruyter.

Giles, H. and Bourhis, R. Y. (1975). Voice and racial categorization in Britain. *Communication Monographs*, 43, 108–14.

Gluszek, A. and Dovidio, J. F. (2010). The Way They Speak: A Social Psychological Perspective on the Stigma of Nonnative Accents in Communication. *Personality and Social Psychology Review,* 14(2), 214–237.

Goldstein, L. and Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In: N. O. Schiller and A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159–208). Berlin: Mouton de Gruyter.

Gordon, E. C. (1998). *Signal and Linear System Analysis*. New York: John Wiley and Sons Ltd.

Gosselin, L., Martin, C. D., Navarra–Barindelli, E., and Caffarra, S. (2021). The presence of a foreign accent introduces lexical integration difficulties during late semantic processing. *Language, Cognition and Neuroscience*, 36(9), 1086–1106.

Grabe, E. and Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven and N. Warner (Eds.), *Laboratory Phonology 7* (pp. 515–546). Berlin: De Gruyter.

Graesser, A. C., Lu, S., Jackson, G. T., Hite Mitchell, H., Ventura, M., Olney, A., and Loewerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, and Computers*, 36(2), 180–192.

Granlund, S., Baker, R., and Hazan, V. (2011). Acoustic–phonetic characteristics of clear speech in bilinguals. In W. S. Lee and E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 763–766). Hong Kong, 17–21 August 2011.

Grantham O'Brien, M. and Fagan, S. M. B. (2016). *German Phonetics and Phonology: Theory and Practice*. New Haven and London: Yale University Press.

Graves, A., Mohamed, A., and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)* (article 3478, pp. 1–5). Vancouver, 26–31 May 2013.

Greenberg, J. (1978). Some generalizations concerning initial and final consonant clusters. In J. Greenberg et al. (Eds.), *Universals of human language* (Vol. 2, pp. 243–279). Stanford, CA: Stanford University Press.

Grieser, D. and Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech–sound prototypes. *Developmental Psychology,* 25(4), 577–588.

Grosjean, F. (2001). The bilingual's language modes. In J. Nicol (Ed.) *One mind, two languages: Language processing* (pp. 1–22). Oxford: Blackwell.

Gruhn, R. E., Minker, W., and Nakamura, S. (2013). *Statistical Pronunciation Modeling for Non–Native Speech Processing*. Berlin: Springer Berlin.

Gut, U. (2007). Foreign Accent. In C. Müller (Ed.), *Speaker Classification I Fundamentals, Features, and Methods* (pp. 75–87). Berlin: Springer.

Gut, U. (2009). *Non–native speech: A corpus–based analysis of phonological and phonetic properties of L2 English and German*. Frankfurt am Main: Peter Lang.

Hamilton, W. S. (1980). *Introduction to Russian Phonology and Word Structure*. Bloomington: Slavica Publishers.

Hammarberg, B. (2001). Roles of L1 and L2 in L3 production and acquisition. In J. Cenoz et al. (Eds.) *Cross–Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives* (pp. 21–41). Bristol, Blue Ridge Summit: Multilingual Matters.

Hancin–Bhatt, B. and Bhatt, R. M. (1997). Optimal L2 syllables: Interactions of Transfer and Developmental Effects. *Studies in Second Language Acquisition*. 19(3), 331–378.

Hansen, J. and Arslan, L. (1995). Foreign accent classification using source generator based prosodic features. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1995)* (pp. 863–839). Detroit, 9–12 May 1995.

Hanson, B. and Applebaum T. (1990). Robust speaker–independent word recognition using static, dynamic and acceleration features: Experiments with Lombard and noisy speech. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1990)* (pp. 857–860). Albuquerque, 3–6 April 1990.

Hardison, D. (2004). Generalization of computer–assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology*, 8(1), 34–52.

Harmer, J. (2001). *The Practice of English Language Learning*, London: Longman.

Harris, J. (1969). *Spanish Phonology*. Cambridge: MIT Press.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A K–Means Clustering Algorithm. *Journal of the Royal Statistical Society*, Series C., 28(1), 100–108.

Heinzmann, S. (2009). Girls are better at language learning than boys: Do stereotypic beliefs about language learning contribute to girls' higher motivation to learn English in primary school? *Bulletin Suisse de linguistique appliqué,* 89, 19–36.

Hoequist, C. (1983). Syllable Duration in Stress–, Syllable– and Mora– Timed languages. *Phonetica*, 40, 203–237.

Hönig, F. (2017). *Automatic Assessment of Prosody in Second Language Learning*. Berlin: Logos Verlag Berlin GmbH.

Hönig, F., Batliner, A., Nöth, E., Schneider, S., and Krajewski, J. (2014a). Acoustic–Prosodic Characteristics of Sleepy Speech – between Performance and Interpretation. In *Proceedings of Speech Prosody 7* (pp. 864–868). Dublin, 20–23 May 2014.

Hönig, F., Batliner, A., Nöth, E., Schneider, S., Booklet, T., and Krajewski, J. (2014b). Are men more sleepy than women or does it only look like – Automatic analysis of sleepy speech. In *International Conference on Acoustics,*

Speech, and Signal Processing (ICASSP 2014) (pp. 995–999). Florence, 4–9 May.

Hou, J., Liu, Y., Zheng, T. F., Olsen, J., and Tian, J. (2010). Multi–layered features with SVM for Chinese accent identification. In *Proceedings of the 2010 International Conference on Audio, Language and Image Processing* (pp. 25–30). Shanghai, 23–25 November 2010.

Hualde, J. I. (2014). *The sounds of Spanish*. Cambridge: Cambridge University Press.

Huang, Y., Yu, D., Liu, C., and Gong, Y. (2014). Multi–Accent Deep Neural Network Acoustic Model with Accent–Specific Top Layer Using the KLD–Regularized Model Adaptation. In *Interspeech 2014* (pp. 2977–2981). Singapore, 14–18 September 2014. ISCA.

Huckvale, M. (2007). ACCDIST: An Accent Similarity Metric for Accent Recognition and Diagnosis. In C. Müller (Ed.), *Speaker Classification II Selected Projects* (pp. 258–275). Berlin: Springer.

Hugo, V. (1859). *The memoirs of Victor Hugo* (J. W. Harding, Ed.; P. Meurice, Trans.). New York: G.W. Dilliingham.

Jalalvand, S., Akbari, A., and Nasersharif, B. (2012). A classifier combination approach for Farsi accents recognition. *20th Iranian Conference on Electrical Engineering (ICEE2012)* (pp. 716–720). Tehran, 15–17 May 2012.

Jenner, B. (1976). Interlanguage and Foreign Accent. *Interlanguage Studies Bulletin,* 1(2/3), 166–195.

Jiang, N. (2018). *Second Language Processing: An Introduction.* London and New York: Routledge.

Jilka, M. (1999). Intonational foreign accent in the speech of American speakers of German. *The Journal of the Acoustical Society of America,* 105(2), 1094–1094.

Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent: Identifying intonational deviations by means of F0 generation and resynthesis.* Stuttgart: IMS, Lehrstuhl für Experimentelle Phonetik. PhD Dissertation.

Johansson, S. (1978). *Studies in Error Gravity*. Göteborg: Acta Universitatis Gothoburgensis.

Jones, D. and Trofimov M. V. (1923). *The pronunciation of Russian.* Cambridge: Cambridge University Press.

Jones, D. and Ward, D. (1969). *The Phonetics of Russian*. Cambridge: Cambridge University Press.

Kang, O. and Yaw, K. (2021). Social judgement of L2 accented speech stereotyping and its influential factors. *Journal of Multilingual and Multicultural Development*, article 1931247, 1–16.

Kaye, A. (2005). Gemination in English. *English Today*, 21(2), 43–55.

Kelly, G. (2000). *How to teach pronunciation.* London: Pearson Education Limited.

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *The Modern Language Journal*, 97(3), 720–744.

Kitashov, F., Svitanko, E., and Dutta, D. (2018). Foreign English Accent Adjustment by Learning Phonetic Patterns. *arXiv*, article 1807.03625, 1–5.

Klein, W. (1996). Language Acquisition at different ages. In D. Magnusson (Ed.), *Individual Development over the Lifespan: Biological and Psychosocial Perspectives* (pp. 88–108). Cambridge: Cambridge University Press.

Kneser, R., Essen, U., and Ney, H. (1995). On the Use of the Leaving–One–Out Method in Statistical Language Modelling. In A. J. R. Ayuso and J. M. L. Soler (Eds.), *Speech Recognition and Coding* (pp. 256–259). Berlin/Heidelberg: Springer.

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., and Goel, S. (2020). Racial disparities in automated speech recognition. In *Proceedings of the National Academy of Sciences*, 117(14), 7684–7689.

Koffi, E. (2021). *Relevant Acoustic Phonetics of L2 English: Focus on Intelligibility*. London and New York: Routledge.

Kohler, K. (1999). German. In International Phonetic Association (Ed.), *Handbook of the International Phonetic Association* (pp. 86-89). Cambridge: Cambridge University Press.

Kolly, M. J. and V. Dellwo. (2013). (How) can listeners identify the L1 in foreign accented L2 speech? *Travaux neuchâtelois de linguistique*, 59, 127–148.

Kotzor, S., Wetterlin, A., Roberts, A. C., and Lahiri, A. (2016). Processing of phonemic consonant length: Semantic and fragment priming evidence from Bengali. *Language and Speech*, 59(1), 83–112.

Krashen, S. and Terrell, T. D. (1983). *The Natural Approach. Language Acquisition in the Classroom*. San Francisco: Alemany Press.

Kuhl, P. K. (1980). Perceptual constancy for speech–sound categories in early infancy. In G.H Yeni–Komshian et al. (Eds.), *Child phonology* (Vol. 2, pp. 41–66). New York: Academic Press.

Kuhl, P. K. (1992). Infants' perception and representation of speech: Development of a new theory. In J. Ohala et al. (Eds.), *Proceedings of the International Conference on Spoken Language Processing* (pp. 449–456). Alberta, 12–16 October 1992.

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, 4(6), 812–822.

Kuhl, P. K. (2000). Language, mind, and brain: experience alters perception. In Gazzaniga, M. (Ed.), *The new cognitive neurosciences* (pp. 99–115). Cambridge: MIT Press.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831–843.

Kuhl, P. K., Conboy, B., Coffey–Corina, S., Padden, D., Rivera–Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM–e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 979–1000.

Kulshreshtha, M. and Mathur, R. (2012). *Dialect accent features for establishing speaker identity: A case study*. New York: Springer Verlag.

Kumar, K., Kim, C., and Stern, R. M. (2011). Delta–spectral cepstral coefficients for robust speech recognition. *International Conference on Acoustics, Speech*

and Signal Processing (ICASSP 2011) (pp. 4784–4787). Prague, 22–27 May 2011.

Ladefoged, P. (2001). *Vowels and Consonants*. Oxford: Blackwell.

Ladefoged, P. and Maddieson, I. (1996). *Sounds of the World's Languages*. Oxford: Blackwell.

Lado, R. (1957). *Linguistics across cultures*. Michigan: University of Michigan.

Landis, J. and Koch, G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174.

Larsen–Freeman, D. and Long, M. (1991). *An Introduction to Second Language Acquisition Research*. London: Longman.

Lenneberg, E. (1967). *The biological foundations of language*. New York: Wiley and Sons.

Leone, G. R., Paci, G., and Cosi, P. (2012). LUCIA: An Open Source 3D Expressive Avatar for Multimodal h.m.i. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering Intelligent Technologies for Interactive Entertainment,* 78, 193–202.

Lev–Ari, S. and Keysar, B. (2010). Why don't we believe non–native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093–1096.

Levi, S. V., Winters, S. J., and Pisoni, D. B., 2007. Speaker–independent factors affecting the perception of foreign accent in a second language. *The Journal of the Acoustical Society of America*, 121(4), 2327–2328.

Levis, J. M. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics,* 27, 184–202.

Levis, J. M. and Moyer, A. (2014). *Social dynamics in second language accent*. Boston: De Gruyter Mouton.

Liberman, A. M, Cooper, F. S, Shankweiler, D. P, and Studdert–Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology,* 22(140)*,* 55.

Lippi–Green, R. (2012). *English with an accent language, ideology and discrimination in the United States*. London and New York: Routledge.

Lipsitz, G. (2011). *How racism takes place*. Philadelphia: Temple University Press.

Liu, H. M., Kuhl, P. K, and Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, 6, F1–F10.

Liu, K., Wu, C., Tseng, S., and Tsai, Y. (2015). Voice Helper: A Mobile Assistive System for Visually Impaired Persons. In *International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing* (pp. 1400–1405). Liverpool, 26–28 October 2015

Lloyd, P. M. (1987). *From Latin to Spanish*. Philadelphia: American Philosophical Society.

Luo, B. (2016). Evaluating a computer–assisted pronunciation training (CAPT) technique for efficient classroom instruction. *Computer Assisted Language Learning*, 29(3), 451–476.

Macaulay, T. B. (1974). *The letters of Thomas Babington Macaulay* (T. Pinney, Ed.). Cambridge: Cambridge University Press.

MacKay, I. R. A., Flege, J. E., and Imai, S. (2006). Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics*, 27, 157–183.

Maier, A., Hönig, F., Zeissler, V., Batliner, A., Yamanaka, N., Ackermann, P., and Nöth, E. (2009). A Language–Independent Feature Set for the Automatic Evaluation of Prosody. In *Interspeech 2009* (pp. 600–603). Brighton, 6–10 September. ISCA.

Major, R. C. (1994). Chronological and stylistic aspects of second language acquisition of consonant clusters. *Language Learning*, 44, 655–680.

Major, R. C. (1996). Markedness in second language acquisition of consonant clusters. In R. Bayley, R. and D. R. Preston (Eds.), *Second Language Acquisition and Linguistic Variation* (pp. 75–96). Amsterdam: John Benjamins.

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Major, R. C. (2002). The phonology of the L2 user. In Cook, V. (Ed.), *Portraits of the L2 user* (pp. 65–92). Clevedon: Multilingual Matters.

Major, R. C. (2007). Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition*, 29(4), 539–556.

Mangold, M. (2005). *Das Aussprachewörterbuch*. Mannheim: Dudenverlag.

Mannepalli, K., Sastry, P. N., and Suman, M. (2015). MFCC–GMM based accent recognition system for Telugu speech signals. *International Journal of Speech Technology*, 19(1), 87–93.

Manssor, S. A. F., Osman, A. A., and Awadalkareem. S. D. (2015). Controlling home devices for handicapped people via voice command techniques. In *International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE)* (pp. 374–378). Khartoum, 7–9 September 2015.

Markham, D. (1997). *Phonetic Imitation, Accent and the Learner*. Lund: Travaux de l'Institut de Linguistique de Lund.

Marotta, G. (2008). Sulla percezione dell'accento straniero. In R. Lazzeroni et al. (Eds.), *Diachronica et synchronica: Studi in onore di Anna Giacalone Ramat* (pp. 327–347). Pisa: ETS.

Marotta, G. (2009). Remarks on Prosody and Foreign Accent. In M. Bertuccelli Papi et al. (Eds.), *Threads in the Complex Fabric of Language: Linguistica and Literary Studies in Honour of Lavinia Merlini Barbaresi* (pp. 191–205). Pisa: Felici Editore.

Marotta, G. (2016). Prosodic structure. In A. Ledgeway and M. Maiden (Eds.), *The Oxford Guide to the Romance Languages* (pp. 484–494). Oxford: Oxford University Press.

Marotta, G. and Barth, M. (2008). Sonoranti allo specchio: Italiano L2 – Tedesco L1. In L. Costamagna and G. Marotta (Eds.), *Processi fonetici e categorie fonologiche nell'acquisizione dell'italiano* (pp. 93–120). Pisa: Pacini Editore.

Marotta, G. and Boula de Mareuil, P. (2010). Persistenza dell'accento straniero. Uno studio percettivo dell'italiano L2. In S. Schmid et al. (Eds.), *La dimensione temporale del parlato: Atti del V Convegno Nazionale AISV – Associazione Italiana di Scienze della Voce* (pp. 475–494). Zürich, 4–6 February 2009. Torriana: EDK Editore.

Martínez Celdrán, E., Fernández Planas, A. M., and Carrera Sabaté, J. (2003). Castilian Spanish. *Journal of the International Phonetic Association*, 33(2), 255–259.

Marx, N. (2002). Never quite a 'native speaker': Accent and identity in the L2–and the L1. *Canadian Modern Language Review/La Revue canadienne des langues vivantes,* 59(2), 264–281.

Massaro D. W. (2005) The Psychology and Technology of Talking Heads: Applications in Language Learning. In: J. C. van Kuppevelt et al. (Eds.), *Advances in Natural Multimodal Dialogue Systems. Text, Speech and Language Technology* (pp. 183–214). Dordrecht: Springer.

Massaro, D. W., Ouni, S., Cohen, M.M., and Clark, R. (2005). A Multilingual Embodied Conversational Agent. In R. H. Sprague (Ed.), *Proceedings of 38th Annual Hawaii International Conference on System Sciences, (HICCS'05)* (CD–ROM, article 296b, pp. 1–8). Big Island, 6 January 2005. Los Alimitos: IEEE Computer Society Press.

Massaro, D. W., Liu, Y., Chen, T. H., and Perfetti, C. A. (2006). A Multilingual Embodied Conversational Agent for Tutoring Speech and Language Learning. In *Interspeech 2006* (pp. 825–828). Pittsburgh, 17–21 September 2006. ISCA.

Mattys, S. L, Jusczyk, P. W., Luce, P. A., and Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38, 465–494.

Mayfield Tomokiyo, L. and Burger, S. (1999). Eliciting Natural Speech from Non–native Users: Collecting Speech Data for LVCSR. In *Proceedings of the ACL–IALL joint workshop on computer mediated language assessment and evaluation in NLP* (pp. 5–11). Maryland, 22 June 1999.

McAllister, R., Flege, J. E., and Piske, T. (2002). The influence of the L1 on the acquisition of Swedish vowel quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, 30, 229–258.

McCully, C. (2009). *The Sound Structure of English.* Cambridge: Cambridge University Press.

McMahon, A. (2002). *An Introduction to English Phonology*. Oxford: Oxford University Press.

Mehrpour, S., Alavi Shoushtari, S., Haghighat, and Nezhad Shirazi, P. (2016). *Computer Assisteted Language Learning – Electronic Journal*, 17(1), 97–112.

Mermelstein, P. (1976). Distance measures for speech recognition, psychological and instrumental. In C. H. Chen (Ed.), *Pattern Recognition and Artificial Intelligence* (pp. 374–388). New York: Academic Press.

Miller, J. (2003). *Audible difference*. Clevedon: Multilingual Matters.

Missaglia, F. (1999). Contrastive prosody in SLA: An empirical study with adult Italian learners of German. In *Proceedings of the XIVth International Congress of Phonetic Sciences (ICPhS99)* (pp. 555–558). San Francisco, 1–7 August 1999.

Mora, J. C., Keidel, J. L., and Flege, J. E. (2015). Effects of Spanish use on the production of Catalan vowels by early Spanish–Catalan bilinguals. In J. Romero and M. Riera (Eds.), *The Phonetics–Phonology Interface: Representations and methodologies* (pp. 33–54). Amsterdam: John Benjamins.

Mori, L. (2007). *Fonetica dell'italiano L2. Un'indagine sperimentale sulla variazione nell'interlingua dei marocchini.* Roma: Carocci Editore.

Mori, L. and Barkat–Defradas, M. (2005). Acoustic Properties of Foreign Accent: VOT Variations in Moroccan–Accented Italian. In *Interspeech 2005* (pp. 2909–2912). Lisbon, 4–8 September 2005. ISCA.

Moyer, A. (1999) Ultimate attainment in L2 phonology: The critical factors of age, motivation and instruction. *Studies in Second Language Acquisition*, 21, 81–108.

Moyer, A. (2004). *Age, Accent and Experience in Second Language Acquisition. An Integrated Approach to Critical Period Inquiry*. Clevedon: Multilingual Matters.

Moyer, A. (2007). Do language attitudes determine accent? A study of bilinguals in the US. *Journal of Multilingual and Multicultural Development*, 28, 1–17.

Moyer, A. (2008) Input as a means to a critical end: Quantity and quality in L2 phonological attainment. In T. Piske and M. Young–Scholten (Eds.), *Input Matters in SLA* (pp. 159–174). Bristol: Multilingual Matters.

Moyer, A. (2013). *Foreign accent: The phenomenon of non–native speech*. Cambridge: Cambridge University Press.

Munro, M. (1998). The effects of noise on the intelligibility of foreign–accented speech. *Studies in Second Language Acquisition*, 20, 139–154.

Munro, M. J. and Derwing, T. M. (1995), Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, 45, 73–97.

Munro, M. and Derwing, T. (2001). Modeling Perceptions of the Accentedness and Comprehensibility of L2 Speech. The Role of Speaking Rate. *Studies in Second Language Acquisition,* 23(4), 451–468.

Munro, M. and Mann, V. (2005). Age of immersion as a predictor of foreign accent. *Applied Psycholinguistics,* 26(3), 311–341.

Muñoz, C. and Singleton, D. (2011). A critical review of age–related research on L2 ultimate attainment. *Language Teaching,* 44(1), 1–35.

Nejjari, W., Gerritsen, M., van Hout, R., and Planken, B. (2020). Where does a 'foreign' accent matter? German, Spanish and Singaporean listeners' reactions to Dutch–accented English, and standard British and American English accents. *PLoS ONE*, 15(4), article e0231089.

Nespor, M. (1993). *Fonologia*. Bologna: Il Mulino.

Neumeyer, L., Franco, H., Abrash, V., Julia, L., Ronen, O., Bratt, H., Bing, J., Digalakis, V., and Rypa, M. (1998). Webgrader (TM): a multilingual pronunciation practice tool, In *Proceedings of ETRW on Speech Technology in Language Learning (STiLL 1998)* (pp. 61–64). Marholmen, 25–27 May 1998.

Newport, E. L. and Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non–adjacent dependencies. *Journal of Cognitive Psychology*, 48, 127–162.

Nicora, F., Mcloughlin, L. I., and Gili Fivela, B. (2018). Impact of prosodic training on Italian as L2 by Hiberno–English speakers. In *Speech Prosody 9* (pp. 970–974). Poznań, 13–16 June 2018.

Nittrouer, S. (2001). Challenging the notion of innate phonetic boundaries. *The Journal of the Acoustical Society of America*, 110, 1598–1605.

Nodari, R., Galatà, V., and Calamai, S. (2021). Italian school teachers' attitudes towards students' accented speech. A case study in Tuscany. *Italiano LinguaDue*, 13(1), 72–102.

Orrù, P. (2017). *Il discorso sulle migrazioni nell'Italia contemporanea. Un'analisi linguistico–discorsiva sulla stampa (2000–2010)*. Milano: Franco Angeli.

Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder Education.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, 5(3), 261–283.

Padgett, J. (2001). Contrast dispersion and Russian palatalization. In E. V. Hume and K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 187–218). San Diego: Academic Press.

Palermo, M. (2009). ADIL2 come strumento per la ricerca. In M. Palermo (Ed). *Percorsi e strategie di apprendimento dell'italiano lingua seconda: sondaggi su ADIL2* (pp. 3–26). Perugia: Guerra.

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics*, 11, 73–89.

Pedersen, P. (1965). The Mel Scale. *Journal of Music Theory*, 9(2), 295–308.

Pellegrino, E. (2012a). The perception of foreign accented speech. Segmental and suprasegmental features affecting degree of foreign accent in Italian L2. In H. Mello et al. (Eds.) *Proceedings of the VIIth GSCP International Conference – Speech and Corpora* (pp. 261–267). Firenze: Firenze University Press.

Pellegrino, E. (2012b). Quando parlo italiano capiscono subito che sono straniero. Parametri soprasegmentali e tempi di reazione al grado di accento straniero. In M. Falconi and A. Paoloni (Eds.) *La voce nelle applicazioni, Atti del 8° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce* (pp. 81–92), Roma, 25–27 January 2012. Roma: Bulzoni.

Pellegrino, E., Caruso, V., and De Meo, A. (2015). Discriminating between pathological and foreign– accented speech: an acoustic–perceptual study on Italian. In M. Vayra et al. (Eds.), *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio* (pp. 271–287). Milano: Officinaventuno.

Pellegrino, E., Salvati, L., and Vitale, G. (2015). Non voglio parlare in napoletano per questo vado a [ʃkwoːla]. L'influenza del dialetto napoletano sui tratti segmentali dell'italiano L2 di immigrati. *Atti del XLVI Congresso Internazionale di Studi della SLI, Plurlinguismo e Sintassi* (pp. 113–126). Siena, 27–29 September 2012. Roma: Bulzoni.

Pellegrino, E and Vigliano D. (2015). Self imitation in prosody training: A study on Japanese learners of Italian. In S. Steidl et al. (Eds.) *Workshop on Speech and Language Technology in Education* (pp. 53–57). Leipzig, 4–5 September 2015. ISCA.

Pellegrino, E. and Dellwo, V. (2015). Listeners pay attention to rhythmic cues when judging the nativeness of speech. *The Journal of the Acoustical Society of America,* 138(3), 1946.

Penfield, W. and Roberts, L. (1959). *Speech and brain mechanisms*. Princeton: Princeton University Press.

Peukert, H. (2015). *Transfer effects in multilingual language development*. Amsterdam: John Benjamins.

Pharies, D. (2007). *A Brief History of the Spanish Language*. Chicago: University of Chicago Press.

Pinget, A., Bosker, H. R., Quené, H., and Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, 31(3), 349–365.

Piske, T. (2008). Phonetic Awareness, Phonetic Sensitivity and the Second Language Learning. In J. Cenoz (Ed.), *Encyclopaedia of Language and Education*, (Vol. 6, pp. 155–166), Berlin: Springer.

Piske, T. (2012). Factors Affecting the Perception and Production of L2 Prosody: Research Results and Their Implications for the Teaching of Foreign Languages. In J. Romero–Trillo (Ed.), *Pragmatics and Prosody in English Language Teaching* (pp. 41–59). Dordrecht: Springer.

Piske, T., MacKay, I. and Flege, J. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29(2), 191–215.

Polat, N. and Mahalingappa, L. G. (2010). Gender Differences in Identity and Acculturation Patterns and L2 Accent Attainment. *Journal of Language, Identity, and Education*, 9, 17–35.

Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F–Measure to ROC, Informedness, Markedness and Correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.

Postman, L. (1971). Transfer, interference and forgetting. In J. W. Kling and L. A. Riggs (Eds.), *Woodworth and Schlosberg's Experimental Psychology* (pp. 1019–1132). New York: Holt, Rinehart and Winston.

Preston, D. R. (Ed.). (1999a). *Handbook of Perceptual Dialectology*. Amsterdam: John Benjamins.

Preston, D. R. (1999b). A language attitude approach to the perception of regional variety. In D. R. Preston (Ed.), *A Handbook of Perceptual Dialectology* (pp. 359–373). Amsterdam: John Benjamins.

Preston, D. R. (2021). Métodos em Linguística Popular (aplicada): o que pensa o povo? *Estudos da Língua(gem)*, 19(2), 9–42.

Purcell, E. T. and Suter, R. W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, 30, 271–287.

Radzikowski, K., Wang, L., Yoshie, O., and Nowak, R. (2021). Accent modification for speech recognition of non–native speakers using neural style transfer. *Journal of Audio, Speech, and Music Processing*, 11, article 001993, 1–10.

Rakić, T. and Steffens, M. C. (2013). Language attitudes in Western Europe. In H. Giles and B. Watson (Eds.), *The social meanings of language, dialect, and accent: international perspectives on speech styles* (pp. 45–64). Bern: Peter Lang.

Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In *Speech Prosody 1* (pp. 115–120). Aix–en–Provence, 11–13 April 2002.

Rämö, A., Nurminen, J., Himanen, S., and Heikkinen, A. (2004). Segmental speech coding model for storage applications. In *Interspeech 2004* (pp. 2677–2680). Jeju Island, 4–8 October 2004. ISCA.

Reimann, A. (2018). Behaviorist Learning Theory. In J. I. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1–6). New York: John Wiley and Sons Inc.

Rivera–Campos, A. and Boyce S. (2013). Describing alternative articulations of the Spanish trill /r/ by ultrasound technology. *Proceedings of Meetings on Acoustics*, 19(1), article 060251, 2–5.

Rizzolatti, G. and Craighero, L. (2004). The Mirror–Neuron system. *Annual Review of Neuroscience*, 27, 169–192.

Roessel, J., Schoel, C., Zimmermann, R., and Stahlberg, D. (2019). Shedding New Light on the Evaluation of Accented Speakers: Basic Mechanisms Behind Nonnative Listeners' Evaluations of Nonnative Accented Job Candidates. *Journal of Language and Social Psychology*, 38(1), 3–32.

Romito, L. (2013). *Manuale di linguistica forense*. Roma: Bulzoni.

Romito L. and Tarasi A. (2012). Rhythmical–prosodic analysis of Italian L2. In A. De Meo and M. Pettorino (Eds.), *Prosodic and Rhythmic Aspects of L2 Acquisition: The Case of Italian* (pp. 137–152). Cambridge: Cambridge Scholar Publishing.

Romito, L., Savy, R., Tarasi, A. and Lio, R. (2012). Analisi ritmica dell'italiano L2 prodotto da parlanti albanesi, cinesi, polacchi e rumeni. In *Atti del VIII Convegno AISV* (pp. 93– 106). Torriana: EDK Editore.

Rossini Favretti, R, Tamburini, F., and De Santis, C. (2003). *A corpus of written Italian: a defined and a dynamic model*. In A. Wilson et al. (Eds.), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World* (pp. 27–38). München: Lincom.

Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., and Landay, J. A. (2018). Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*,1(4), 1–23.

Russo, M., Islam, G., and Koyuncu, B. (2017). Non–native accents and stigma: How self–fulfilling prophesies can affect career outcomes. *Human Resource Management Review*, 27(3), 507–520.

Saito, K. (2011). Examining the role of explicit phonetic instruction in native–like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology. *Language Awareness*, 20, 45–59.

Saito, K. and Lyster, R. (2012). Effects of form–focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, 62, 595–633.

Schmid, S. and Pedrazzini, G. (2016). La pronuncia delle occlusive nel Tedesco L2 di apprendenti italofoni: un esperimento didattico. In R. Savy and V. Alfano (Eds.), *La fonetica nell'apprendimento delle lingue* (pp. 45–60), Milano: Officinaventuno.

Schwarz, P., Matejka, P., and Cernocky, J. (2006). Hierarchical structures of neural networks for phoneme recognition. In *International Conference on Acoustics, Speech, and Signal Processing* (*ICASSP 2006)* (pp. 325–328). Toulouse, 14–19 May 2006.

Scovel, T. (1969). Foreign Accents, Language Acquisition, and Cerebral Dominance. *Language Learning*, 19, 245–253.

Scovel, T. (2000). A critical review of the Critical Period research. *Annual Review of Applied Linguistics*, 20, 213–223.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209–231.

Setter, J. (2008). Theories and Approaches in English Pronunciation. In R. Monroy and A. Sánchez (Eds.), 2*5 años de Lingüística aplicada en España: hitos y retos* (pp. 447–457). Murcia: Editum.

Siddhant, A., Jyothi, P., and Ganapathy, S. (2017). Leveraging native language speech for accent identification using deep Siamese networks. In *Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 621–628). Okinawa, 16–20 December 2017.

Silke, M. W. (2012). Automatic Error Detection in Pronunciation Training: Where we are and where we need to go. In O. Engwall (Ed.), *Proceedings of International Symposium on automatic detection on errors in pronunciation training* (pp. 1–8), Stockholm, 6–8 June 2012. Stockholm: KTH.

Singleton, D. (2005). The Critical Period Hypothesis: A coat of many colours. *International Review of Applied Linguistics in Language Teaching*, 43, 269–285.

Sisinni, B. (2016). *Fonetica e fonologia della seconda lingua: Teorie, metodi e prospettive per la didattica*. Roma: Carocci.

Skinner, E. (1990). *Speak with Distinction*. Milwaukee: Applause Theatre Book Publishers.

Soky, K., Li, S., Mimura, M., Chu, C., and Kawahara, T. (2021). On the Use of Speaker Information for Automatic Speech Recognition in Speaker–imbalanced Corpora. In *Asia–Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)* (pp. 433–437). Tokyo, 14–17 December 2021.

Southwood, H. and Flege, J. E. (1999). The validity and reliability of scaling foreign accent. *Clinical Linguistics and Phonetics*,13, 335–349.

Steidl, S., Riedhammer, K., Bocklet, T., Hönig, F., and Nöth, E. Java Visual Speech Components for Rapid Application Development of GUI based Speech Processing Applications. In *Interspeech 2011* (pp. 3257–3260). Florence, 27–31 August. ISCA.

Stevens, S. S., Volkmann, J., and Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, 8(3), 185–190.

Stocker, L. (2016). The Impact of Foreign Accent on Credibility: An Analysis of Cognitive Statement Ratings in a Swiss Context. *Journal of Psycholinguistic Research*, 46(3), 617–628.

Studdert–Kennedy, M. and Goldstein, L. (2003). Launching language: The gestural origin of discrete infinity. In M. H. Christiansen and S. Kirby (Eds.), *Language evolution* (pp. 235–254). Oxford: Oxford University Press.

Sturim, D. E., Torres–Carrasquillo, P. A., and Campbell, J. P. (2016). Corpora for the Evaluation of Robust Speaker Recognition Systems. In *Interspeech 2016* (pp. 2776–2780). San Francisco, 8–12 September 2016. ISCA.

Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning*, 26, 233–253.

Tahta, S., Wood, M., and Loewenthal, K. (1981). Age Changes in the Ability to Replicate Foreign Pronunciation and Intonation. *Language and Speech*, 24(4), 363–372.

Tamburini, F. (2022). *Neural models for the automatic processing of Italian*. Bologna: Pàtron Editore.

Tanner, M. and Landon, M. (2009). The Effects of Computer–assisted Pronunciation Readings on ESL Learners' Use of Pausing, S., Intonation, and Overall Comprehensibility. *Language Learning and Technology*, 13(3), 51–65.

Tatman, R. (2017). *Modeling the Perceptual Learning of Novel Dialect Features*. University of Washington. PhD Dissertation.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, 41, 177–204.

Thomson, R. I. and Derwing, T. M. (2015). The Effectiveness of L2 Pronunciation Instruction: A Narrative Review. *Applied Linguistics*, 36(3), 326–344.

Timming, A. R. (2017). The effect of foreign accent on employability: a study of the aural dimensions of aesthetic labour in customer–facing and non–customer–facing jobs. *Work, Employment and Society*, 31(3), 409–428.

Tranel, B. (1987). *The Sounds of French: An Introduction*. Cambridge: Cambridge University Press.

Tyler, M., Best, C., Faber, A., and Levitt, A. (2014) Perceptual assimilation and discrimination of non–native vowel contrasts. *Phonetica*, 71, 4–21.

Van Patten, B. and Benati, A. G. (2015). *Key Terms in Second Language Acquisition*. London: Bloomsbury Publishing.

Vasiliu, E. (1968). *Fonologia istorică a dialectelor dacoromâne*. București: Editura Academiei Republicii Socialiste România.

Vedovelli, M. (1991). Apprendimento linguistico nel contesto dell'immigrazione. *Scuola Democratica*, 1/1991, 119–130.

Vedovelli, M. (1994). Fossilizzazione, cristallizzazione, competenza di apprendimento spontaneo. In A. Giacalone Ramat and M. Vedovelli (Eds.), *Italiano: lingua seconda, lingua straniera: Atti del XXVI Congresso della Società di linguistica italiana* (pp. 519–547). Siena, 5–7 November 1992. Roma: Bulzoni.

Verma P. and Das, P. K. (2015). i–Vectors in speech processing applications: a survey. *International Journal of Speech Technology*, 18(4), 529–546.

Vieru, B. and Boula de Mareüil, P. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica*. 63(4), 247–267.

Vieru, B., Boula de Mareüil, P., and Adda–Decker, M. (2007). Identification of foreign–accented French using data–mining techniques. *International Workshop on Paralinguistic Speech* (pp. 47–52). Saarbrücken, 3 August 2007.

Vieru, B., Boula de Mareüil, P., and Adda–Decker, M. (2011). Characterisation and identification of non–native French accents. *Speech Communication*, 53(3), 292–310.

Vitale, M., De Meo, A., and Pettorino, M. (2012). La competenza prosodico–intonativa nell'italiano L2: analisi e sintesi del segnale fonico di cinesi, giapponesi e vietnamiti. In S. Ferreri (Ed.), *Linguistica educativa* (pp. 329–342). Roma: Bulzoni.

Vitale, M., Boula de Mareüil, P., De Meo, A. (2014). An acoustic–perceptual approach to the prosody of Chinese and native speakers of Italian based on yes/no questions. In *Speech Prosody 7* (pp. 648–652). Dublin, 20–23 May 2014.

Vitale, M., Boula de Mareüil, P., and De Meo, A. (2017). Chiedere in italiano: le domande polari e lo sviluppo della competenza prosodica in parlanti cinesi di italiano L2. In A. Gundmundson et al. (Eds.), *Romance Languages: Multilingualism and Language Acquisition* (pp. 143–159). Bern: Peter Lang.

Walker, D. C. (1975). Word Stress in French. *Language*, 51(4), 887–900.

Walker, D. C. (2001). *French sound structure*. Calgary: Hignell Book Printing.

Walker, R. (2014). Technology for pronunciation. *English Teaching Professional*, 95, 29–31.

Wang, X. and Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, 32, 539-552.

Ward., J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58, 236–244.

Weninger, F., Sun, Y., Park, J., Willett, D., and Zhan, P. (2019). Deep Learning based Mandarin Accent Identification for Accent Robust ASR. In *Interspeech 2019* (pp. 510–514). Graz, 15–19 September 2019. ISCA.

Weinreich, U. (1953). *Languages in contact: Findings and problems*. New York: Mouton.

Werker, J. F., Gilbert, J. H., Humphrey, K., and Tees, R. C. (1981). Developmental aspects of cross–language speech perception. *Child Development*, 52(1), 349–355.

Werker, J. F. and Tees, R. C. (2002). Cross–language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 25(1), 121–133.

Whaley, L. J. (1997). *Introduction to Typology. The unity and diversity of language*. Thousand Oaks: SAGE Publications.

Wolfson, N. (1976). Speech events and natural speech: Some implications for sociolinguistic methodology. *Language in Society*, 5(2), 189–209.

Wottawa, J. and Adda–Decker, M. (2016). French Learners Audio Corpus of German Speech (FLACGS). In *Proceedings of the 10th Conference on Language Resources and Evaluation (LREC 2016)* (pp. 3215–3219). Portorož, 23–28 May 2016. European Language Resources Association (ELRA).

Yeni–Komshian, G., Flege, J. E., and Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean–English bilinguals. *Bilingualism: Language and Cognition*, 3(2), 31–149.

Zárate–Sández, G. (2017). Reexamining Foreign–Accent: How much Can Personality Explain? *Ilha Desterro*, 70(3), 227–243.

Zheng, Y., Sproat, R., Gu, L., Shafran, I., Zhou, H., Su, Y., Jurafsky, D., Starr, R., and Yoon, S. (2005). Accent detection and speech recognition for shanghai–accented mandarin. In *Eurospeech 2005* (pp. 217–220). Lisbon, 4–8 September 2005.

Zhi, N. and Li, A. (2021). Acoustic salience in the evaluations of intelligibility and foreign accentedness of nonnative vowel production. *Lingua*, 256, article 103069, 1–19.

Materiali linguistici
a cura dell'Università di Pavia, Dipartimento di Studi Umanistici - Sezione
Linguistica Teorica e Applicata

Ultimi volumi pubblicati:

IRENE FIORAVANTI, *Tra le parole nella mente.* Studi interdisciplinari sulle collocazioni lessicali (disponibile anche in e-book).

CLAUDIA MARZI, *Modelling the morphological lexicon.* A computational approach to mono- and bilingual learning and processing of verb inflection (disponibile anche in e-book).

AISHA NASIMI, *L'arabo L2.* Valutazione e certificazione delle competenze (disponibile anche in e-book).

PATRIZIA SORIANELLO, *Linguaggio e sindrome di down* (disponibile anche in e-book).

DANIELA MEREU, *Il sardo parlato a Cagliari.* Una ricerca sociofonetica (disponibile anche in e-book).

MARINA CHINI, PIERLUIGI CUZZOLIN (a cura di), *Tipologia, Acquisizione, Grammaticalizzazione.* Typology, Acquisition, Grammaticalization Studies.

MARINA CHINI, CECILIA MARIA ANDORNO (a cura di), *Repertori e usi linguistici nell'immigrazione.* Una indagine su minori alloglotti dieci anni dopo (disponibile anche in e-book).

MAURIZIO BAGATIN, *Il Mi'at 'Āmil Fī An-Nahw Di 'Abd Al-Qāhir Al-Ǧurǧānī.* Un trattato didattico sugli operanti grammaticali in arabo (disponibile anche in e-book).

ILARIA FIORENTINI, *Segnali di contatto.* Italiano e ladino nelle valli del Trentino-Alto Adige (disponibile anche in e-book).

ANNA GIACALONE RAMAT, PAOLO RAMAT, *Scripta linguistica minora.*

MANANA TOPADZE GAUMANN, *Mezzi di espressione dell'evidenzialità in Georgiano* (disponibile anche in e-book).

MARIA ROSA CAPOZZI, *La comunicazione pubblicitaria.* Aspetti linguistici, sociali e culturali.

LORENZO FILIPPONIO, CHRISTIAN SEIDL (a cura di), *Le lingue d'Italia e le altre.* Contatti, sostrati e superstrati nella storia linguistica della Penisola.

MARINA CHINI (a cura di), *Il parlato in (italiano) L2.* Aspetti pragmatici e prosodici (disponibile anche in e-book).

ALEXANDRU LAURENTIU COHAL, *Mutamenti nel romeno di immigrati in Italia* (disponibile anche in e-book).

EMANUELE MIOLA, *Innovazione e conservazione in un dialetto di crocevia.* Il kje di Prea (disponibile anche in e-book).

GIOVANNA ALFONZETTI, *Il discorso bilingue.* Italiano e dialetto a Catania.

MARIA CHIARA FELLONI, *Prosodia sociofonetica.* L'italiano parlato e percepito a Parma (disponibile anche in e-book).

MARINELLA CARUSO, *Italian language attrition in Australia.* The verb system (disponibile anche in e-book).

MARINA CHINI (a cura di), *Topic, struttura dell'informazione e acquisizione linguistica* (disponibile anche in e-book).

FABIANA ROSI, *Learning Aspect in Italian L2.* Corpus annotation, acquisitional patterns, and connectionist modelling (disponibile anche in e-book).

FABIO MONTERMINI, *Il lato sinistro della morfologia.* La prefissazione in italiano e nelle lingue del mondo (disponibile anche in e-book).

GIORGIO FRANCESCO ARCODIA, *La derivazione lessicale in cinese mandarino* (disponibile anche in e-book).

LIDIA COSTAMAGNA, STEFANIA SCAGLIONE (a cura di), *Italiano acquisizione e perdita.*

ANDREA SANSÒ (a cura di), *Language resources and linguistic theory.*

CRISTINA MARIOTTI, *Interaction strategies in english-medium instruction.*

CARLOTTA VITI, *Strategies of subordination in Vedic.*

ANNA CILIBERTI (a cura di), *La costruzione interazionale di identità.* Repertori linguistici e pratiche discorsive degli italiani in Australia.

BARBARA TURCHETTA, *Lingua e diversità.* Multilinguismo e lingue veicolari in Africa occidentale.

PIERLUIGI CUZZOLIN, MARIA NAPOLI (a cura di), *Fonologia e tipologia lessicale nella storia della lingua greca.* Atti del VI incontro internazionale di linguistica greca.

MARIA NAPOLI, *Aspect and Actionality in Homeric Greek.* A contrastive analysis.

DOMENICA ROMAGNO, *Il perfetto omerico.* Diatesi azionalità e ruoli tematici.

ALESSANDRO MENGOZZI (a cura di), *Studi afroasiatici.* XI incontro italiano di linguistica camitosemitica.

FEDERICA DA MILANO, *La deissi spaziale nelle lingue d'Europa.*

ANNALISA BAICCHI, CRISTIANO BROCCIAS, ANDREA SANSÒ (a cura di), *Modelling thought and constructing meaning.* Cognitive models in interaction.

NICOLA GRANDI (a cura di), *Morfologia e dintorni.* Studi di linguistica tipologica ed acquisizionale.

ANDREA TROVESI, *La genesi di articoli determinativi.* Modalità di espressione della definitezza in ceco, serbo-lusaziano e sloveno.

CHIARA CELATA, *Acquisizione e mutamento di categorie fonologiche.* Le affricate in italiano.

CARLA BAGNA, *La competenza quasi-bilingue/quasi-nativa.* Le preposizioni in italiano L2.

MASSIMO VEDOVELLI, STEFANIA MASSARA, ANNA GIACALONE RAMAT (a cura di), *Lingue e culture in contatto.* L'italiano come L2 per gli arabofoni.

ANNA GIACALONE RAMAT, EDDO RIGOTTI, ANDREA ROCCI (a cura di), *Linguistica e nuove professioni.*

ANDREA SANSÒ, *Degrees of event elaboration.* Passive construction in Italian and Spanish.

EVA-MARIA THUNE, SIMONA LEONARDI (a cura di), *Telefonare in diverse lingue.* Organizzazione sequenziale, routine e rituali in telefonate di servizio, di emergenza e fatiche.

LUISA AMENTA, *Perifrasi aspettuali in greco e in latino.* Origini e grammaticalizzazioni.

Vi aspettiamo su:

www.francoangeli.it

per scaricare (gratuitamente) i cataloghi delle nostre pubblicazioni

DIVISI PER ARGOMENTI E CENTINAIA DI VOCI: PER FACILITARE
LE VOSTRE RICERCHE.

Management, finanza,
marketing, operations, HR

Psicologia e psicoterapia:
teorie e tecniche

Didattica, scienze
della formazione

Economia,
economia aziendale

Sociologia

Antropologia

Comunicazione e media

Medicina, sanità

Architettura, design,
territorio

Informatica, ingegneria

Scienze

Filosofia, letteratura,
linguistica, storia

Politica, diritto

Psicologia, benessere,
autoaiuto

Efficacia personale

Politiche
e servizi sociali

FrancoAngeli

La passione per le conoscenze

FrancoAngeli

a strong international commitment

Our rich catalogue of publications includes hundreds of English-language monographs, as well as many journals that are published, partially or in whole, in English.

The **FrancoAngeli**, **FrancoAngeli Journals** and **FrancoAngeli Series** websites now offer a completely dual language interface, in Italian and English.

Since 2006, we have been making our content available in digital format, as one of the first partners and contributors to the **Torrossa** platform for the distribution of digital content to Italian and foreign academic institutions. **Torrossa** is a pan-European platform which currently provides access to nearly 400,000 e-books and more than 1,000 e-journals in many languages from academic publishers in Italy and Spain, and, more recently, French, German, Swiss, Belgian, Dutch, and English publishers. It regularly serves more than 3,000 libraries worldwide.

Ensuring international visibility and discoverability for our authors is of crucial importance to us.

FrancoAngeli

torrossa
Online Digital Library

Materiali Linguistici

This book investigates non-native Italian speech, focusing on Russian, English, German, French, Romanian, and Spanish accents, while also drawing comparisons to the native varieties of the language. By using a corpus that includes over 8 hours of read and spontaneous speech, collected from 122 young adult speakers of L1 and L2 Italian, the work examines various factors that influence how foreign accents are perceived and recognized. A mixed-method approach that combines quantitative and qualitative analyses is deployed, incorporating machine learning classification techniques as well as accent perception experiments with Italian listeners. The monograph offers a novel empirical perspective on the phenomenon of foreign accent, examining and discussing the specific segmental and suprasegmental features that contribute to the perception and identification of each accent. The findings yielded by this work hold the potential to serve as a groundwork for future analyses on foreign accent.

Claudia Roberta Combei is a Researcher (RTD-a) in Linguistics at the University of Pavia where she teaches Phonetics and Phonology and Computational Linguistics. She earned her PhD in 2019 from the University of Pisa. Her research employs a mixed-method approach, investigating topics such as non-native speech, patterns of vowel variation and change, and more recently political and (social) media discourse. She co-coordinates the RadioCast-It speech corpus and serves as a member of the Editorial Board of Bibliography of Metaphor and Metonymy – MetBib (John Benjamins). Since 2023, she has been the Coordinator of the Experimental Phonetics Laboratory at the University of Pavia.

FrancoAngeli
La passione per le conoscenze