Effects of age and language exposure on the acquisition of pronunciation:

Focus on primary school Tunisian Arabic learners of French

Amina Affes

A Thesis
In the Department
of
English / Individualized Program

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy Linguistics / Humanities

At

Faculté des Lettres et des Sciences Humaines Sfax, University of Sfax, Sfax, Tunisia

And

Concordia University, Montreal, Quebec, Canada

October 2022

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Amina Affes

Entitled: Effects of age and language exposure on the acquisition of pronunciation: Focus on primary school Tunisian Arabic learners of French

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy (Individualized Program)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

____Chair
Dr. Nina Howe

____ External Examiner
Dr. Kevin Papin

____ Examiner
Dr. Walcir Cardoso

____ Examiner
Dr. Denis Liakin

____ Examiner
Dr. Diane Querrien

____ Examiner
Dr. Mounir Triki

____ Thesis Co-Supervisor
Dr. Adel Jebali

____ Thesis Co-Supervisor
Dr. Chokri Smaoui

Approved by ____
Dr. Felice Yuen, Graduate Program Director

1/27/2023 ____
Dr. Effrosyni Diamantoudi, Dean
School of Graduate Studies

Abstract

Effects of Age and Language Exposure on the Acquisition of Pronunciation: Focus on Primary School Tunisian Arabic Learners of French

Amina Affes, Ph.D.

Sfax University and Concordia University, 2023

Despite considerable progress in research on the impact of age of acquisition (AOA) and language exposure on general second language (L2) learning, pronunciation has received limited attention in the school setting and findings remain inconclusive: while some studies conclude that earlier starters outperformed late starters (e.g., Fullana & Mora, 2007), others find that neither AOA nor exposure have a significant effect on participants' pronunciation (e.g., MacKay & Fullana, 2007). A possible explanation for these inconclusive results could be that researchers adopt different tasks to assess pronunciation, without considering the effects they can have on learners' performance (Saito & Plonsky, 2019).

To address the effects of AOA, language exposure, and task type on the acquisition of L2 pronunciation, this thesis examined the oral production of 68 Tunisian learners of French (age 8-12) considering two AOAs (4 and 8 years), two levels of L2 exposure (1000h and 1600h), using four task types: word list reading, sentence reading, picture identification, and storytelling. The analysis focused on the participants' global (accentedness, comprehensibility, and intelligibility) and specific (four vowels and consonants, and stress) pronunciation knowledge. For each task, global pronunciation was assessed by three non-native listeners (using ratings and transcriptions, Derwing & Munro, 1997), while specific pronunciation was measured using perceptual auditory and acoustic analysis. Results indicate that there was no significant difference between older and younger learners. Also, results demonstrated that the group with more exposure (1600h) produced more target-like forms than the group with less exposure (1000h). Significant task effects were observed, with participants producing more target-like forms in more controlled tasks. These findings suggest that exposure to the target language plays a more important role in determining learners' pronunciation abilities in comparison with AOA, and that task type must be taken into consideration in studies examining L2 phonological development.

Résumé

Les effets de l'âge et de l'exposition à la langue sur l'acquisition de la prononciation : à travers les jeunes apprenants arabophones du français au primaire

Amina Affes, Ph.D.

Sfax University and Concordia University, 2023

Malgré les progrès considérables dans la recherche sur l'impact de l'âge d'acquisition (AA) et de l'exposition à la langue sur l'apprentissage général de la langue seconde (L2), la prononciation a reçu une attention limitée en milieu scolaire et les conclusions restent peu concluantes: tandis que certaines études concluent que les apprenants précoces ont obtenu de meilleurs résultats que les apprenants tardifs (Fullana & Mora, 2007), d'autres trouvent que ni l'AA ni l'exposition n'ont un effet significatif sur la prononciation des participants (MacKay & Fullana, 2007). Une explication possible de ces résultats pourrait être que les chercheurs adoptent des tests différents pour évaluer la prononciation, sans considérer les effets qu'elles peuvent avoir sur la performance des apprenants (Saito & Plonsky, 2019).

Pour examiner les effets de l'AA, de l'exposition à la L2 et du type de tests sur l'acquisition de la prononciation en L2, cette thèse a examiné la production orale de 68 élèves tunisiens apprenants du français (8-12 ans) en considérant deux AA (4 et 8 ans), deux niveaux d'exposition à la L2 (1000h et 1600h), et en utilisant quatre types de tests: lecture de listes de mots, lecture de phrases, identification d'images et narration d'une histoire. L'analyse s'est concentrée sur la prononciation globale (accentuation, compréhensibilité et intelligibilité) et spécifique (quatre voyelles et consonnes et accent tonique) des participants. Pour chaque tâche, la prononciation globale a été évaluée par trois évaluateurs non natifs (en utilisant des évaluations et des transcriptions, Derwing & Munro, 1997), tandis que la prononciation spécifique a été mesurée à l'aide des analyses perceptuelles auditives et acoustiques. Les résultats indiquent qu'il n'y avait pas de différence entre les apprenants précoces et les apprenants tardifs. En outre, les résultats ont montré que le groupe avec plus d'exposition (1600h) a obtenu de meilleurs résultats que le groupe avec le moins d'exposition (1000h). Nous avons également observé des effets significatifs des tests, où les participants ont obtenu de meilleurs résultats dans les tests contrôlés que dans des tets spontanés. Ces résultats suggèrent que l'exposition à la langue cible joue un rôle plus important dans la détermination des capacités de prononciation des apprenants que l'AA, et que le type de tâche doit être pris en compte dans les études examinant le développement phonologique de la L2.

Acknowledgements

I would like to take a moment to express my gratitude to those who have contributed to this dissertation and supported me during this journey in one way or another.

Most importantly, I want to thank my supervisor, Dr. Chokri Smaoui at the Faculty of Letters and Humanities of Sfax, whose passion for second language speech learning inspired me to begin this adventure since my first year at the Faculty of Letters of Sfax. He accepted to be my supervisor without a second thought and believed in me. His constant guidance and support were invaluable. Thank you for supporting me during the hard times especially during COVID. Thank you!

I also wish to express my deepest gratitude to Dr. Adel Jebali, at Concordia University. His unconditional help during various stages of the project, support, and readiness to offer advice, whether it was a working day or not, was invaluable. I am extremely grateful for having had the opportunity to have him as my mentor. Thank you for always checking up on me during all the difficult times that I've been through even when I'm on a different continent. I wouldn't have completed this thesis without your support. I will always be grateful.

I am thankful for Dr. Walcir Cardoso who guided me and was always available to answer my numerous emails even on weekends. His expertise, work ethic, and always-present support were the highlight of my journey. Thank you so much and I am very honored to have you as a member of my thesis committee. I am also grateful to Dr. Denis Liakin whose valuable suggestions guided me through this thesis. Thank you.

I am very grateful to all my little participants. My most sincere thank-you for making this study possible to all the headmasters of the twelve schools in Sfax who accepted to welcome me through their premises and allowed me to collect data during the most difficult times of COVID.

I am especially thankful for father Mohammed, my hero and my mother Hela, my everything. I wouldn't have completed this thesis without you. You've always been there during my ups and downs and always pushed me to bring out the best of me. You illuminated my path with your constant support, guidance, sacrifices and prayers. No words can express my gratitude. I am also incredibly grateful for my sisters Fatma and Feten who have always believed in me, as well as my nephews Youssef and Sofien and to my niece Lina, who are my bundles of joy. I am forever grateful!

Finally, I would like to express my sincere appreciation to my grandparents, to my brother-in-law Mahmoud, to all my family members and to my friends, especially Emna, who stood by me and supported me with their thoughts and advice.

Thank you!

Dedication

Pour papa et maman

Je vous aime jusqu'à l'éternité

Table of Contents

List of Tables

List of Figures

List of abbreviations

AOA	Age of Acquisition
BAE	Beyond Age Effects
BAF	Barcelona Age Factor
CAH	Contrastive Analysis Hypothesis
CPH	Critical Period Hypothesis
DV	Dependent Variable
IPA	International Phonetic Alphabet
IV	Independent Variable
FL	Foreign Language
L1	First Language
L2	Second Language
L3	Third Language
MDH	Markedness Differential Hypothesis
MSA	Modern Standard Arabic
OPM	Ontogeny Phylogeny Model
OT	Optimality Theory
SLA	Second Language Acquisition
SLM	Speech Learning Model
SLM-r	Revised Speech Learning Model
TA	Tunisian Arabic
TL	Target Language
PAM	Perceptual Assimilation Model
VOT	Voice Onset Time

Chapter 1

Introduction

1. Background of the study

Age effects on the acquisition of foreign languages constitute one of the most investigated and debated topics over the past decades in the field of second language acquisition through diverse lenses (Muñoz, 2010, 2019). Age of acquisition is defined as the age at which significant exposure to the target language begins. Traditionally, the majority of the age effects studies have focused on the impact of age of acquisition (AOA) on the development of the phonological abilities in terms of the Critical Period Hypothesis (Lenneberg, 1967). Generally, these studies were administered in natural settings with immigrants. The findings of these naturalistic setting studies have shown that there is an advantage in starting L2 learning from an early age, indicating that learners can attain a native-like pronunciation. As a result, the long-standing assumption “the younger, the better” has been reinforced (e.g., Muñoz, 2008). This belief states that young children are intrinsically better language learners, and the younger they are, the quicker they will foster a swift and successful language proficiency.

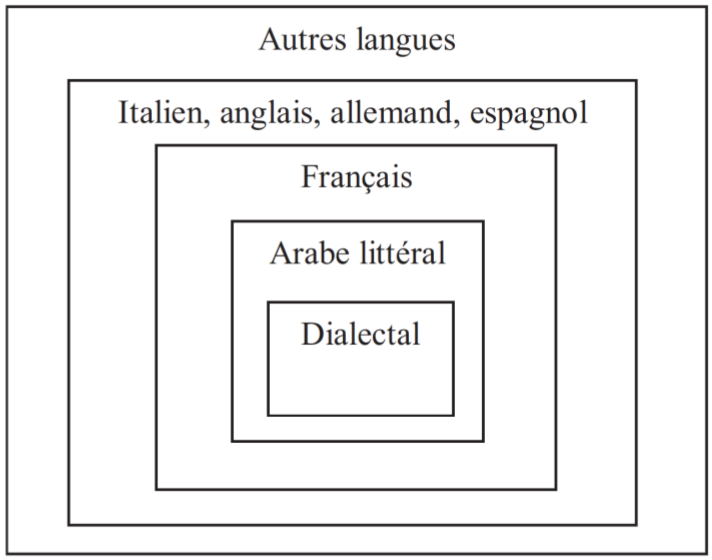
Recent studies have shown that findings from the naturalistic settings cannot be applied to the formal learning contexts due to the differences between the two learning settings in terms of the learning mechanisms and the quantity and the quality of input offered to the learner (Jaekel, Schurig, Florian & Ritter, 2017; Muñoz, 2008). Indeed, students in the formal setting are exposed to a limited number of speakers (teachers) of the target language (TL) in the classroom setting and to a limited amount of L2 (Muñoz, 2006a), whereas the amount of exposure to the TL and the contact with native speakers is unlimited in the naturalistic setting. Therefore, it would be difficult to compare the two settings.

Nevertheless, there have been several misunderstandings about the age factor (Singleton & Pfenninger, 2019) and the advantages found in natural settings were overgeneralized (Muñoz & Singleton, 2011). Added to that, there is a growing demand for proficient L2 speakers for the job market (Jaekel et al., 2017). According to Fraser (2010), pronunciation for L2 learners is "simultaneously the most difficult of the language skills and the one they most aspire to master" (p. 358). All of these reasons have influenced policymakers in different parts of the world to lower the AOA of foreign languages in public schools (de Bot, 2014; Murphy, 2014; Nikolov & Mihaljević Djigunović, 2011, Singleton & Pfenninger, 2022). The rapidly growing wave of early foreign language (FL) programs in elementary schools demonstrates its global appeal and conviction in its educational usefulness (Jaekel, Schurig, van Ackern, & Ritter, 2022)**.** As Jaekel et al. (2022) cited, this trend for lowering AOA is spread in Europe, Asia, South and Central America, and Australia. Concerning North Africa, the Tunisian educational system is not an exception since it has also been influenced by this global wave by decreasing the onset of French L2 instruction in the elementary schools from the third grade (9 years) to the second grade (8 years).

In fact, the Modern Standard Arabic (MSA) is Tunisia's official language according to the orientation legislation n° 2002-80 of July 23, 2002, related to education and school teaching (Official Journal of the Tunisian Republic, 2002). French language is the language of education alongside to the classical Arabic in elementary schools. It is the first FL taught in the Tunisian educational system, followed by English. In high school, students may choose to study German, Italian, Chinese, Turkish, Spanish or Russian as optional courses. In general, the Tunisian educational system favors the multilingual approach (Mejri, Said, & Sfar, 2009). Figure 1 summarizes the Tunisian linguistic system.

Figure 1

The Linguistic System in Tunisia



Note. Adapted from "Pluringuisme et diglossie en Tunisie," by S. Mejri, M. Said, and I. Sfar, 2009, *Synergies Tunisie*. *1*, p. 57.

However, with the relatively recent colonial history (1881-1956), French occupies the first position among the other languages as it has become the dominant language of scientific subjects in schools, university education and administrations (Mejri et al., 2009). It is recognized as co-official language, a foreign language and a privileged foreign language (Kossentini, 2021; Ministère de l'Éducation, 2004). French is used in everyday conversations and the extent to which it is used varies according to the socio-cultural, socio-professional, geographical and affective contexts (Ftita, 2019).

Indeed, French competencies are considered as the vehicle of communication in some economic and educational areas, and they are directly and indirectly required by social demand and employment (Miled, 2007). On the economic level, the proliferation of French-speaking call centers and the employment offers from overseas organizations have resulted in a growing need for the recruitment of proficient French speakers (Miled, 2007). On the social level, the use of French is the prerogative of urban or affluent socioeconomic groups, as well as some specific professional circles (Miled, 2007). Indeed, its use is associated with a high degree of sophistication and prestige (Aouina, 2013; Ballais et al., 2018; Ftita, 2019, Mejri et al., 2009). Recently, there has been a trend on social media in which certain accounts were created with the intention of making fun of influencers who make pronunciation mistakes in French, creating societal pressure to perfect French pronunciation on social media.

As a result of French's particular significance in Tunisian society, teaching French is one of the major axes of the Tunisian educational policy (Kossentini, 2021). Most high school and university subjects are taught in French, with the exception of history, geography, philosophy, Islamic studies and Arabic literature which are taught in standard Arabic (Badwan, 2021). Concerning elementary schools, French occupies an important status.

From Independence (1958) until the present day, French instruction in elementary schools has undergone changes with respect to the age at which it is initiated. From 1993 to 2018, French was introduced in the third grade onwards in elementary schools (age eight) and was granted a significant time slot in the educational system. Table 1 presents the weekly language instruction allotment according to UNESCO – International Bureau of Education (2016).

However, despite the importance of French in the Tunisian educational curriculum, some linguists and employers have observed a continuous decline in the level of linguistic competency for this language (Boukhari, 2006; Langerová, 2012, Miled, 2007). As a result of social and economic pressure, a recent decision has been made in 2019 to lower the starting age of French-language instruction to the second grade (7 years) in public schools (Kossentini, 2021).

Table 1

Number of Hours per Week of Language Training in Tunisian Elementary Schools

Matière	Nombre d'heures par semaine					
	1er degré		2e degré		3e degré	
	1re	2e	3e	4e	5e	6e
Langue arabe	11h	11h	6h	6h	6h	6h
Langue française	–	–	8h	8h	9h	8h
Langue anglaise	–	–	–	–	–	2h

Note. Adapted from *Données mondiales de l'éducation: Septième édition 2010–11* (p. 18), by UNESCO Bureau international d'éducation, 2016. http://www.ibe.unesco.org/fileadmin/user_upload/Publications/WDE/2010/pdf-versions/Tunisia.pdf

Primary schools in the private sector have different curriculum for teaching French L2 in terms of the starting age. French initiation and teaching begin in pre-schools at the age of four or five, with private schools predominately located in the big cities. The private elementary schools are overtaking the public schools in demand, owing to the onset of instruction of French, among other factors. According to the polling institute Sigma Conseil (2018), statistics revealed that between 2010 and 2015, the number of private elementary schools in Tunisia more than doubles, rising from 109 schools to 263 (as cited in Tuniscope, 2018). In 2017, the number of private schools climbed to more than 500.

The worldwide trend toward lowering the starting age of L2 instruction has increased interest in the age effect field during the past few years. As a result, large-scale studies investigating the age effects on the acquisition of various linguistic dimensions have emerged across different European countries: in Spain, the Barcelona Age Factor (BAF) project (García Mayo & García Lecumberri, 2003; Muñoz, 2006a), in Switzerland (Pfenninger & Singleton, 2017a), in Germany (Jaekel et al., 2017) and in North America (Boyson, Semmer, Thompson, & Rosenbusch, 2013). These studies investigated the effects of starting age on different developmental skills in the school setting such as the general oral production (Boyson et al., 2013; Cenoz, 2002, 2003; Miralpeix, 2006; Muñoz, 2003a), fluency (Fullana & MacKay, 2010; Mora, 2006; Torras, Navés, Celaya, & Pérez-Vidal, 2006) and morphosyntax (Larson-Hall, 2008; Mayo, 2003; Muñoz, 2006b; Pfenninger, 2011; Zarobe, 2005).

These studies revealed conflicting results. The outcomes of some studies have reported different results for early starters who surpassed older learners by the end of elementary school (Boyson et al., 2013; Goorhuis-Brouwer & de Bot, 2010; Graham, Courtney, Marinis, & Tonkyn, 2017; Hopp, Vogelbacher, Kieseier, & Thoma, 2019; Larson-Hall, 2008; Mihaljević Djigunović et al., 2008; Oller & Nagato, 1974; Unsworth, Persson, Prins, & de Bot, 2015)

However, other studies reported that later starters progress faster than younger learners and that an early start does not lead to higher proficiency levels nor maintain the advantage of their early start on the long term (Baumert, Fleckenstein, Leucht, Köller, & Möller, 2020; Huang, 2016; Jaekel et al., 2017; Lambelet & Berthele, 2015; Muñoz & Singleton, 2011; Pfenninger & Singleton, 2017a, 2019; Singleton & Pfenninger, 2018). These studies have also shown that late beginners who had less learning exposure time to L2 proved to be equal or superior to early starters and demonstrated quicker development across different linguistic measures (Muñoz, 2006a; Muñoz & Singleton, 2011). Pfenninger and Singleton (2019) even argued that early-start programs are built on the before-mentioned belief: "There is no real dispute about the scientific facts, which are that primary school instruction in L2 fails to equip learners with a level of L2 proficiency which by the end of secondary schooling is superior to that of those whose instruction begins later" (p. 30).

The results of these studies were explained by the explicit L2 input received in the school setting and by the more advanced cognitive maturity, literacy knowledge and the metalinguistic awareness of older learners (Cadierno & Eskildsen, 2018; Muñoz & Singleton, 2011; Nikolov & Mihaljević Djigunović, 2006). Older L2 learners demonstrate higher levels of L2 competency in tasks that require high cognitive abilities such as in morphosyntax. By contrast, the younger learners have less developed cognitive and verbal skills and learn more instinctively and implicitly (Muñoz & Spada, 2018). However, the advantage related to the older learners is no longer relevant when it comes to the communicative skills (particularly oral fluency and pronunciation) which are unrelated to the cognitive abilities (Muñoz, 2006b).

2. Statement of the problem

Despite the extensive research on different linguistic skills, studies examining the age effects of L2 on the acquisition of pronunciation in the school setting are limited. Muñoz and Spada (2018) presented possible explanations for the paucity of research in this area. The first possible explanation is the fact that it is difficult to evaluate the L2 development of learners at a young age where their cognitive and psychological characteristics can impose significant constraints when choosing the appropriate tasks for data collection. Besides, the lack of motivation to investigate the effectiveness of early start programs comes from the fact that policy makers have already taken decisions. Plus, the studies that aim to assess the oral skills require one-on-one assessments, which is time-consuming and costly (Pinter, 2011, p. 211).

These studies that investigated the effects of AOA and L2 pronunciation proficiency in the school setting have yielded inconclusive results ranging from an advantage for late starters (Fullana, 2006; Fullana & MacKay, 2003; Garcia Lecumberri & Gallardo, 2003), to no effects of starting age (Fullana & Mora, 2007; Mackay & Fullana, 2007; Mora & Fullana, 2007). The inconclusiveness of these results could be explained by the fact that all the relevant findings emerged from a limited number of measuring tasks (either repetition tasks, imitation tasks, or storytelling tasks) to assess pronunciation, without taking into consideration the effects they can have on learners' outcome. Indeed, Norris and Ortega (2012) and Saito and Plonsky (2019) confirmed the fact that researchers apply different tasks to assess pronunciation without clear consideration of the impact that different choices in measurement may have on performance or study outcomes. Suzukida and Saito (2019) noted that the impact of the task type has received less attention in the context of L2 pronunciation research.

Added to that, these studies mainly examined accentedness and the production accuracy segments; in addition, there is only a study by Garcia Lecumberri and Gallardo (2003) that investigated intelligibility. In fact, Saito and Plonsky (2019) indicated that L2 pronunciation is a multilayered phenomenon that is composed of the global construct that encompasses comprehensibility, intelligibility and accentedness and the specific construct that englobes the segmental and suprasegmental features of the TL. Furthermore, recent research has stressed the importance of helping students achieve a realistic and achievable goal of L2 pronunciation proficiency in terms of enhanced intelligibility and comprehensibility and to focus on what really matters for the real-life L2 usage (Isaacs, Trofimovich, & Foote, 2018; Saito & Plonsky, 2019).

Consequently, it would be important to investigate AOA and L2 exposure effects on the global and specific constructs of L2 pronunciation.

Besides, all the studies that investigated the AOA effects and L2 pronunciation in the school setting were part of the BAF project which was held in Spain with Spanish/ Catalan learners of English as third language L3. Singleton and Pfenninger (2019, p.114) noted that the degree of distance between L1 and L2 could play a role on diverse aspects of L2 performance with respect to starting age outcomes. Also, Collins and Muñoz (2016) insisted on the need to explore the FL classes for young learners across different geographical sites, beyond the confines of North America and Europe.

To overcome these limitations, and to expand the range of geographical contexts that investigated the age effects, this dissertation examined the effects of AOA and L2 exposure on the global (accentedness, comprehensibility and intelligibility) and specific (segmental and prosodic features) constructs of pronunciation proficiency across different types of tasks (controlled and spontaneous) for young Tunisian learners of French L2.

3. Aims of the study

This dissertation attempts to provide some new insights into the age factor field by investigating the effects of AOA and L2 exposure on the acquisition of French L2 pronunciation by comparing the specific and general pronunciation constructs of three groups of young Tunisians ($N = 68$) that had different starting ages and different amounts of exposure to L2 (G1: AOA 4 years, 1600h exposure L2 / G2: AOA 8 years, 1000h exposure L2 / G3: AOA 4 years, 1000h exposure L2). Moreover, this dissertation aims to explore the effects of task types (controlled vs spontaneous) on the participants' L2 pronunciation outcomes by using four different task types, as will be discussed later.

4. Research questions

The aim of this dissertation is to answer the following research questions:

- Will an earlier starting age result in a better global and specific French L2 pronunciation proficiency across different task types in the school setting?
- Will a longer exposure to French L2 result in a better global and specific L2 pronunciation proficiency across different task types in the school setting?
- Will learners have better results in the controlled or in the spontaneous tasks?

5. Outline of the thesis

The literature review in chapter 2 starts by the examination of the theoretical background and the empirical research for the age effects studies and second language acquisition (SLA) in the first section. Within this section, previous research examining the effects of age and SLA in the naturalistic setting and in the school setting will be presented. Additionally, the studies investigating age effects and L2 pronunciation in the school setting will also be discussed. A second section on L2 pronunciation dimensions is presented. This section covers the main phonological learning theories, and the pronunciation constructs utilized in the thesis. In this section, the global and the specific constructs of L2 pronunciation will be detailed. In addition, the phonemic inventories of Tunisian Arabic and French will be described in detail. The third section will cover task type effects on L2 pronunciation.

Chapter 3 describes the methodology of the study. This chapter begins by presenting the research design and the variables of the study. Then, the characteristics of the speakers and the listeners will be described. Afterwards, the details of the pilot study will be reported. Following that, the four instruments designed to elicit data will be provided. Thereafter, the procedures for data collection, processing and scoring methods will be explained. Finally, the statistical analysis approach will be presented.

Chapter 4 reports the results obtained. Results will be presented in three main sections. The first section provides the statistical results for AOA and L2 exposure according to task type. Section two presents the results for AOA and L2 exposure according to L2 pronunciation constructs. The third section displays the results of task type effects on L2 pronunciation constructs. Lastly, a summary of results is offered.

Chapter 5 aims to provide answers to the research questions by discussing the results reported in chapter 4. The first section of this chapter is concerned with the results obtained for AOA and exposure to L2 effects on pronunciation. The second section discusses the findings of task type effects on L2 pronunciation in terms of AOA and L2 exposure.

Chapter 6 offers a few of concluding remarks about the effects of age of acquisition of L2 pronunciation, exposure to the FL, and task effects will be presented. This chapter also provides the limitations and contribution to knowledge and suggests a number of implications for future research. Finally, a list of bibliographical references and appendices close this dissertation.

Chapter 2

Literature Review

1. Theoretical and empirical underpinnings for age and SLA

1.1. Age effects and SLA in the naturalistic setting

1.1.1. The critical period hypothesis

The age of acquisition has frequently proven to be a relatively robust predictor of the L2 acquisition end state (Abrahamsson & Hyltenstam, 2009; Saito, 2015). It is defined as the chronological age at which the significant exposure to the target language starts (Birdsong, 2006; Flege & MacKay, 2011). According to DeKeyser (2013) "It has often been observed that age effects are stronger in pronunciation than in grammar and barely noticeable in vocabulary; some even hypothesized that age effects were limited to pronunciation" (p. 444).

The effects of age on the acquisition are considered to be one of the most controversial issues in the language acquisition field (Muñoz, 2010). This issue started few decades ago when Penfield and Roberts (1959) formulated the Critical Period Hypothesis (CPH) for the acquisition of the mother tongue (L1) and then was developed for L2 by Lenneberg in 1967. Singleton and Pfenninger (2022) defined the CPH as "the notion that biological aging constrains what is attainable by language acquirers beyond a certain point." (p. 253). This hypothesis claims that age is the main predictor of native-like proficiency of SLA. According to this hypothesis, there is an ideal period to acquire a new language, namely before puberty, during which children can speak the TL with a native accent (Johnson & Newport, 1989; Lenneberg, 1967; Penfield & Roberts, 1959; Selinker, 1972). Penfield and Roberts (1959) stated that this period starts from the age of four and ends at the age of ten. After puberty, language learning becomes increasingly challenging as learners lose their neurological plasticity and as a result, they cannot attain a native-like level of proficiency in many aspects and mainly pronunciation. Lambelet and Berthele (2015) cited that there are four variations of the CPH:

1. Partial L2 development after the critical age: The first version of the hypothesis of a critical period for L2 development is based on the aforementioned concept of L1 holding that language learning after a certain age results only in partial language development (in L1 or L2). With regard to L2, this version of the hypothesis also maintains that students who begin to learn an L2 after the age defined as the end of the critical period are unable to achieve a native-like level of proficiency.
2. Nonlinear decline in ultimate attainment as a function of age of onset: The second variation of the critical period postulates nonlinearity of learning before, during, and after the ages defined as delimiting the critical period.
3. Differences in terms of rate of acquisition: The third variation postulates differences in the speed with which students (early and late) integrate the structures of L2.
4. Neurobiological differences in the brain: The fourth variation of the critical period hypothesis focuses on neurobiological aspects, postulating different cognitive processes and/or differentiated activation of the brain, depending on the age at the start of L2 development. (p.21)

Johnson and Newport (1989) pointed that some scholars referred to ideal period for FL learning as the "sensitive period" because they view the decline in capacity to sustain native-like pronunciation as progressive rather than abrupt. Nevertheless, there was no consensus among researchers about the cut-off point for the critical period i.e., the exact ages that determine the beginning and the end of the critical period. For Lenneberg (1967), the critical period begins at the age two and ends at the age of twelve. Long (1990) suggested that the critical period starts at the age of six and ends at eight. Schwartz (2004) proposes that the critical period is from four to seven and Meisel (2009) believes that the period starts from three to four years old. Granena and Long (2013) suggested that there are three sensitive periods: first for phonology, then for lexis, and finally for syntax.

1.1.2. Empirical findings

Since the inception of the CPH, the majority of studies are retrospective immersion studies in which the participants were typically immigrants who arrived at the hosting country at different ages ranging from infancy to adulthood and acquired the target language to different levels of proficiency (Muñoz, 2019). The CPH and the age effect on language learning were assessed along three dimensions: rate, route, and ultimate attainment. Lambelet and Berthele (2015) stated that it is important to distinguish between the studies that investigates the rate of acquisition and the ultimate attainment as they have different methodological implications. Lambelet and Berthele (2015) defined the ultimate attainment as "the level of proficiency attained by an individual who is deemed to have achieved his or her highest level in the second or foreign language." (p.8).

There are three possible types of scenarios that are employed to investigate the ultimate attainment (Lambelet & Berthele, 2015): First, by comparing the ultimate attainment of learners who started FL learning at different ages. Second, by comparing the ultimate attainment of learners who started FL instruction at different ages but are not grouped according to the age of onset, and age is scrutinized as a continuous variable in this case. Third, by comparing the ultimate attainment of learners who started FL instruction before and after the cut-off point of the critical period to see whether a native-like level can be attained. The tests used to assess the ultimate attainment varied from surveys, formal testing and evaluating the learner's skills through the intuitive judgment of native speakers. The major challenge for researchers is determining the period needed to achieve ultimate attainment in a certain region or a country.

When it comes to the rate of L2 acquisition in relation to AOA, Lambelet and Berthele (2015) cited that this type of studies consisted of testing the linguistic dimensions after a certain amount of time spent in the L2 environment. They also noted that in comparison to the studies that evaluated the ultimate attainment, this type of studies has greater control over the variables and the external factors that may influence favorably or negatively learners' outcome. Nevertheless, the problem with the studies that examined the rate of L2 acquisition is that the groups being compared use the number of hours of exposure to TL as their defining criteria. Consequently, the members of each group are seen as having the same background in terms of time and they differ only in their biological age (Lambelet & Berthele, 2015). To solve the problems with the studies examining the ultimate attainment and the rate of L2 acquisition, some researchers combined the two elements by making sure to evaluate learners of different ages with the same number of hours of exposure (rate of acquisition), as well as learners with the same age but with different amounts of exposure to TL (L2 attainment).

The CPH is confronted with mixed evidence and counterevidence (Singleton & Pfenninger, 2022). Some researchers provided supporting evidence by arguing that there are fundamental and qualitative differences in the quantity of input, i.e., early arrivals are exposed to more input than the late arrivals (Flege & MacKay, 2010). Additionally, earlier arrivals have more "plastic brains" whereas late arrivals progressively lose access to their language-specific cognition system (Saito, 2015). Furthermore, empirical findings demonstrated that learners who arrived after the age of puberty to the host country were not able to gain a native-like pronunciation and that their ultimate attainment patterns are not associated with their AOA profiles (Abrahamsson, 2012; DeKeyser & Larson-Hall, 2005; DeKeyser, Alfi -Shabta, & Ravid, 2010; Scovel, 2000; Stölten, 2005; Verissimo, Heyer, Jacob, & Clahsen 2018). Consequently, AOA is regarded as the key predictor of pronunciation accuracy at L2 ultimate attainment (Hopp & Schmid, 2013).

For example, Abrahamsson and Hyltenstam (2009) examined the oral production of 195 Spanish learners of Swedish L2 with age of arrival varied from under 1 to 47 years old. All the participants identified themselves as having potentially a native-like pronunciation. 20 native speakers of Swedish served as the control group. Abrahamsson and Hyltenstam (2009) evaluated the oral productions of the participants using the intuitive rating of 30 Swedish listeners. The listeners were asked to determine whether the speaker's speech sample was from a native or non-native speaker of Swedish. As a result, only 41 speakers were identified as native speakers of Swedish. This first part of the study served as a screening for the second part of the study. In the second part of the study, using phonetic tools, Abrahamsson and Hyltenstam (2009) investigated the oral production of 41 participants. Among these participants, 31 began learning Swedish between the ages of one and 11. The rest of the participants learned the target language between the age of 12 and 19. Fifteen native speakers of Swedish participated in the study as the control group. Abrahamsson and Hyltenstam (2009) used 20 different tests to assess the oral production of participants. According to the researchers, these tests were highly complex and were designed to differentiate between a native and a non-native speaker of a language. As a result, Abrahamsson and Hyltenstam (2009) found negative correlations between age and nativelikeness. They also found that only some participants who started learning Swedish before 11 were able to attain a native-like level but no participant who started learning the target language after the age of 12 succeeded in all the tests, which supports the CPH.

Granena and Long (2013) also provided supporting evidence for the CPH. They examined the oral production of 65 Chinese learners of Spanish L2. Participants varied in age from 3 to 29 years old when they immigrated, and their length of residence ranged from 8 to 31 years. Participants were divided into three groups according to their age of arrival: 3-6, 7-15, 16-29. Granena and Long (2013) assessed the L2 phonological, lexical, and collocational abilities of the participants, as well as the morphology and syntax. To collect data, participants were asked to read aloud a paragraph, undertook a word completion task and a grammaticality judgment test. Twelve native speakers of Spanish rated the pronunciation of participants using a nine-point Likert scale. As a result, Granena and Long (2013) reported that the group of participants whose age of arrival ranged from 3 to 6 had the best results, followed by those whose age of arrival ranged from 7-15 and ultimately the last group (16-29). Findings also indicated that all participants who arrived after the age of 12 did not have a native-like pronunciation.

Contrastingly, some other researchers expressed reservation with regard to the CPH and questioned its validity by arguing that SLA is biologically based and does not have maturational constraints (Ellis, 1985). Also, the variability in the cut-off age for the termination of the critical period is another element that questions the validity of the CPH (Birdsong, 2006). Furthermore, some researchers presented supportive evidence that some language learners reached an ultimate attainment that is undistinguishable from that of the native speakers and that older learners progress faster in language acquisition due to their cognitive maturity (Birdsong, 2007; Moyer, 1999; van Boxtel, Bongaerts, & Coppen, 2005). As an illustration, Birdsong (2003) investigated the oral production of 22 French learners of English L2. The mean age of the participants when they started learning English is 24.5 years old. Seventeen native speakers of French served as the control group. The participants were asked to read a list of words that contained challenging sounds for French learners of English (p/, / l/, /k/, /i/, /e/, /o/ et /u/). The oral productions were then analyzed in terms of aspiration (VOT) and length of pronunciation of vowels and in terms of the ratings of three judges on a five-point Likert scale. As a result, Birdsong (2003) found that from the 22 participants, 2 speakers were identified as having a near-native pronunciation.

Likewise, Muñoz and Singleton (2007) further confirmed that some learners could attain a native-like pronunciation after the age of puberty. In their research, the authors examined the oral production of 12 Spanish/Catalan learners of English. The participants' mean age when they started learning English was 22.5 years old and their mean age at the moment of testing is 35. Five native speakers of English served as the control group. In order to collect extemporaneous speech, the participants were asked to narrate a story of a movie. The oral productions were presented afterwards to four native speakers of English in order to rate on a 5-point Likert scale the degree of perceived foreign accent. As a result, Muñoz and Singleton (2007) found that two participants were judged as having a native-like accent. As a further step, the authors questioned these two participants to understand how they succeeded in attaining a native-like pronunciation. Muñoz and Singleton (2007) stated that the excessive use of the target language on a daily basis and the high motivation helped the participants to attain such level.

Johnson and Newport (1989) furthered Lenneberg's hypothesis in order to investigate the critical period and SLA in the school setting. Their investigation is considered as a landmark study as it introduced the exercise hypothesis and the maturational state hypothesis. Johnson and Newport (1989) defined the exercise hypothesis as:

> Early in life, humans have a superior capacity for acquiring languages. If the capacity is not exercised during this time, it will disappear or decline with maturation. If the capacity

is exercised, however, further language learning abilities will remain intact throughout life. (p.64)

The maturational state hypothesis cites that humans have a greater potential for language acquisition early in life, and this ability diminishes with maturity. Johnson and Newport (1989) examined the syntactic structures in English of 46 native speakers of Chinese and Korean. The participants were all university students who started learning English L2 for at least five years and lived in the U.S for at least three years before testing. They were divided into two groups: early arrivals who came to the U.S between the ages 3 and 15, and late arrivals who came to the U.S starting from the age of 17. A group of 23 native speakers of English served as a control group. All the participants took a grammaticality judgment test and rated their L2 experience according to a 5-point Likert scale. Results demonstrated that as age of arrival increased, L2 proficiency decreased and that learners who came to the U.S between the ages three and seven had a native-like scores. Johnson and Newport (1989) concluded that age is the most dominant predictor of ultimate L2 proficiency.

Similar AOA effects have also been observed in studies examining phonology (Flege, Yeni-Komshian, & Liu et al., 1999; Huang & Jun, 2011; Munro, Flege, & MacKay, 1996; Oyama, 1976). However, there has been a debate about which linguistic domains that the critical period affects (Huang, 2014). According to Scovel (1988), phonology is the only linguistic dimension that could be affected by the critical period, whereas some other scholars posit that the critical period could affect morphosyntax and semantics (Long, 1990, 2005; Newport, Bavelier, & Neville, 2001).

Birdsong (2018), Huang (2014) and Stölten, Abrahamsson, and Hyltenstam (2014) stated that the discrepancies in the findings of the studies examining CPH and AOA could be the result of the diverse study designs, duration of exposure to TL, and the complex nature of the measuring methodologies, statistical methods, and sampling methods. Acquiring a language is a complex process that necessitates a combination of the learner's linguistic, cognitive, and social talents, along with exposure to quantitatively and qualitatively adequate input (Esteve-Gibert & Muñoz, 2021). Indeed, there are several internal and external factors that could mitigate the ultimate attainment of L2 acquisition (Marinova-Todd, Marshall, & Snow, 2000) such as motivation (Derwing & Munro, 2013; Moyer, 1999), aptitude (DeKeyser et al., 2010), linguistic distance between L1 and L2 (Birdsong & Molis, 2001; Flege, 2003), and the quality and quantity of input (Flege & Liu, 2001).

To summarize, in the natural setting, late learners demonstrate successful FL rate during the first months, after which early starters outperform them and subsequently maintain their advantage. Yet, these findings could not be generalized to the school setting. The following section explores the age effect studies and SLA in the school setting.

1.2. Age effects and SLA in the school setting

During the few past decades, there has been a constant growth of research on the age effects in the instructed L2 research (Rokita-Jaksow & Ellis, 2019). Understanding the age factor is important for SLA research as well as for pedagogy (Singleton & Pfenninger, 2022). In the school setting, the acquisition of L2 phonology is associated with AOA, the quality and amount of input accessible to learners, as well as the type of instruction they encounter (Trofimovich & Gatbonton, 2006). Input is operationalized as the amount of time L2 learners spend practicing the TL both within and outside of FL classrooms (Saito, 2019).

The results of studies conducted in the naturalistic settings demonstrating that young learners could attain native-like proficiency level were often mistakenly generalized to the school setting (Dekeyser & Larson-Hall, 2005). Applying the results from the naturalistic setting to early foreign language instruction is not appropriate due to the difference between the two environments (Singleton & Muñoz, 2011). Immersion programs and naturalistic settings provide significant amounts of L2 interaction compared to the structured and limited quantity of input that schools provide (Muñoz, 2006b, 2008).

1.2.1. Studies with positive impact of early L2 start programs

Some studies demonstarted that there is indeed a positive effect for implementing early foreign language education (Goorhuis-Brouwer & de Bot, 2010; Graham et al., 2017; Hopp et al., 2019; Oller & Nagato, 1974; Unsworth et al., 2015). Baumert et al. (2020) even stated that "There is thus no doubting the effectiveness of early-start programs at the point of leaving elementary school." (p. 1094).

One of the earliest studies that provided evidence of positive effects of early start programs is that of Oller and Nagato (1974). The researchers examined the English proficiency of 223 student girls at a secondary school in Japan. The group of early starters began learning English at the age of 6 (year 1 of elementary school) and the group of late starters learned English at the age of 13 (year 7 of secondary school). From year 8 of secondary school, early and late learners had joint English classes. Tests were administered at the years 7, 9, and 11. Oller and Nagato (1974) found that the gap between early and late learners decreased over the years and that an early start had a long-term positive effect on learners' proficiency. This finding is consistent with the results of Larson-Hall (2008).

In her study, Larson-Hall (2008) examined the performance of 200 Japanese university students of English who began learning English at ages 8 and 12.5. The purpose of the study was to examine the effects of age and exposure to the TL on the long run. Participants were divided into two groups: those who started studying English between the ages 3 and 12 and had had 1923 hours of exposure to the target language and those who started learning at the age of 12/13 and had had 1764 hours of exposure. Participants were tested at the age of 19. Larson-Hall (2008) administered a phonemic discrimination task, a grammaticality judgment task and an aptitude task. The researcher found that earlier starters obtained higher scores on the phonemic but not the morphosyntactic measures. Larson-Hall (2008) concluded her study with the following statement:

> Because this study is in the minority of studies in that it found a positive effect for a younger starting age in a situation of minimal input, readers should be cautious in interpreting it too broadly. However, my recommendation would be to begin foreign language study as young as possible, with as many hours of input as are possible. (p.59)

Another study that demonstrated a positive effect in implementing early foreign language teaching is the study by Mihaljevic Djigunovic, Nikolov, and Otto (2008) who scrutinized the general English proficiency (listening, reading, and writing skills) of 717 Croatian and Hungarian learners who started learning the TL at two different ages: 10 years old and before 10 years old. Participants' mean age at testing was 14 years old. A reading task, a listening task, and a writing task were used to collect data. Mihaljevic Djigunovic et al. (2008) found moderate to strong correlations between AOA and test scores and concluded that early starters performed significantly better on the English foreign language proficiency tests than late starters.

These results were further confirmed by the study of Boyson et al. (2013). In the 5-year longitudinal study of Boyson et al. (2013), researchers examined the oral and listening comprehension proficiency levels of 120 English students who started learning Spanish at two different ages (5/6 years old vs 10/11 years old). The oral proficiency tests were administered at the end of grades 5 and 8. As a result, Boyson et al. (2013) found that students who began in kindergarten had statistically higher proficiency levels than those who began in Grade 5. In the conclusion, Boyson et al. (2013) strongly supported early foreign language learning from the kindergarten "The study provides empirical data to support what has often previously been assumed: that younger students will, in the long term, attain higher proficiency levels than those who begin later" (p. 259).

The findings in Hopp et al. (2019) also revealed that there is a benefit for early L2 instruction. The researchers examined the fluency of 23 early starters and 29 late starters. Participants' age at testing was 12 years old in the final grade of a Dutch elementary school. A picture description task was used to collect data and the speech samples were analyzed using PRAAT. Hopp et al. (2019) found that there is an advantage for early starters in terms of oral fluency.

1.2.2. Studies with no positive impact of early L2 start programs

Some other scholars demonstrated different results where older learners surpass the younger learners since they learn faster (Baumert et al., 2020; de Bot, 2014; Huang, 2016; Jaekel et al., 2017; Lambelet & Berthele, 2015; Muñoz, 2008; Muñoz & Singleton, 2011; Pfenninger & Singleton, 2017b, 2019; Singleton & Pfenninger, 2018). Older learners progress faster than young learners due to their cognitive maturation and advanced literacy skills (Jaekel et al., 2017; Pfenninger, 2014).

One of the earliest studies that presented evidence of positive effects of late start programs is that of Burstall, Jamieson, Cohen, and Hargreaves (1974). The researchers examined the effects of early French instruction in England and Wales. Burstall et al. (1974) examined the proficiency (no specifics were given) of three cohorts of participants who started learning French at the ages of 8 and 11. The total number of participants was 17,000. The results of the several tests administered at the end of the compulsory education in elementary school indicated that the early FL instruction program had only a marginally positive effect on listening comprehension, whereas late starters performed better on the other tests in unspecified dimensions. Burstall et al. (1974) concluded that there were no long-term benefits to start L2 instruction from the early grades when the overall amount of exposure to TL is relatively limited.

In a more recent study, Jaekel et al. (2017) investigated the proficiency of the early receptive skills of German-speaking students learners of English L2 in the North Rhine-Westphalia elementary school, where there was a shift in the onset of English from year 3 to year 1. Participants were divided into two cohorts: early starters with AOA 5 years old (N = 2,468) and a control group with AOA 8 years old (N = 2,632). The factors that distinguished between the two groups were the AOA of the participants and the amount of language exposure. Participants were assessed in year 5 and two years later in year 7. Jaekel et al. (2017) found that the effects of AOA were noticeable in year where early starters outperformed their peers in reading and listening comprehension with effect sizes $d = 0.28$ and $d = 0.34$, respectively. Yet, in year 7, older learners had better results than the younger learners with effect sizes of $d = -0.35$ in reading and $d = -0.17$ in listening comprehension. These results suggest that there is a disadvantage for a late start, which is in accordance with research with large-scale studies.

The evidence for "the later = the better" came from two major projects: the BAF project in Spain in the early 2000s and the Beyond Age Effects project (BAE) in Switzerland from 2008 to 2016. The BAF project is one of the most significant longitudinal large-scale experiments that investigated the age effects issue (Muñoz, 2006a). The study took place from 1999 to 2002 when the Catalan educational system underwent a change where students begun learning English at the year 3 instead of the year 6 (Muñoz, 2006b). The study followed the progress of language development of bilingual Spanish-Catalan learners of English L3 with different AOAs: at the age of 8 (N = 164) and at the age of 11 (N = 107) years, 14 and +18. Participants were tested after 200, 416, and 726 hours of instruction of English. Learners were tested on different linguistic skills: written skills (e.g., Navés, Torras, & Celaya, 2003; e.g., Torras & Celaya, 2001), verbal proficiency (e.g., Muñoz, 2003b; Mora, 2006), morphosyntax (Alvarez, 2006), and both verbal and written proficiency (e.g., Miralpeix, 2006)

Participants underwent a battery of tests ranging from grammaticality judgment tests, Cloze tests and interviews. After 200 and 416 hours of instruction, late learners demonstrated faster rates and outperformed the earlier learners in almost all language skills. After 726 hours of instruction, there was no significant disparities between the older groups (+ 11 years old) and the advantage for late starters began either to diminish or remained constant depending on the linguistic dimension in question (Muñoz, 2006c). Yet, the youngest group of learners (AOA= 8) lagged behind the older groups without surpassing them. Muñoz (2006a) concluded that "age effects are not uniform across measures of language abilities" (p. 30).

To cite an example of a study from the BAF project, Cenoz (2003) examined the oral and written proficiency levels of 135 Spanish/ Catalan students who started learning English at three different ages 4, 8 and 11. Their age at testing was 10, 13 and 16 and they all had 600 hours of instruction of English. After the same number of hours of exposure, Cenoz (2003) found that older learners had better results in comparison with the younger learners.

Likewise, Muñoz (2003b) compared the oral production of 136 Spanish/ Catalan students who started their L3 at the age of 8 and the oral production of 104 students whose AOA was 11. Muñoz (2003b) tested the participants after having the same number of hours of exposure to the target language (T1: after 200 hours and T2: after 416 hours) and when they all had the same age (12 years) at the moment of the testing. The researcher found that older learners had better results than younger learners. These results were also confirmed by the study of Mora (2006) who analyzed the oral production and fluency of 30 Spanish/ Catalan students whose AOA was 8 and 30 Spanish/ Catalan learners of English whose AOA was 11. After having the same number of hours of instruction of English (726 hours), the older learners had statistically better results than the younger learners.

The second large-scale project is the BAE study conducted by Pfenninger and Singleton (2017b) in the German speaking Swiss canton of Zurich. The project took place when a shift in the educational program of elementary schools in Zurich in the year 2004/2005 was initiated, where the onset of instruction of English was lowered from year 7 (AOA = 13) to year 2 (AOA = 8). Participants were tested in the middle of year 7 and at the end of year 12. The main goals of the project were to identify the factors that hinder early learners from benefiting from the extended exposure to TL and to understand the mechanisms that allow late learners to catch up with their counterparts. The cross-sectional data were collected from 636 students and the longitudinal data from almost 200 students. Pfenninger and Singleton (2017b) found that there was no evidence for an advantage of an early start of FL teaching where in the first year, early starters demonstrated better results in the lexical knowledge but after six months, later starters caught up and occasionally surpassed them in some domains. At the end of the secondary school, all age disparities between the two groups vanished.

1.2.3. Long-term effects of early start programs

Another important issue that scholars have attempted to solve regarding the long-term effects of early L2 exposure can be summarized as following:

> If late starters display a faster rate of learning than early beginners, do the early starters catch up to and surpass the late starters by after a certain number of hours of instruction, as is the case in natural contexts? (Lambelet & Berthele, 2015, p. 62)

To answer this question, various strands of research were interested in the long-term effects of early foreign language instruction in elementary schools (Baumert et al., 2020). The results of some empirical studies have shown that students who began learning an L2 at an early age do not maintain on the long run the advantage of their early start (Muñoz, 2003a, 2014) and that late starters prove to be equal or even superior to the early starters (Muñoz & Singleton, 2011). For example, Muñoz (2014) compared the oral performance of 160 Spanish learners of English whose AOA of English ranged between 3 and 15 years old. Data were collected using a film retelling task and were then analyzed in terms of fluency, lexical diversity, and syntactic complexity. Muñoz (2014) concluded that age did not correlate positively with the oral performance and that several measures of input (e.g., the number of hours of instruction or the frequency of contact with the target language) are better predictors of learners' proficiency and oral performance in the target language than AOA.

Another recent study that investigated the long-term effects of early vs late FL instruction is that of Baumert et al. (2020). This cross-sectional study was administered on the national level of Germany. Baumert et al. (2020) assessed the receptive skills (reading and listening comprehension) of a random sample of 19,858 students learning English as an L2. The participants were divided into three groups: early starters (from year 1), middle starters (from year 3) and late starters (from year 5). Participants were tested when they were at year 9 and the data were collected from 1,431 classes. Baumert et al. (2020) found that by year 9, there was a light difference in the performance of the three groups. The researchers found that there was no evidence for an early start after five years of instruction in the secondary school.

Lambelet and Berthele (2015) explained that an early FL start does not provide a long-term advantage with regard to the level of proficiency attained because of the number of compulsory hours of FL instruction that do not usually exceed 800 hours, which is far less than the number of hours required for early learners to compensate for their initial delay and catch up to adults in natural contexts. Consequently, Lambelet and Berthele (2015) highly recommended more research examining AOA on the impact of increased number of hours of instruction of FL in the school setting.

1.3. Age effects and pronunciation in the school setting

Despite the significance of pronunciation mastering for FL learners, only a limited number of studies addressing the effects of age and pronunciation were identified. Indeed, Thomson and Derwing (2014) noted that pronunciation has not received sufficient attention in instructed SLA settings. According to Lambelet and Brethele (2015), the majority of research investigating AOA and phonology focused either on learners' ability to distinguish between L2 phonemes or L2 accentedness. The findings of these studies suggest that late learners have better pronunciation proficiency levels over their younger counterparts, whereas others found that neither AOA nor exposure to TL has any effect on learners' performance.

For example, the subproject of the BAF project conducted by Fullana (2006) evaluated the oral perception and production of Spanish/ Catalan participants. Participants were divided into two groups: the first group had different AOAs (8, 11, 14, and + 18) and the second group had different amounts of exposure to English (200 hours, 416 hours, and 726 hours). A phonemic discrimination task of 13 contrast minimal pairs and an imitation task were administered to the participants. The study focused on seven problematic segments (i, ɪ, ɛ, æ, ɒ, u, ʌ) for Spanish learners of English. The words with the target segments were evaluated in terms of foreign accent. Results revealed that there is an effect of AOA and exposure on participants' proficiency levels. The results for the phonemic discrimination task demonstrated that early learners obtained lower scores than the older groups after 200 and 416 hours. Yet, after 726 hours, early starters caught up and became slightly better than the other groups. Similarly, when it comes to the production task, the younger learners with an AOA of 8 years were rated as less accented after 200 and 416 hours and after 726 hours, 11-year-old starters were found to be less accented. Fullana (2006) concluded that, unlike perception, AOA and exposure to TL were not a conclusive determinant for attaining native-like pronunciation production proficiency.

Fullana and MacKay (2003) found similar results by examining the pronunciation of /i/ and /ɪ/ by 135 Spanish/ Catalan learners of English L3. Being part of the BAF project, the participants in this study were the same as in the previous research of Fullana (2006). The participants were divided into groups according to their starting age (8, 11, 14, and + 18) and the amount of exposure to the TL (2.5 years, 4.5 years, and 7.5 years that corresponded to 200 hours, 415 hours and 726 hours of formal instruction, respectively). Participants performed an imitation task and raters evaluated participants' production of words with the target vowels in terms of accentedness using a nine-point Likert scale. Results revealed that the listeners tended to rate older learners' productions of /i/ and / ɪ / higher than younger learners.

Better scores by participants who started learning English at an advanced age were also found in the study of Garcia Lecumberri and Gallardo (2003). The scholars compared the speech perception and the phonological production of Spanish/ Catalan students who began learning English at the age of four with two other groups with different AOAs (8 and 11 years). The participants (n= 60) were also exposed to the same amount of English (6 years) during their compulsory education. The learners were 9–11, 13–15, and 16–18 years old at the moment of testing. Participants undertook a phonological discrimination test for /i/ and /ɪ/ and a storytelling task. Participants' extemporaneous productions were evaluated by native speakers in terms of the degree of foreign accent and intelligibility by using a 9-point Likert scale. As a result, after having the same number of hours of exposure to English, later starters were found to be more intelligible and to have less pronounced foreign accent than those who started learning English at the age of 4. The same results were also observed for the phonemic discrimination task.

On the other hand, MacKay and Fullana (2007) found neither AOA nor exposure had a significant effect on participants' pronunciation. The scholars examined the accentedness of 148 Catalan/Spanish learners of English that had different AOAs (8, 11, 14, and 18+ years) and varying amounts of exposure to the TL (2.5, 4.5, and 7.5 years). A word repetition task was administered to collect data. Seven Canadian listeners evaluated the production of English vowels /i ɪ æ u/ in terms of foreign accent by using a 9-point Likert scale. They were also asked to identify the vowel sounds that participants had produced. Results demonstrated that there were no positive effects for starting age nor for exposure to the TL on the accentedness ratings and vowel identification scores.

In another study, Fullana and Mora (2007) aimed at investigating the perception and production of English final-word obstruents (/s/-/z/, /p/-/b/, and /t/-/d/) of 49 Catalan-Spanish learners of English L3. Participants were divided into two groups according to their starting age (before and after 8 years-old) and experience in the TL (amount of formal instruction and extracurricular exposure). Data of segmental production were collected through a categorical AXB discrimination task and a delayed sentence repetition task using a delayed repetition technique. To analyse data, Fullana and Mora (2007) administered acoustic analyses with PRAAT to measure the accuracy of final-word obstruents pronunciation and to measure perception, the mean percent of correct discrimination scores was calculated. Researchers found that neither age nor experience / exposure had a significant effect on discrimination scores. For segmental production, earlier starters produced more native-like pronunciation. Yet, researchers concluded that AOA of exposure to TL failed to determine the perception and production proficiency levels of English final-word obstruents for Catalan/ Spanish bilinguals.

These results were further confirmed by Mora and Fullana (2007) who examined the perception and production of two vowel contrasts (/iː/-/ɪ/ and /ʌ/- /æ/) of 49 Catalan/Spanish learners of English L3 that had different AOA and varying FL experience in the school setting. Mora and Fullana (2007) used a delayed sentence repetition task to collect the oral productions and then run acoustical analyses using PRAAT. Although a late-starting age advantage was observed for the vowel discrimination task, results revealed that neither AOA nor exposure to English had a significant effect on the participants' vowel production.

These studies examining AOA and exposure effects regarding the pronunciation proficiency did not find and unequivocal evidence for a younger learner advantage. Yet, the findings of these studies were obtained by analyzing two pronunciation dimensions (intelligibility and segmental production in terms of accentedness) and by using two tasks to collect data (storytelling and repetition tasks). Nevertheless, Lambelet and Berthele (2015) noted that the effects of the task type on the findings is worth considering, especially in studies examining the rate of FL learning. Additionally, pronunciation is a multilayered phenomenon that is composed of global and specific constructs. Hence, it would be interesting to examine the effects of AOA and exposure to TL on the global and specific constructs of pronunciation while taking into consideration the possible effects of task types. Furthermore, considering these above-mentioned studies employed data from the one large project with homogeneous populations (Spanish, Basque, or Catalan speakers), it is uncertain if the findings can be generalized to learners with other L1s (Huang, 2016). In fact, Nagle (2022) stated most L2 pronunciation studies focused on English as the target language. However, it would be interesting to explore if the findings from these studies could be extended to different contexts with participants with different L1s and L2s other than English.

The sections that follow will present the dimensions of L2 pronunciation and discuss the effects of task on pronunciation assessment.

2. Dimensions of L2 pronunciation

The acquisition of L2 phonology is a complex process that requires a thorough understanding of how learners acquire a new phonological system with taking into consideration the differences between L1 and TL. Yet, the capacity to properly communicate in L2 has become increasingly important as encounters across cultural and linguistic barriers became increasingly common (Bergeron & Trofimovich, 2017). L2 pronunciation is considered as a part of global L2 oral proficiency (Saito & Plonsky, 2019). Pronunciation is very salient to listeners and speakers. It is a fundamental part of communication as it is the first stop for listeners to detect the small deviations of vowels and consonants and form social judgements based on these deviances (Levis, 2018a, p. 22). Pronunciation could be an indicator of the speaker's social status, educational level, and regional background (Derwing & Munro, 2022). The mispronunciations could even have a strong influence on future opportunities for interaction (Derwing & Munro, 2013) and even employability and promotion opportunities (Newton, 2018).

Pronunciation is defined according to Derwing and Munro (2015, p. 2) as "the way in which speakers use their articulatory apparatus to create speech". Richards and Schmidt (2010) provided another definition of pronunciation as "the way a certain sound or sounds are produced. Unlike ARTICULATION, which refers to the actual production of speech sounds in the mouth, pronunciation stresses more the way sounds are perceived by the hearer" (original emphasis, p. 469). For Levis (2018a), pronunciation is an "interrelated system of sounds and prosody that communicates meaning through categorical contrasts (e.g., phonemes), systematic variations (e.g., allophones), and individual variations that may mark gradient differences such as gender, age, origin, etc." (p. 35).

Despite the importance of pronunciation, there are no dominant theoretical framework and models (Foote & Trofimovich, 2018, p. 75). Few pronunciation researchers framed their studies within a theoretical framework because the existing theories are either too broad or too restricted in scope to perform a prominent role as a viewpoint on pronunciation development. Furthermore, Foote and Trofimovich (2018) added that SLA theories have seldom been applied to pronunciation, thus few pronunciation researchers farmed their studies within a specific theoretical framework. Consequently, despite the evident relevance of pronunciation, L2 pronunciation research and phonological development have been historically underrepresented and has received less attention in comparison to the other areas of SLA research (Derwing & Munro, 2005; Foote, 2015; Parlak, 2018).

This section presents the phonological learning hypotheses and models as they pave the way to a better understanding of the mechanisms of perception and the production of L2. Next, the constructs of L2 pronunciation will be discussed. In this subsection, an overview of the

Tunisian Arabic and French linguistic systems will be provided. Finally, the pronunciation assessment methods and task types will be explored.

2.1. Phonological learning theories

2.1.1. Contrastive analysis hypothesis

The Contrastive analysis hypothesis (CAH) was formulated by Lado (1957) based on the behaviorist tradition. This notion holds that the difference between L1 and TL is necessary and sufficient to account for the difficulties / pronunciation errors that L2 learners can face. The following quote embodies this viewpoint:

> We assume that the student who comes in contact with a foreign language will find some features of it quite easy and others extremely difficult. Those elements that are similar to his native language will be simple for him, and those elements that are different will be difficult. (Lado, 1957, p. 2)

Similarity means that L1 and TL are close in their language family relationship where some features of the grammatical, writing, or sound systems are similar (Aldurayheem, 2020). CAH suggests that the closer a phonemic inventory of the TL is to L1, the easier it is to perceive and to produce the TL. Hence, the new sounds are difficult to learn.

For example, a contrastive analysis between the Arabic and French would assume that Arabic speakers would face difficulties in producing the voiceless bilabial stop /p/ since that bilabial stop is not included in the Arabic sound inventory. Therefore, according to the CAH, the Arabic speakers are predicted to present difficulties with /p/, resulting in pronunciation errors. Much of the previous pronunciation research focused on comparing diverse L1 and L2 phonological inventories point-by-point in an attempt to forecast probable learners' mistakes.

However, by the early 1970s, support for the CAH began to erode (Mehrdad & Ahgbar, 2015). Many researchers questioned the feasibility and the usefulness of the CAH as it lacked error predictability (Whitman & Jackson, 1972). Some predicted errors did not occur and some errors that were not predicted occurred. To cite an example, an Arabic learner of English could successfully pronounce the English fricative /ʒ/, which does not exist in Arabic. Just as importantly, the CAH did not take into consideration the possibility that there could be differing levels of difficulty due to the intermediating variables (Parlak, 2018) such as the learners' performance, attention, etc. Added to that, the similar sounds in two languages might lead to confusion due to the differences at the phonetic level where the language learner might use the similar form correctly or overuse it. To cite an example, an English learner of French could make confusions between the sound /u/ in 'vous' and the sound /y/ in 'vu' which is a front rounded sound in French and back-rounded sound in English. Finally, the contrastive analysis approach assumes that L2 pronunciation errors are attributed to faulty articulation rather than to incorrect sounds caused by inaccurate perception (Flege & Bohn, 2021)

With regards to these flaws in the CAH, further explanations were needed to explain and detect the difficulties for L2 learners. To address these inadequacies, Eckman (1977) suggested the markedness differential hypothesis.

2.1.2. Markedness differential hypothesis

Major (2008) indicated that "[t]ransfer should not be studied as an isolated phenomenon, but rather only in relationship to other factors, including Markedness, similarity/dissimilarity, and other universals" (p. 82). The term "markedness" was proposed by the Prague School of Linguistics phonologists Nikolai Trubetzkoy (1939) and Roman Jakobson (cited in Eckman, 1996). This notion refers to the relationship between linguist elements and considers that certain linguistic elements can be seen as marked, i.e., complex, exceptional, or difficult while other elements can be seen as unmarked, i.e., simple, common, or easy to produce. The unmarked forms are assumed to be easier to acquire and consequently easier to produce while the marked forms are harder to acquire and to produce. For example, Durand and Prince (2015) reported that from research on the vocalic systems throughout the world, the nasal vowels are rarer and more complex than the oral vowels where their production involves a supplementary resonator, hence they are regarded as marked. The markedness value can be determined by a variety of factors such as the phonetic factors (the more elements, the more complex), the cross linguistic frequency distributuion of the sound types, etc. (Alqattan, 2015)

The markedness differential hypothesis (MDH) was introduced by Eckman (1977) as an attempt to address some of the problems with the CAH. Eckman (1985) suggested that the markedness continuum was not enough but rather the markedness differential. The MDH states "The areas of difficulty that a language learner will have can be predicted on the basis of a systematic comparison of the grammars of the native language, the target language and the markedness relations stated in universal grammar" (Eckman, 1977, p. 321). The areas in the TL that are marked compared to the L1 will be more difficult to acquire than the unmarked ones. As a result, not all the differences between the TL and the L1 will cause equal difficulty, but the relative degree of difficulty of the more marked TL forms corresponds to the relative degree of Markedness (Eckman, 2013).

For example, Dinnsen and Eckman (1975) (cited in Gass & Selinker, 2008) proposed a Markedness Differential Hierarchy for voicing contrasts. Table 2 presents their proposed hierarchy for voicing contrasts from the less marked or frequent position to the most marked or less frequent position. This hierarchy reflects that the contrast of voicing in the initial position is unmarked and that the contrast in the final position is the most marked. The suggested hierarchy also implies that a speaker of Arabic with a more marked TL structure (voicing in final position) will have no difficulty to produce Catalan words with no voicing contrast in the final position (less marked feature than L1) whereas a Catalan learner of Arabic is expected to face difficulties when learning the contrast in the final position in Arabic words (a more marked structure than Catalan).

Table 2

Markedness Differential Hierarchy of The Voicing Contrasts

Description	*Languages*	
Languages that maintain a superficial voice contrast in initial, medial, and final positions	English, Arabic, Swedish	More frequent
Languages that maintain a superficial voice contrast in initial and medial positions, but fail to maintain this contrast in final position	German, Polish, Greek, Japanese, Catalan	↓
Languages that maintain a superficial voice contrast in initial position, but fail to maintain this contrast in medial and final positions	Corsican, Sardinian	↓
Languages that maintain no voice contrast in initial, medial, or final positions	Korean	Less frequent

Note. From *Second Language Acquisition. An Introductory Course. Third Edition* (p. 180) by S. Gass & L. Selinker, 2008, Routledge.

Yet, Archibald (2021) was contemplating whether it is the markedness effect or the frequency effect that contributes to the success of the acquisition of the new forms. In other words, do learners succeed in acquiring the new TL forms because these forms are more frequent or because they are less marked. Cardoso (2007) investigated the effects of markedness and input frequency on the acquisition of L2 English consonant clusters ([st], [sn] and [sl]) by L1 speakers of Brazilian Portuguese. These clusters are the most frequent structures and the most marked as well. According to the markedness hypothesis, the participants are predicted to acquire the clusters from the least marked to the most marked in the following order: /sl/ > /sn/ > /st/. Yet, according to the frequency hypothesis, the developmental order should be the following: /st/ > /sl/ > /sn/. As a result, Cardoso (2007) found that [st] is the most frequent and the most marked cluster and that learners were less accurate on the most clusters. Cardoso (2007) concluded that markedness was more explanatory than frequency.

The MDH has been criticized on the conceptual and empirical levels. Some researchers argue that markedness is too broad and makes predictions about universal patterns (Hume, 2011). MDH has also been criticized as it does not match what learners really generate (Altenberg & Vago, 1983; Stockman & Pluut, 1992, as cited in Cadette-Blasse, 2006)

2.1.3. Speech learning model

The previous hypotheses of CAH and MDH had major shortcomings, as they treated the phonemes as a set of objects to be analyzed, thus ignoring learner-related factors. Consequently, the speech learning model (SLM) was developed by Flege (1995), providing a more elaborated explanation for L2 sound production difficulties/ errors as it is concerned exclusively with L2 phonology and specifically with age of exposure and the ultimate attainment of the production of L2 sounds. The SLM is intended to explain a variety of phenomena.

First, SLM predicted that perception precedes production, and that production and perception will converge over the course of learning. L2 learners will form auditory equivalence classes derived from the elaboration of new phonetic L2 categories to which they have been exposed (Flege & Bohn, 2021). SLM proposed that L2 learners at any age develop equivalence classes. This means that learners classify the phonemes TL as "new' and 'similar' through the L1 phonological system. The new phonemes that are distinct from L1 are established into a new category (dissimilation), while the similar sounds are classified as phonetic realizations of their corresponding L1 counterparts (assimilation). This equivalence classification is considered as a mapping for the TL phonemes onto the L1. Still, some sounds are identical or near identical to their L1 counterparts as they are perceived and produced in terms of the L1 category (e.g., [b] in French and Arabic) (Cebrian, Gorba, & Gavaldà, 2021; Flege, 1992)

This classification of L2 sounds to 'new', 'similar' and 'identical' was based on three factors: the transcription in International Phonetic Alphabet (IPA), the acoustic proximity of TL and L1 and the perceptual similarity with L1. The new sounds are classified as identical to L1 when they have the same IPA transcription, close acoustic properties and there is no discernable perceptual difference. The L2 sounds are considered as similar to their L1 counterparts when they have the same IPA symbol but different acoustic and perceptual features. Finally, the L2 sounds are regarded as new by the learners when the IPA symbol is different and is acoustically and perceptibly different from the L1 phonemes.

In contrast to the CAH and the MDH, the SLM (Flege, 1995, 1999) predicts that new sounds will be difficult to acquire but successfully mastered, and that the similar sounds will be the most difficult to acquire and will be produced with an accent, i.e., difficult to master (Munro et al., 1996). This SLM prediction has been confirmed in a study by Busà (1992), who examined the production of English /u/ and /ʊ/ vowels by Italians living in the United States. Busà (1992) found through an acoustic analysis that most of the participants produced English /uː/, classified as similar to Italian /u/, with lower F2 values than those of the native English speakers (that is, closer to Italian /u/). By contrast, a larger number of Italians produced the new L2 vowel /ʊ/ more accurately with F1 and F2 values than was the case of the vowel /u/.

Moreover, the SLM stipulates that the mechanisms for learning L1 sounds remain active during the life span, and they continually adjust to accommodate all the new learned sounds. This, however, does not mean that learners (children and adults) could attain a native-like proficiency level. The SLM posits that the ability to produce and master new L2 sounds is not confined to a single element, i.e., the biological age at which the learning process started but rather a combination of factors that involve the nature and amount of L2 exposure, L1-TL usage, and the similarities between L1 and L2. In fact, the increased L2 exposure can improve the ability to discern between L1 and TL categories. As the L2 experience increases, the age of exposure decreases, and as result, the likelihood of becoming a native-like increases when the age of exposure decreases. As an illustration, Flege and Eefting (1987) found that Spanish learners of English L2 whose age of exposure began at an early stage of their lives succeeded in producing an English /t/ with VOT values that are comparable to that of the native English speakers. The same result was confirmed by Jun and Cowie (1994) who compared the production of the new sound /ɪ/ by Korean learners of English. As a result, Jun and Cowie (1994) found that the experienced Korean learners with a length of residence of 16-31 years produced more accurately the target sound than the inexperienced learners with a length of residence of 1.3 - 5.3 years.

The overarching aim of the SLM (Flege, 1995) provided a clearer understanding of the relationship between AOA, experiential variables such as quantity and quality of L2 input, and

L2 pronunciation attainment of highly proficient L2 users. In sum, the Flege (1995) offered four main postulates to form the SLM:

P1- The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning.

P2- Language-specific aspects of speech sounds are specified in long-term memory representations called *phonetic categories*.

P3- Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as the realization of each category.

P4- Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space. (p. 239)

The SLM also proposed seven hypotheses to account for the various settings of L2 acquisition based on the results of several studies (e.g., Flege, 1993; Flege Munro, & MacKay, 1995; Sheldon & Strange, 1982; Takagi, 1993):

H1: Sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level.

H2: A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds.

H3: The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned.

H4: The likelihood of phonetic differences between L1 and L2 sounds, and between L2 sounds that are noncontrastive in the L1, being discerned decreases as AOL increases.

H5: Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones). Eventually, the diaphones will resemble one another in production.

H6: The phonetic category established for L2 sounds by a bilingual may differ from a monolingual's if: 1) the bilingual's category is 'deflected' away from an L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; or 2) the bilingual's representation is based on different features, or feature weights, than a monolingual's.

H7: The production of a sound eventually corresponds to the properties represented in its phonetic category representation. (Flege, 1995, p. 239)

2.1.4. Perceptual assimilation model

Best (1994, 1995) expanded on the earlier SLM hypothesis of Flege (1995) by introducing the perceptual assimilation model (PAM). This model concerns the L2 phonological perception mapping of non-native sounds for naïve L2 learners. PAM theorises that, at early stages of L2 acquisition, learners perceive and categorize non-native sounds in relation to their similarities and dissimilarities to their native phones in the L1 phonological space, i.e., phonetic contrast between L1 and L2 phonemes. When a new sound is detected by the novice learner, the learner will compare it with the existing closest exemplars in L1 and then he will decide how it is interpreted. In other terms, novice listeners have difficulty to categorize and discriminate phonemes in non-native languages, especially when the contrasts needed for classification do not exist in their L1.

PAM proposes three perceptual assimilation patterns (Best, 1995). The first pattern states that L2 sounds can be assimilated to the listeners' L1 sounds as good, acceptable, or deviant. The second perceptual pattern indicates that L2 sounds can be assimilated to L1 sounds as uncategorizable which means that the L2 sounds are recognized as speech sounds but do not fall within the phonological space of the L1. The third perceptual pattern implies that L2 sounds are not assimilable where they are acoustically similar to any L1 sounds and consequently, they are not categorized as speech sounds (Best, McRoberts, & Goodell, 2001).

To further illustrate the PAM mechanisms, Best (1994, 1995) suggested six different types to assimilate patterns for L2 contrasts depending on the degree of perceptual similarities and dissimilarities to the L1 exemplars. The first is the "two-category assimilation" where two L2 speech sounds are assigned into two different L1 categories and as consequence, the discrimination of the L2 phones is predicted as excellent and will be perceived as good exemplars of L1 phonemes.

The second is "category-goodness difference" designates that two non-native sounds are assimilated to the same L1 phonological category where one of them is considered as an acceptable exemplar and the other as bad or unacceptable exemplar. The discrimination in this second pattern is projected to be moderate-to-good.

The third pattern is the "single-category assimilation" that indicates that two L2 sounds that differ from an L1 sound are assimilated equally to one L1 phonetic category. The L2 sounds can be equally different or similar to an L1 phoneme, thus, the discrimination is expected to be poor.

The fourth pattern is the “uncategorized- uncategorized” assimilation where L2 sounds are perceived as speech sounds and fall into the phonological space, but neither can be assimilated to any existing L1 category, hence, the discrimination of these sounds is predicted to range from bad to very good, depending on whether the L2 sounds are perceived as similar to the same L1 category or not.

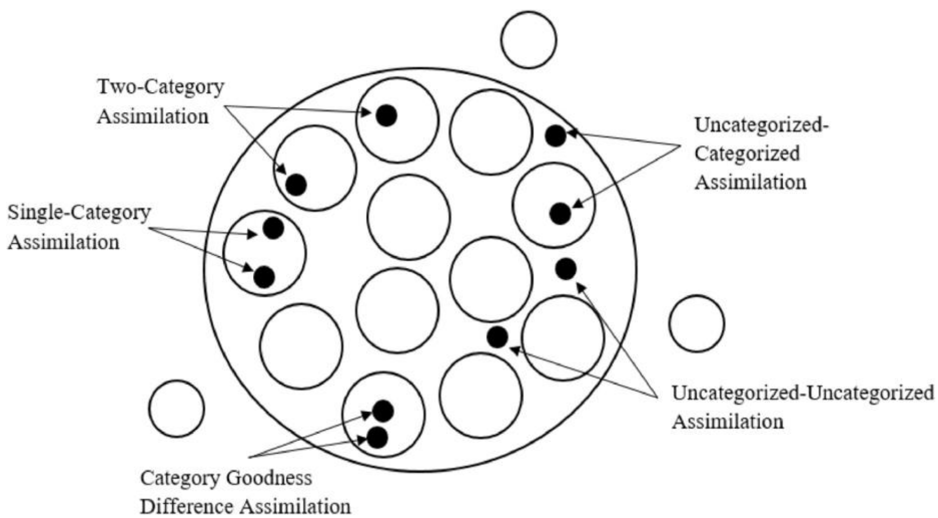
The last but one pattern is the “uncategorized-categorized” assimilation pattern and it yields that only one non-native sound can be assimilated to an L1 category while the other cannot be assimilated into any L1 phonetic category and falls in the phonological space. The discrimination of non-native sounds in this assimilation is expected to be very good.

Finally, the sixth pattern is the “non-assimilated” pattern that suggested that L2 sounds are perceived as non-speech sounds and they cannot be assimilated to any L1 phonetic categories. The discrimination of the L2 sounds discrimination could range from poor to excellent. Figure 2 illustrates the main assimilation types proposed by the PAM model.

Later, Best and Tyler (2007) further developed the PAM model into PAM-L2. Best and Tyler (2007) stated that the PAM-L2 model differs from SLM as it **“**addresses equivalence not only at the phonetic level addressed by the SLM, but also at the phonological level” (p. 27). PAM-L2 proposes that, in order to discern linguistic differences, the phonetic details of a speech should be linked to the phonological contracts. Accordingly, PAM-L2 predicted the perceptual mechanisms of advanced level sequential bilinguals from a lexical-functional perspective rather than focusing on the degree of resemblance between L1 and L2 phonemes. PAM-L2 theorizes that L2 sounds can be assimilated to the same sound category of the L1, even if the speaker is able to differentiate between the two sounds. For example, learners of French can distinguish between the French uvular fricative [ʁ] and the English postalveolar approximant [ɹ] but he categorizes them into the same lexical-functional category /r/ (Best & Tyler, 2007).

Figure 2

Assimilation types proposed by PAM



Note. From "Perceptual realization of Greek consonants by Russian monolingual speakers,", by G. P. Georgiou et al., 2020, *Speech Communication*, *125*, p. 8.

Tyler (2019) noted that the PAM-L2 prediction cannot be directly applied to the school context as the model was originally proposed based on empirical research with naïve novice learners in the naturalistic setting. The disparities in the quantity and quality of input in the school setting differ substantially from those in the naturalistic setting, making learners less likely to develop single-category assimilations and category-goodness assimilations.

2.1.5. Revised speech learning model

Recently, Flege and Bohn (2021) proposed a revised model of the SLM (SLM-r) motivated by the fact that some assumptions of the SLM needed clarifications and that some parts of the original model had become unproductive. Flege and Bohn (2021) summarized the SLM-r approach as follows: “If one needed a two-word summary of the SLM-r approach those two words would be that, there is no change in how the vowels and consonants found in an L1 and in an L2 are learned” (Flege & Bohn, 2021, p. 23).

Flege and Bohn (2021, p. 23) indicated that the SLM-r has three core assumptions. The first assumption is that the phonetic categories that are used for word recognition are based on statistical input distributions. The second assumption is that L2 learners of all ages use the same mechanisms and techniques that the children use when acquiring the L1 to learn L2 speech. The third assumption is that the differences in L2 production between native and non-native speakers are prevalent not because of the limited neuro-cognitive developmental abilities at a certain stage but because learners apply the mechanisms and processes that they used in L1 acquisition to the sounds of L2 and this does not yield to the same results.

Flege and Bohn (2021) proposed eleven hypotheses for their SLM-r. The first states that perception and production coevolve. This hypothesis stipulates that L2 segmental perception and production coevolve without precedence (Flege & Bohn, 2021, p. 29). This hypothesis is based on the findings in L2 research (Baker & Trofimovich, 2006; Flege, Bohn, & Sunyoung, 1997) that demonstrated a significant bidirectional relationship between production and perception where some L2 learners could generate contrasts in L2 sounds that they are unable to detect. This means that perception and production should mirror one another during L2 learning process.

The second hypothesis concerns L2 input. As SLM proposed that L2 learners discover progressively L1-L2 phonetic distinctions through daily life experience, this model did not provide a method to measure how much phonetic input is required to form new phonetic categories nor how to measure these phonetic knowledge accumulates. SLM-r redefined input as the sensory stimulus associated with perceived L2 speech in meaningful conversations. This second hypothesis emphasizes on the importance of measuring the quality and the quantity of input and proposed innovative measures to quantify these elements (Flege, Aoyama, & Bohn, 2021).

The next hypothesis is about the perceived cross-language dissimilarity. Just as the SLM, this hypothesis claims that learners subconsciously and automatically associate L2 sounds with L1 phonetic categories, and the greater the perceived phonetic dissimilarity of L2 phonemes from the realizations of L1 category, the more likely new phonetic categories will be formed. Flege and Bohn (2021) indicated that the optimal way to quantify cross-language dissimilarity must be determined and they suggested that the dissimilarity should be assessed perceptually rather than acoustically.

The fourth hypothesis stipulates that the “age hypothesis” is replaced with the “L1 category precision hypothesis” As the SLM proposed that the chronological age and the state of development of L1 phonetic categories of L2 learners at the time of exposure to L2 may impact L2 speech learning, Flege and Bohn (2021) stated that it is impossible to dissociate the state of development of L1 phonemes from the neurocognitive developmental state of L2 learners. For that, SLM-r claims that learners with relatively precise L1 phonetic categories are better than learners with relatively imprecise L1 phonetic categories in distinguishing phonetic differences between L2 sounds and their closest L1 equivalents. This ability to discern the cross-language phonetic difference will have an influence on the production and the perception of L2 sounds. Flege and Bohn (2021) added that it is pointless to attribute the disparities in the precision of L1 category to the age of learners when L2 learning begins.

The following hypothesis states that the capacity to form phonetic categories remain intact over the lifespan, regardless the age of first exposure. According to the SLM-r, the creation of L2 sound categories is defined by the statistical features of input distributions and it is predominantly influenced by three factors: the degree of perceived similarity between L1 and L2 sounds, the quality and quantity of L2 input that the learner obtained from meaningful conversations, and the accuracy of existing L1 categories at the moment of L2 learning. The SLM-r proposed that the formation of new L2 phonetic categories has three stages: the identification of the phonetic differences between the realizations of L2 sounds and the L1 sounds that is nearest to its phonetic space; the emergence of a functional "equivalence class" of speech symbols which are similar; and finally, the dissolution of the perceptual link between the L2 equivalence class and the L1 category at an undetermined point in phonetic development.

The sixth hypothesis suggests that L1-on-L2 and L2-on-L1 effects are unavoidable because the phonetic elements of the L1 and L2 subsystems of the learner reside in the same phonetic space. Added to that, the extent of cross-language phonetic impact may be affected by how learners employ the phonetic categories. It is expected that L2-on-L1 effects to increase as L2 proficiency develops as a result of the frequent use of L2. The subsequent hypothesis concerns cue weighting. SLM-r posits that as input distributions change, the structure of new L2 categories as well as composite of L1–L2 categories change. These changes are motivated by the need to categorize the phonetic segments quickly and effectively. Also, Flege and Bohn (2021) indicated that individual differences occur among monolinguals, indicating that individual differences will be visible in the production and perception of L2 sounds.

The individual differences in speech learning ability were discussed in the eighth point. The SLM-r posits that the formation of a new L2 phonetic category may be modulated by difference among learners, i.e., their endogenous factors. These factors include, but are not limited to, auditory acuity, auditory processing, the ability to accurately mimic sounds, auditory working memory , musical ability, selective attention, and phonemic coding ability.

The individual differences in L1 phonetic categories were also discussed by Flege and Bohn (2021) in the ninth hypothesis. In their previous model, the SLM looked at the performance of “early” and “late” immigrants or between “high” and “low” achievers. This approach implicitly presupposes that all native speakers who are monolingual with the same L1 have identical, or at least very similar phonetic categories, which was proven to be incorrect. SLM-r suggested that L1 phonetic categories differ between monolinguals as they could be exposed to different dialects, and this could impact L2 speech learning.

The penultimate hypothesis implies that while SLM focused on between-group differences, the SLM-r deals with how individuals acquire L2 sounds and how L2 learning impacts their production and perception of L1 (Flege & Bohn, 2021, p. 58). The decision to make individuals the primary unit of analysis was prompted by the fact that it is difficult to form groups based on a single variable (“low” vs “high”) and by the fact that drawing meaningful inferences from grouped data might be difficult. Individuals might bring a variety of L1 categories during the L2 learning process and focusing on groups may hide these distinctions among individuals.

Finally, the SLM-r hypothesizes that the findings from research on adult and child L2 learners would be similar, albeit longer periods of time for adults. This was explained by the fact that both adults and children employ the same capacities to learn L2 speech where the phonetic categories and realization rules used in L1 and L2 phonetic subsystems remain malleable throughout time.

To sum up, SLM-r is distinct from SLM in that it focuses on how individuals learn L2 and how it affects their perception and production of L1 phonemes rather than on the differences between groups, i.e., child vs adult learners. The SLM-r no longer focuses on the comparison between early and late learners because the CPH no longer offers a reasonable explanation for the age-related effects. Also, SLM-r no longer focuses on highly experienced L2 learners and whether they are distinguishable or not from native speakers because it is impossible for L2 learners to perceive and produce TL sounds exactly as the native speakers. The SLM-r has yet to be evaluated empirically. Although it offers valid starting points for examining perception-production connections, there are still some problems that need to be addressed (Nagle & Baese-Berk, 2021).

2.1.6. Optimality theory

Optimality theory (OT) is a constraint-based theory that was first proposed for phonology theory, but later extended to other aspects of language such as syntax and semantics (Gass, Behney, & Plonsky, 2013). This theory was originally proposed by Prince and Smolensky (2004) and its purpose is to explain how the underlaying representation of any given language is mirrored onto the surface representation (Osifeso, 2022).

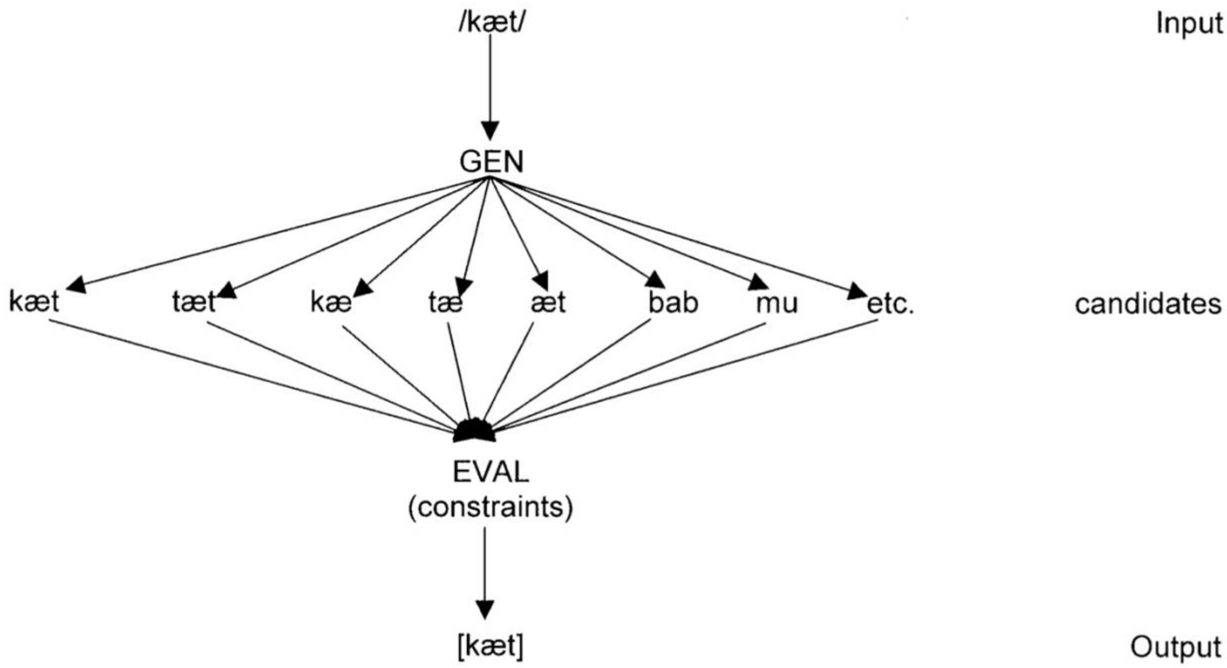
OT states that the forms of a language arise from the interaction between conflicting constraints that are governed by the articulatory and physiological capacities and limits. These constraints are innate where they are ranked, violable and apply to all languages (universal). The differences between the languages result from the varying rankings of these constraints during the acquisition process. The objective of these constraints is to match the pronounced form (output) with the lexical representation (input) in the most harmonic form possible.

OT is a mechanism that has three basic universal components: GEN (from generator), CON (from constraint) and EVAL (from evaluator) (Osifeso, 2022). GEN is the operational component and signifies that infinite set of candidate inputs are generated. Then, these input candidates are evaluated (EVAL) against a set of hierarchically ranked constraints (CON). The input candidate with the most harmonic and with the fewest number of violations and meets a high-ranking constraint is named the optimal output (McCarthy, 2007). Figure 3 schematises the mechanism of the OT.

OT proposed that there are two types of constraints: faithfulness (positive constraint) and markedness (negative constraint). Markedness constraints evaluate the surface forms and demands their well formedness, prohibiting the marked structures of the input. Faithfulness demands the faithful mapping between the input and output forms, i.e., the identical match between the input and their respective output. An element is produced if faithfulness is marked higher otherwise; if markedness is ranked higher, the element is not produced (Bernhardt & Stemberger, 2020)

Figure 3

The optimality theory



Note. Adapted from "Case Study: Optimality Theory and the Assessment and Treatment of Phonological Disorders," by J. Barlow, 2001, *American Speech- Language-Hearing Association*. *32*(4), p. 243.

2.1.7. Ontogeny phylogeny model

The Ontogeny phylogeny model (OPM) was proposed by Major (2001) and its goal is to capture the fundamental patterns of the interlanguage at the different stages of L2 phonological development and to explain the reason for foreign accent. OPM explains the relationship between L1-L2 transfer and universals (Stefanich & Cabrelli, 2021). These three linguistic systems will influence learners' output. It is to be noted that "universals" was defined by Major (2001) as the "universals of language that are not already part of the L1 or L2 system" (p. 83).

Major (2001) summarized the OPM model as follow: “L2 increases, L1 decreases, and U [universals] increases and then decreases” (p. 82). That being said, at the early stage of the interlanguage, the learner masters the L1 system and the influence of L2 and universals are not present. As the process of L2 acquisition begins, the influence of transfer is greatest at the beginning and L2 learners tend to substitute their L1 sounds for new L2 sounds and as a result, L1 decreases as the awareness of the L2 phonological sounds develops (Saito & Plonsky, 2019). Hence, the interlanguage pronunciation development starts, and learners seek to employ some composite forms between L1 and L2. During this stage, regardless of the learner’s L1 background, some universal characteristics of the learner’s L2 pronunciation forms may emerge and in later stages these universals decrease as the learner’s L2 knowledge gets stronger. The level of interference and transfer will vary depending on the stage of acquisition. This chronological development of the interlanguage is intitled the “Chronological Corollary of the OPM”.

OPM proposed three other corollaries to its model: the stylistic, similarity and markedness corollaries. The stylistic corollary of the OPM states that universals increase and then decrease according to the speech style (Major, 2001). When the style becomes informal, the influence of transfer becomes more dominant and vice-versa. Namely, L2 learner’s forms tend to be more native-like when their performance is prompted in formal (controlled) tasks then in informal tasks or spontaneous speech tasks. Majors (2001) explains this by the fact that learners are more conscious and monitor their production resulting in less L1 transfer. In the spontaneous tasks, learners pay attention on the meaning rather than on form and consequently, L1 transfer increases.

The third corollary of the OPM is similarity, where it discusses the claim that similar sounds between L1 and L2 are difficult to acquire in the course of development than the less similar or dissimilar sounds because it is difficult for learners to distinguish the small differences between the similar sounds. Major and Kim (1996) argue that the similar sounds are not difficult to acquire, but they are learned at a slower rate. When the learner acquires the similar sounds, the interlanguage develops in the following steps: L2 increases slowly, L1 decreases slowly, and the universals increase slowly and then decrease slowly. However, when acquiring dissimilar sounds, the rate of learning will be more rapid (Major, 2008). For this third corollary, the role of L1 is much greater than the universals for the similar form.

The fourth corollary of the OPM concerns markedness. This corollary stipulates that the more marked a sound, the slower the rate of acquisition will be. When the learner acquires the marked sounds, the interlanguage develops in the following steps: L2 increases slowly, L1 decreases and then decreases slowly, and the universals increase rapidly and then decrease slowly. Contrastingly to the third corollary, the role of the universals is much greater than L1 for the marked forms, compared to the less marked phenomena (Khalifa, 2020).

2.2. Pronunciation constructs

According to Saito and Plonsky (2019), the L2 pronunciation constructs encompass global and specific features. The global features include comprehensibility, accentedness, and intelligibility and the specific features include segmental, prosodic, and temporal features. Huang and Jun (2011) stated that the acquisition of L2 pronunciation requires the mastery of the segmental and suprasegmental features.

Saito and Plonsky (2019) reported that the global and specific constructs of L2 pronunciation are to some extent independent where the correlation between segmental accuracy and accentedness could be considerably strong but may become weak for accentedness and suprasegmentals. The following sections explore each the global and specific pronunciation constructs.

2.2.1. Global L2 pronunciation constructs

2.2.1.1. Comprehensibility

Comprehensibility is a valuable construct for both speakers and listeners that takes time and effort to develop (Trofimovich, Isaacs, Kennedy, & Tsunemoto, 2022). It is defined as “the listeners’ perception of how difficult it is to understand an utterance” (Derwing & Munro, 2005, p. 385). Isaacs and Trofimovich (2012) defined comprehensibility as “listeners’ perceptions of how easily they understand L2 speech” (p. 476) while Saito, Trofimovich, and Isaacs (2016) defined the term as “ease or difficulty in raters’ understanding of L2 speech” (p. 8). In a more recent study, Saito (2021) defined comprehensibility as “the *actual* effort made to understand” (p. 29). Munro and Derwing (1995a) view comprehensibility as the amount of effort required by listeners to understand the utterances. Likewise, Levis (2006) defines comprehensibility as “a measure of how comfortable a speaker is to listen to” (p. 254).

Comprehensibility is usually operationalized by the intuitive evaluation of raters of a speech sample using Likert scales (Trofimovich et al., 2022). Munro and Derwing (1995a) measured comprehensibility using a 9-point scale, where 1 = extremely easy to understand and 9 = impossible to understand (p. 79), while Derwing and Munro (1997) used slightly different descriptors with 1 = not difficult to understand at all and 9 = very difficult to understand. Some studies reversed the scale where for example, Trofimovich and Isaacs (2012), Isaacs and Trofimovich (2012) and Isaacs and Thomson (2013) used a 9-point scale to operationalize comprehensibility with 1 = hard to understand and 9 = very easy to understand. Some other researchers used a 7-point Likert scales (Ludwig & Mora, 2017) and a 10-point Likert scales (Caspers, 2010). Yet, Munro (2017) and Derwing (2018) concluded that using a 9-point scale is the most appropriate scale to rate comprehensibility and that fewer scalar points would be too restrictive for judges.

Comprehensibility ratings are generally reliable among listeners regardless of how it is measured. However, comprehensibility ratings could differ from intelligibility (see forthcoming discussion about this construct), where they are usually more complex to assess as listeners can rate an utterance as less comprehensible when it requires additional effort or longer processing time (Derwing & Munro, 1997; Munro & Derwing, 1995a). Comprehensibility judgements can be useful for researchers as a practical measure of L2 comprehension, since they present an assessment of the listeners' real understanding of speech (Trofimovich et al., 2020). Furthermore, comprehensibility could inform researchers about the linguistic dimensions that could hinder the comprehension process and help teachers to target those linguistic dimensions through instruction (Trofimovich et al., 2022). According to Crowther, Trofimovich, Saito, and Isaacs (2018), comprehensibility is mainly associated with suprasegmentals. Meanwhile, Munro and Derwing (1995a) found that comprehensibility is associated with phonetic substitutions, intonation and morphosyntactic errors. In more recent research, Bergeron and Trofimovich (2017), Crowther et al. (2018), and Saito et al. (2016) confirmed that comprehensibility is associated to multiple linguistic correlates. By examining the extracts from a picture narrative task by 40 Spanish speakers of French L2, Bergeron and Trofimovich (2017) found that comprehensibility was linked not only to overall pronunciation assessment, but also to discourse richness, fluency, lexis, and grammar.

Crowther et al. (2018) cited that besides the linguistic dimensions, the speaker's linguistic background and task type also have an effect on learner's comprehensibility. Nevertheless, the findings of Crowther, Trofimovich and Isaacs (2016) and Saito, Trofimovich, Isaacs, and Webb (2017) have demonstrated that the evaluators' background (non-native vs native listeners) has no major effects on the global ratings of comprehensibility. Nevertheless, despite the importance of comprehensibility evaluation, Trofimovich et al. (2022) noted that the target speakers in comprehensibility studies are mainly university students, seldom school-aged learners, and almost never children (the target population of this thesis).

2.2.1.2. Accentedness

Accentedness was emphasized in literature given that accented speech could be associated with negative appraisals of intelligence, successfulness and social standing (Fuertes, Gottdiener, Martin, Gilbert, & Giles, 2012). A foreign accent or the notion of accentedness can be defined in terms of how the listeners perceive it. Derwing and Munro (2005) defined accentedness as "the listeners' perception of how different the speaker's accent is from that of the L1 community" (p. 385). Saito et al. (2016) described accent as "listeners' perceptions of the degree to which L2 speech is influenced by his/her native language and/or colored by other non-native features" (p. 8).

Some other researchers define accentedness in terms of how much L2 speech differs from the TL (Thomson, 2018) or as a measure of nativeness (Crowther et al., 2016). Saito (2021) described accentedness as "how closely L2 speech approximates the phonological norm of native speakers" (p.3). Similarly, Isaacs and Thomson (2013) identified this linguistic dimension as "how different the speaker sounds from a NS [native speaker]" (p. 141).

Numerous studies operationalize accentedness in a similar way through scalar judgments by the means of human raters. Munro and Derwing, (1995a, 1995b) evaluated foreign accent using 9-point scales, where 1 = no foreign accent and 9 = very strong foreign accent. Saito et al. (2016) used a 9-point rating scale with slightly different descriptors to measure extemporaneous speech samples where 1 = no accent, and 9 = heavily accented.

Regardless of learner's orientation toward nativeness (acquiring a native-like pronunciation) or intelligibility (see next section), the key to assisting L2 learners attain their goals is by identifying the linguistic dimensions of L2 speech that are associated with accentedness, and that could impede comprehensibility and intelligibility. This is because the differences in the speaker's oral productions can sometimes have no impact on the listener' understanding while in other cases they can result in communication breakdowns (Bergeron & Trofimovich, 2017; Derwing & Munro, 2022). This will assist L2 teachers in selecting pedagogical foci that are more compatible with the learner's needs.

Consequently, several studies were carried out to determine which linguistic dimensions interfere with accentedness. Research demonstrated that accentedness is principally linked to segmental accuracy, i.e., consonant and vowel errors (Crowther et al., 2018; Munro & Derwing, 2006; Trofimovich & Isaacs, 2012), and to some suprasegmental dimensions of speech such as speech rate, pause, intonation, and word stress (Bergeron & Trofimovich, 2017; Kang, 2010; Kang, Rubin, & Pickering, 2010; Munro & Derwing, 2001). Bergeron and Trofimovich (2017) and Nagle and Huensch (2020) inquired if these linguistic correlates of accentedness differed for L2s other than English because there has been little research in languages other than English and they advised further research into other L2s and in different contexts of learning for a better understanding.

Some other non-linguistic features were also found to impact accentedness ratings. For example, Piske, MaKay, and Flege (2001) found that the strongest predictor of the degree of accentedness is age. Bergeron and Trofimovich (2017) and Crowther et al. (2018) reported that task complexity has a substantial impact on accentedness evaluation, demonstrating that accentedness could be distinguished from comprehensibility in a simple task (picture description task) but not in a more cognitively demanding task (interview).

2.2.1.3. Intelligibility

Munro (2011) stated that: "Intelligibility is the single most important aspect of all communication. If there is no intelligibility, communication has failed. In language pedagogy this […] is an empirically sound concept that will provide a basis for a wide range of pedagogically-oriented research in the future" (p. 13).

Intelligibility is the most important dimension for L2 speakers, teachers, and researchers (Derwing, 2018, p. 13). It is defined as the extent to which a speaker's message is actually understood by the listener (Munro & Derwing, 1995a, 2011). Saito (2021, p. 29) presents a close definition where he defines intelligibility as "the actual outcome of understanding". Suprasegmental features of L2 speech play an important role for intelligibility proficiency (Derwing, Munro, & Wiebe, 1998; Field, 2005; Liu & Reed, 2019). From speakers' point of view, intelligibility refers to the ability to communicate in a way that listeners can understand by paying attention to both segmental and prosodic features (Levis & Silpachai, 2022).

Levis and Silpachai (2022) state that intelligibility involves several types of understanding: at the word level, at the message level and at the interpretation of the message level. At the word level, when the words in a message can be identified, the speech is considered as intelligible. However, if the words are mispronounced, or the noise of that word is difficult to understand, then the word can be misheard or not being understood at all. Levis and Silpachai (2022) noted that the intelligibility of isolated words is different from the intelligibility of sentences due to fundamental differences in perception and production. The word-level intelligibility problems often occur with segmental and stress placement inaccuracies. At the message level, when the intended meaning of an utterance is not understood, then the message is considered as unintelligible. At the interpretation level, intelligibility occurs when the listener grasps the meaning behind a message or “its illocutionary force” (p. 160). Levis and Silpachai (2022) stated that this type of intelligibility is rarely studied as the indented message of the L2 learners is often ambiguous.

Intelligibility scores can vary across task types and the nature of speech samples (Kang, Thomson, & Moran, 2018). Levis and Silpachai (2022) reported that intelligibility evaluation can also be affected by differences in dialects, speed of speech, L2 speaker’s limited proficiency level, unexpected phonological, grammatical, or lexical error, noise, attitudes towards L2 accentedness, and when non-native speakers use the same target language to communicate. Yet, intelligibility can be facilitated by a shared L1, where speakers with the same L1, and who share the same interlanguage, phonological and phonetic forms are more intelligible to the listeners of that same L1 (Edwards, Zampini, & Cunningham, 2019).

Nevertheless, this facilitative effect of the shared L1 between speakers and listeners was not supported by Munro, Derwing, and Morton (2006). These researchers examined the production of 48 participants and 12 non-native speakers of Cantonese, Japanese, Spanish and Polish. After the analysis of the extemporaneous narratives, Munro et al. (2006) found that Japanese listeners considered the native Japanese’s English as more intelligible than that of the Cantonese, Mandarin and English speakers. Yet, the Cantonese listeners did not find the English produced by Cantonese speakers as more intelligible in comparison to the other languages. Munro et al. (2006) concluded that the shared L1 background effects are outweighed by other factors such as the proficiency level of the speakers.

Intelligibility is typically measured through transcriptions of speech in which listeners are asked to write down exactly what they had heard and then the score is calculated as the percentage of exact word matches with the speakers' production per sentence (Levis, 2018a). The scoring procedure can include counting the exact word matches of all words (Derwing & Munro, 1997; Munro & Derwing, 1995a; Munro et al., 2006), or only the content words (Bent, Bradlow, & Smith, 2007).

Additionally, intelligibility can also be operationalized through comprehension questions (Hahn, 2004) or by providing summaries of what was understood (Kennedy & Trofimovich, 2008; Munro & Derwing, 1995b; Saito, 2021). The construct can also be evaluated by verbal repetition (Fogerty & Humes, 2012), where listeners repeat aloud the sentence that they hear, and the score is determined by counting the number of correctly identified words in a sentence as judged by trained evaluators. Moreover, intelligibility might be assessed using a Likert scale (Gooch, Saito, & Lyster, 2016), but this measure is rarely used. According to Thomson (2018, p. 19), using a Likert scale to measure intelligibility is not a reliable tool as it could reflect listeners' subjective impression rather than demonstrating that they can match what they have heard.

2.2.1.4. Relationship between accentedness, comprehensibility, and intelligibility

While speaking an L2, communication can become challenging due to some non-target-like pronunciation features (Thomson, 2018, p. 12). In a broader sense, successful L2 communication is seen to be associated with the speakers' ability to sound nativelike (Bergeron & Trofimovich, 2017). The ideologies of nativeness and near-nativeness are deeply embedded in

L2 pronunciation. Nativeness has frequently been used to explain how non-monolinguals (L2 learners or bilinguals) differentiate from native monolinguals (Levis, 2020).

However, considerable strides have been done in L2 pronunciation research over the last two decades on whether intelligibility or nativeness should be the goal of L2 pronunciation learning (Derwing & Munro, 2015; Jenkins, 2000, 2007; Levis, 2005; Munro & Derwing, 2015; Murphy, 2014). Levis (2005) presented the Intelligibility Principle that stipulates that L2 learners' primary concern is the communicative success and to be understood despite the presence of a detectable accent (Crowther et al., 2018). This principle recognizes that accent diversity is marginally connected to communication problems, values the enormous abilities of non-native language teachers, and recognizes that not all pronunciation dimensions are equally important (Levis, 2020). Additionally, this principle stated that being a native speaker is not a sufficient qualification for teaching L2 pronunciation. According to Derwing and Munro (2005), adequate pedagogical training, proficiency and linguistic knowledge are sufficient criteria for language teaching. Lastly, the intelligibility principle asserts that a wide range of accents are appropriate as teaching models and that learners can develop and embrace their own accents (Levis, 2018b).

The intelligibility principle stands in contrast to the Nativeness Principle. The nativeness principle upholds high standards of L2 performance, where acquiring a native-like pronunciation is possible and desirable. According to Levis (2020), the nativeness principle is predicated on a myth that stipulates that there are perfect and inadequate ways to pronounce an L2, with the latter not being tolerated. It is also based on the belief that mastering flawless pronunciation ensures successful communication, greater respect, less discrimination, and better professional opportunities (Derwing & Munro, 2009).

This principle also posits that all aspects of pronunciation are equally important and that any unmastered component of pronunciation shows that the learner has failed (Levis, 2020). Furthermore, the nativeness principle assumes that only native or native-like teachers are qualified and trusted to teach the TL pronunciation, whereas non-native teachers may be considered as inadequate examples by their students and colleagues, leading to eventual discriminatory practices (Levis, 2020). Finally, this principle claims that only certain accents are acceptable (for example, the American and the British English accents).

Most rigorous research goes contra the Nativeness Principle by demonstrating that acquiring a native-like pronunciation is biologically conditioned, and that very few learners with specific cognitive abilities and whose L1 is linguistically close to the TL are able to attain native-like pronunciation norms (Abrahamsson & Hyltenstam, 2009; Saito, 2021). Derwing and Munro (2015) emphasized on the fact that attaining L2 native-like pronunciation is exceedingly difficult and even an unrealistic goal. L2 students should be encouraged to target achievable and realistic goals of intelligibility and comprehensibility instead of focusing on accent reduction (Levis, 2018a, 2020; Saito, 2021). The most important goal is for the listener to grasp the words, messages, and the intention of the speaker (Levis & Silpachai, 2022). Instead of attempting to eliminate pronunciation errors in order to promote accent-free speech, teachers should concentrate solely on features of pronunciation that affect comprehensibility and intelligibility in order to facilitate L2 communication (Derwing & Munro, 2005; Saito & Lyster, 2011).

The role of intelligibility and comprehensibility as useful developmental goals has been consolidated with the pioneering work of Derwing and Munro (1997, 2005, 2015), Munro and Derwing (1995a) and Munro et al. (2006) who provided evidence of the partial independence of the three linguistic dimensions. Comprehensibility and intelligibility are closely related to each other than they are to accent (Derwing & Munro, 2015, p. 5). In contrast, the relationship between intelligibility and accentedness is generally weak and more inconsistent (Nagle & Huensch, 2020). In fact, it is quite possible that L2 speakers produce a heavily accented speech and yet be fully intelligible and easy to understand (Derwing & Munro, 1997, 2022; Munro & Derwing, 1995a). Still, it is also possible to be difficult to understand and yet fully intelligible. Derwing and Munro (2022) stated that accentedness is less significant than comprehensibility and intelligibility, which are far more critical for successful communication.

Levis (2006, 2020) point out that the terms intelligibility and comprehensibility are used interchangeably but do not refer to the same thing. Intelligibility and comprehensibility have in some ways similar aspects where they can be affected by the pronunciation and the lexico-grammatical features of the language, and both can fluctuate even while accentedness remains constant (Levis & Silpachai, 2022). However, while both constructs are measures of understanding of L2 speech, they are operationalized differently. Comprehensibility is measured by the amount of effort that listeners undertake to understand a speech and not by the success and failure to understand the utterance (Levis & Silpachai, 2022).

2.2.2. Specific L2 pronunciation proficiency

According to Isaacs and Trofimovich (2016), L2 pronunciation encompasses segments that is composed of individual consonants and vowels; and the prosodic features that englobe word stress, rhythm, and intonation. Segments of the language are distinct from the suprasegmentals, yet they do overlap to some extent (Levis, 2018b). Segmental and the suprasegmental constructs of L2 will be presented in the following sections.

2.2.2.1. L2 Segments

Segments encompass the vowel and consonant sounds of a language (Levis, 2018b). In order to speak clearly and be understood, an L2 learner has to master a sufficient, but not perfect, accuracy of the sounds of the target language in order to match what listeners expect to hear (Levis, 2018b). The L2 segmental features demonstrate learners' capacities to pronounce new consonantal and vocalic sounds without deleting or substituting them with their L1 counterpart. Segments can be measured through expert raters' scalar judgments and acoustic analyses (Saito & Plonsky, 2019).

Levis (2018b) described segments according to four areas: phonemes, allophones, phonotactic constraints and syllable structures. Segments can be described phonologically. A phoneme is the individual unit of sounds in a word (a set of allophones). Phonemes are language-specific, and they influence how a spoken language is perceived and produced. For example, French has three voiceless stops /p/, /t/ and /k/ whereas Arabic has only two voiceless stops /t/ and /k/.

Segments can also be described phonetically. An allophone is the articulation of one phoneme. As an illustration, the allophone /p/ can be pronounced as [p] in "apple" or aspirated as [p^h] in "pit". Levis (2018b) stated that allophones, rather than phonemes, could affect learners' intelligibility. Another way to describe segments is through the linguistic constraints that exist in a language. The phonotactic constraints are the rules and restrictions that govern how syllables can be formed in language. These constraints are language-specific. For instance, the allophone [k^h] is pronounced at the beginning of stressed syllables in English but not after [s] (Levis, 2018b).

Lastly, the fourth method of describing segments is to determine how they fit into syllable structures (Levis, 2018b). Syllable structures are the arrangements of sounds in words, and they are also language-specific. For example, some languages allow multiple clusters of consonants (English for example) while in some other languages consonant clusters are not allowed (Arabic for example prohibits more than two consecutive consonants).

Due to the differences in the phonetic inventories of the speakers' L1 and TL errors of pronunciation can occur (Derwing & Munro, 2015). Segmental errors are common in L2 pronunciation. These mispronunciations can lead listeners to hear different words which can result in some cases in challenges for listeners to understand speakers' intended message. Errors of pronunciation can also lead to loss of intelligibility where words are not understood at all (Levis, 2018b, p. 72). However, not all pronunciation errors have the same impact where some errors are less serious in their effect. Errors or mispronunciations originate from deviations such as substitutions (replacing one phoneme for another), deletions (eliminating expected sounds), distortions (changing phonemes in a way that do not match expected sounds) or epenthesis (addition of extra sounds) (Derwing & Munro, 2015).

Regarding the framework of this dissertation, the L2 pronunciation proficiency for Tunisian Arabic (TA) learners of French will be examined. Accordingly, it would be necessary to differentiate TA from the Modern Standard Arabic, by comparing the two phonemic inventories and presenting the distinctive phonological features of TA. Following that, the phonemic inventory of French will be provided. Finally, the main segmental difficulties for Tunisian learners of French will be discussed in the last section.

2.2.2.1.1. Phonemic inventory of Modern Standard Arabic

Arabic is the sixth most spoken language in the world by nearly all the people of the 22 Arab countries and by more than 200 million speakers globally (Miller & Caubet, 2010) and it is also the liturgical language of over a billion Muslims around the world. Arabic is a complex entity that consists of three main varieties: Classical Arabic; MSA and colloquial or dialectal Arabic. Classical Arabic is the language found in the Holy Quran and is used by more than 1 billion Muslims around the world (Alqattan, 2015). MSA is the official language of all Arab countries. It is taught at school and used in any type of formal written and oral communication (e.g., in news broadcasts, official speeches, conference presentations, etc.). Arab communities rarely or never speak Standard Arabic but rather the local dialect variety of the country in which they live (Ben Romdhane, 2019, p. 33). Dialectal Arabic is used in daily life for spoken communication and in informal spoken media, such as soap operas (Alqattan, 2015). It is the true native form of Arabic.

Given the differences in cultures and geographical proximity, there are many Arabic dialects across the Arab World. The differences among these regional vernaculars are considerable and affect every aspect of the language (syntax, phonology, morphology, semantics). These variances may hinder communication to the extent that Arabs who speak very distinct dialects choose to communicate with each other using Modern Standard Arabic. These dialects can be classified according to geographical areas into five major subgroups: Hijazi (Arabian Peninsula and Yemen); Mesopotamian (Iraq); Levantine (Syro-Lebanese); Egyptian (Egypt and Sudan); and Maghribi (North African) (Youssef, 2013). Tunisian is a subset of the Maghribi dialect.

Embarki (2008) provided an example (table 3) of the variability in pronunciation of the interdentals / θ, ð, $t^{ʕ}$/ and the uvular /q/ in the five major subgroups, demonstrating the disparities in pronunciation of these phonemes in these regions, and that the differences in pronunciation exist even between the Bedouin and city dwellers in the same region.

Table 3

Example of Pronunciation of / θ, ð, $t^{ʕ}$ q/ in the Five Major Areas

Division géographique/ sociologique	Arabique	Mésopotamien	Levantin	Égyptien	Maghrébin
bédouins nomades	ǵ-ğ, ṯ, ḏ, ḏ̣, ī, ū, ē, ō, ā, i, u, a	ğ, ṯ, ḏ, ḏ̣, ī, ū, ē, ō, ā, i, u, a	k, ṯ, ḏ, ḏ̣, ī, ū, ē, ō, ā, i, u, a	g, s, z, ẓ, ī, ū, ē, ō, ā, i, u, e, o, a	g, ṯ, ḏ, ḏ̣, ī, ū, ā, i, u, a, ə
bédouins sédentaires	ğ-g, ṯ, ḏ, ḏ̣, ī, ū, ē, ō, ā, i, u, a	g, ṯ, ḏ, ḏ̣, ī, ū, ē, ō, ā, i, u, a	ḳ-g, ṯ, ḏ, ḏ̣, ī, ū, ē, ō, ā, i, u, a	k, s, z, ẓ, ī, ū, ē, ō, ā, i, u, e, o, a	g, t, d, ḍ, ī, ū, ā, i, u, a, ə
citadins	ğ-g, ṯ, ḏ, ḏ̣, ī, ū, ē, ō, ā, i, u, a	q, t, d, ḍ, ī, ū, ē, ō, ā, i, u, a	ʔ, t-s, d-z, ḍ-ẓ, ī, ū, ē, ō, ā, i, u, a	’, s, z, ẓ, ī, ū, ē, ō, ā, i, u, e, o, a	q, t, d, ḍ, ī, ū, ā, i, u, a, ə

Note. From “Les dialectes arabes modernes : état et nouvelles perspectives pour la classification géo-sociologique,” by M. Embarki, 2008, *Arabica*, *55*(5), p. 529.

2.2.2.1.1.1. Consonants of MSA

Consonants can be categorized as follows: one lateral /l/; one affricate /ʤ/; one trill /r/; two nasals /m, n/; two semi-vowels /w, j/; twelve fricatives /f, θ, ð, s, z, ṣ, ʃ, χ, ɣ, ħ, ʕ, h, $ð^{ʕ}$/ and ten stops /b, d, g, t, k, q, ʔ, $t^{ʕ}$, $d^{ʕ}$/. In MSA, places of articulation are profoundly used in the posterior half of the vocal tract to produce the glottals [h ه, ʔ ء], the pharyngeals [ħ ح, ʕ ع] and the uvulars [ɣ غ, χ خ, q ق] (Odisho, 2005, p. 25). That same region in the mouth is used to articulate emphatic sounds [ṣ ص, ðˤ ط ,ضtˤ]. Table 4 presents the consonant inventory of MSA in IPA and table 5 presents the consonant inventory of MSA in Arabic symbols.

Table 4

Arabic Consonants of MSA in IPA

Manner of Articulation	Place of Articulation		Bilabial	Labio-Dental	Inter-dental	Alveolar	Post-Alveolar	Palatal	Velar	Uvular	Pharyn-geal	Glottal
	Plosive: Plain Emphatic		b			t^{h} d ṭ ḍ			k^{h}	q		ʔ
	Affricate						ʤ					
	Fricative: Plain Emphatic			f	θ ð ð̣	s z ṣ	ʃ			χ ʁ	ħ ʕ	h
	Approximants	**Central**	w					j				
		Lateral				l						
		Nasal	m			n						
	R-sounds: Tap					ɾ						
	R-sounds: Rolled					r						

Note. From *Techniques of Teaching Comparative Pronunciation in Arabic and English* (p. 26), by E. Y. Odisho, 2005, New Jersey: Gorgias Press.

Table 5

Arabic Consonants of MSA in Arabic Symbols[1]

مخارج الاداء			اثنا شفوية	شفوية اسنانية	عبر اسنانية	لثوية	غارية لثوية	غارية	طبقية	لهوية	حلقية	حنجرية
انماط الاداء	شديدة	مستقلة / مطبقة	ب			د ت ض ط			ك	ق		ء
	مركبة						ج					
	رخوة	مستقلة / مطبقة		ف	ذ ث ظ	ز س ص	ش			غ خ	ع ح	هـ
	متمادية	مركزية	و					ي				
		جانبية				ل						
		انفية	م			ن						
	رائية	مستلة				ر						
		مكررة				رّ						

Note. From *Techniques of Teaching Comparative Pronunciation in Arabic and English* (p. 26), by E. Y. Odisho, 2005, New Jersey: Gorgias Press.

---

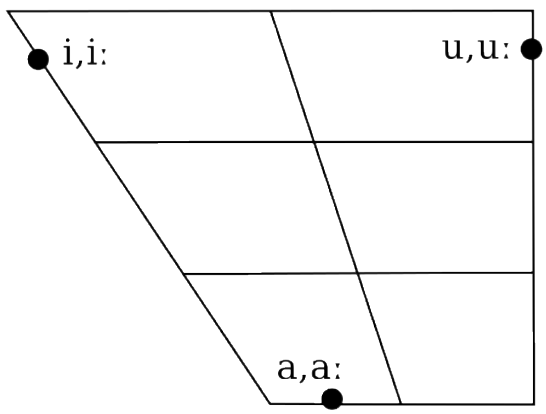
[1] مخارج الأداء = Place of articulation
أنماط الأداء = manner of articulation
ثنائية شفوية = bilabial
شفوية اسنانية = labio-dental
عبر اسنانية = inter-dental
لثوية = alveolar
غارية لثوية = post-alveolar
غارية = palatal
طبقية = velar
لهوية = uvular
حلقية = pharyngal
حنجرية = glottal
مفتلة = plosive
مركبة = affricate
رخوة = fricative
تمادية = approximants
متلة رائية = tap R-sounds

2.2.2.1.1.2.Vowels of MSA

MSA has only three short vowels and three long vowels: /i/, /a/, /u/, /iː/, /aː/, /uː/. This limited vocalic inventory can lead to articulatory difficulties for second language learners, as they have to acquire new sounds and learn how to articulate them. Figure 4 illustrates the vocalic inventory of MSA.

Figure 4

Vocalic Inventory of MSA



Note. From *Investigating the perception and production of the Arabic pharyngealised sounds by L2 learners of Arabic* (p. 18), by H.S.M. Binasfour, 2018.

According to Al-Busaidi and Al-Saqqaf (2015), Arabic speakers depend a lot on the context to interpret the right pronunciation of words. In fact, many Arabic words are spelled the same but are pronounced differently to give different meanings. For example, the words for "camel" and "sentences" are written the < جمل > but native speakers know from the context that the first is pronounced /dʒamal/ and the other is /dʒumal/ (Al-Busaidi & Al-Saqqaf, 2015, p.184). Also, MSA uses diacritical marks, i.e., small marks put on words known as "fatha", "damma" and "kasra" to indicate /a/, /u/ and /i/ vowels respectively. Table 6 presents MSA vowels with diacritical marks and some examples.

Table 6

Arabic Vowels and Diacritical Marks

Transliteration	IPA	Allophones	Arabic Character	Example	Gloss
i	/i/	[i], [ɨ], [ə]	*kasra*	*bint* /bint/	"girl"
ī	/i:/	[i:], [ɨ:]	*yā' al-madd*	*tīn* /ti:n/	"figs"
a	/a/	[æ], [ə], [ʌ]	*fatHa*	*fann* /fann/	"art"
ā	/a:/	[æ:], [a:]	*alif al-madd*	*dār* /da:r/	"house"
u	/u/	[u], [o], [ɒ]	*Damma*	*dub* /dubb/	"bear"
ū	/u:/	[u:]	*wāw al-madd*	*'ūd* /ʕu:d/	"lute"

Note. From "English Spelling Errors Made by Arabic-Speaking Students," by S. Al-Busaidi and A. Al-Saqqaf, 2015, *ELT*, *8*(7), p. 184. https://doi.org/10.5539/elt.v8n7p181

2.2.2.1.1.3. Diphthongs of MSA

A diphthong is a combination of two vowels that glide from one to the other. MSA has two diphthongs /aj/ and /aw/. These diphthongs correspond to the long mod vowels /ee/ and /oo/ in most dialects of Arabic (Mustafawi, 2018).

2.2.2.1.2. Phonemic inventory of Tunisian Arabic

Tunisian Arabic, known as the "Darija" or "Tounsi", is a dialect of MSA. It is spoken in Tunisia by approximately 12 million people. Despite the country's small size, Tunisia represents a rich sociolinguistic mosaic with a long history of bilingualism and language contact. According to Mejri (2003), the Tunisian dialect is strongly influenced by Berber. During the Islamic conquests, Arabic language was introduced. This was followed by the Ottoman Turkish domination of North Africa from the late nineteenth century to the French colonization in 1830 that introduced in its turn a new language (Zribi, Graja, Khmekhem, Jaoua, & Belguith 2013). Besides French, Tunisian is also influenced by other European languages like Italian, Spanish and Maltese.

Tunisian Arabic has a wide range of regional varieties and dialects: the dialect of Tunis spoken by the capital, the dialect of Sahel spoken in some coastal cities, the dialect of Sfax spoken only in Sfax, the dialect of north-occidental spoken in Beja, Kef, etc., the dialect of south-oriental in Mednine, Gabes, etc., and the dialect of south-occidental in Gafsa, Sidi Bou Zid, etc.. (Baccouche, 2009; Hamrouni, 2010). Regional differences affect all levels of language, but Tunisian Arabic is still understood by all Tunisian people (Masmoudi, Khmekhem, Estéve, Bougares, & Belguith, 2014).

The following sections present the consonantal and the vocalic inventories of TA.

2.2.2.1.2.1. Consonants of TA

Just as like Modern Standard Arabic, TA has 28 consonants, six vowels (/iː/, /aː/, /uː/, /i/, /a/, /u/) and two diphthongs (/aj/ and /aw/). TA is considered to be the closest to MSA in terms of consonant pronunciation when compared to other Arabic dialects (Zribi et al., 2013). However, through borrowing from Berber and French, the Tunisian dialect added the /p/, /v/ and /g/ consonants to its inventory (Masmoudi et al., 2014). These three consonants do not exist in the phonological system of MSA. The uvular stop /q/ could be pronounced in certain regions as [g]. The consonantal inventory of TA has the following phonemes: /p, v, g, l, ʤ, r, m, n, f, θ, ð, s, z, ṣ, ʃ, χ, ɣ, ħ, ʕ, h, ðˤ, b, d, g, t, k, q, ʔ, tˤ, dˤ, w, j/. Table 7 presents the consonantal inventory of Tunisian Arabic.

Table 7

Consonantal Phonetic Chart of Tunisian Arabic

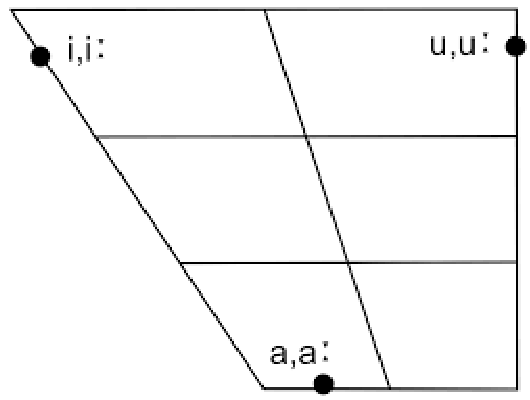
PoA MoA		Bilabial	Dental	Alveolar	palatal	velar	Uvular	Pharyngeal	Glottal
Nasal		[m] م	[n] ن						
Stop	**voiceless**	[p]پ	[t] ت	[tˤ] ط		[k] ك	[q] ق		[ʔ] ء
	voiced	[b] ب	[d] د	[ðˤ] ض					
Fricative	**voiceless**	[f] ف	[θ] ث	[s] س [ṣ] ص	[ʃ] ش	[x] خ [ɣ] غ		[ħ] ح	[h] ه
	voiced	[v] ڤ	[ð] ذ	[z] ز		[g] ڨ		[ʕ] ع	
Affricate				[dʒ] ج					
Lateral				[l] ل					
Trill				[r] ر					
Semi-vowel					[y] ي	[w] و			

2.2.2.1.2.2. Vowels of TA

What distinguishes Tunisian dialect from other Arabic dialects and MSA, is the fact that it maintains the opposition of short and long vowels and retains the vowel system of standard Arabic with the three short vowels /a/, /i/, /u/ and their corresponding long vowels /a:/, /i:/, /u:/ (Marçais, 1977). According to Mejri et al. (2009), this phenomenon created minimal pairs based on a vocalic lengthening of vowels and this does not exist in MSA. As an illustration, [sir] means 'secret' and [si:r] that means 'the imperative form of walk', or [qal] that means 'rare' and [qa:l] that means 'say'. Figure 5 presents the Tunisian vocalic system.

Figure 5

Vowel Diagram of Tunisian Arabic



Note. From *Investigating the perception and production of the Arabic pharyngealised sounds by L2 learners of Arabic* (p. 18), by H.S.M. Binasfour, 2018. https://core.ac.uk/download/pdf/222830826.pdf

According to (Mzoughi, 2015), TA omits short vowels between two consonants ([kataba] ‘write’ becomes [kteb], omits short vowels in the second to the last position in feminine nouns ([ʃajara] ‘tree’ becomes [ʃajra]) and omits the first short vowel in adjectives ([qasi:r] ‘short’ becomes ‘[qsi:r]). Another feature of TA is that it omits short vowels when they are located at the end of the syllable and make some changes to the syllabic structure of the word to become monosyllabic like in the case of " شرب" [ʃariba], "he drank" in MSA that becomes [ʃrab] (Zribi et al., 2013). Interestingly, short vowels have essentially a morphophonemic function in the root structure of Arabic words (Ali, 2013). Speakers of Maghrebian / Tunisian dialects would also be prone to centralize short vowels in closed syllables while speakers of other dialects exhibit a more peripheral distribution (Barkat, Ohala, & Pellegrino., 1999). The vowel /a/ would be the less stable phoneme of the Arabic vocalic system, depending on its contextual environment and regional varieties.

2.2.2.1.3. Phonemic inventory of French

French is a romance language of the Indo-European family that originates from Vulgar Latin. The European French is developed from the dialects of the Gallo-Romance area (Langue d'Oïl) and with other romance and non-romance languages (Langue d'Oc or Occitan, Franco-Provençal or Arpita, Breton, Basque, Dutch, and German) (Delais-Roussarie, Post, & Avanzi, 2015; Walker, 2001).

According to the report of the International Organization of the Francophonie, French is spoken across five continents by approximately 321 million individuals in 2022. It is spoken in more than 75 countries and enjoys the official status in 29 countries such as France, Canada, Belgium, Switzerland, Luxembourg, Cameroon, Ivory Coast, French Guiana, Madagascar, and French Polynesia. In the European Union, French is the second most widely spoken L1 and second most taught language. French is also the language of education in a number of countries such as Algeria, Morocco, Tunisia, and sub-Saharan Africa.

The wide geographical spread of French is correlated with linguistic differences. Walker (2001) stated that the French of Montréal is not the same as that of Liège or Geneva. Even within the borders of France, French recognizes regional varieties, accordingly, French in Normandie differs from that in Alsace. Walker (2001) added that in addition to the geographical diversity, there are social and stylistic varieties of the French language. He explained that the speech of "la haute bourgeoisie", the middle class, the elder and younger generations, as well as speech in classrooms and courtrooms. for example. differ. Notwithstanding all the variants of French, Walker (2001) asserted that there is "Standard French". He described it as a reference point and the most appropriate variant for usage in formal and educational settings. Côté (2021) defined the standard French as a concept that refers to two distinct notions: the pronunciation of a speaker that is characterized by his geographical origin, socioeconomic profile and professional status; and the pronunciation of the reference books.

2.2.2.1.3.1. Consonants of French

French phonemic system has seventeen consonants: /p, t, k, b, d, ɡ, f, s, ʃ, v, z, ʒ, m, n, ɲ, ʁ, l, ŋ/, three semi-vowels: /j, ɥ, w/ and sixteen vowels /i, e, ɛ, a, u, o, ɔ, y, ø, œ, ɑ̃, ɔ̃, ɛ̃, œ̃, ə, ɑ/. Table 8 illustrates the seventeen French consonant phonemes.

Table 8

French Consonant Phonemes

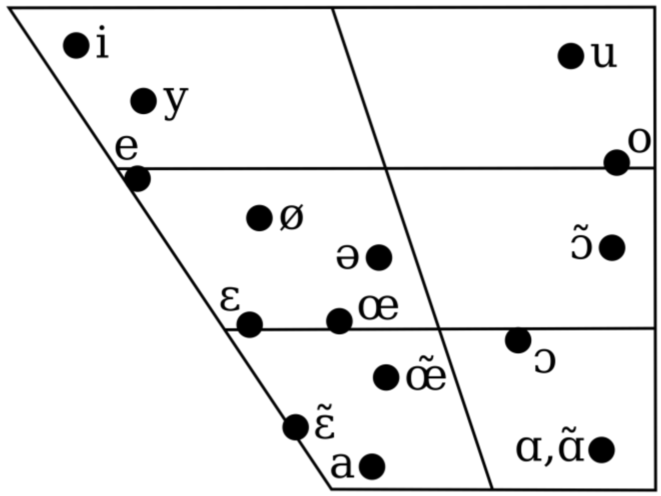
Consonants		Bilabial	Dental	Alveolar	Palatal	Velar	Uvular
Nasal		[m]	[n]		[ɲ]	[ŋ]	
Stop	**voiceless**	[p]	[t]			[k]	
	Voiced	[b]	[d]			[g]	
Fricative	**Voiceless**	[f]		[s]	[ʃ]		
	Voiced	[V]		[z]	[ʒ]		[ʁ]
Lateral				[l]			

2.2.2.1.3.2. Vowels of French

Price (2015, p. 20) describes French vowels as ‘pure’, i.e., the lower jaw of the mouth does not move when pronouncing vowels and are never diphthongized. According to Walter (1976), French vowel system has 16 vowels, with 11 oral short vowels /i, y, e, ø, ɛ, œ, a, ɑ, ɔ, o, u/, one oral long vowel /ɛ:/ and four nasal vowels /ɛ̃, œ̃, ɑ̃, ɔ̃/. Figure 6 presents the standard French vocalic system.

Figure 6

Vowel Diagram of French



Note. From *Practical Phonetics and Phonology: A Resource Book for Students* (p. 225), by B. Collins and I. M. Mees, 2013, London: Routledge.

According to Vaissière (2011), some vowels are to some extent neutralized like /œ̃/ in favor of /ɛ̃/, /ɑ/ in favor of /a/, /œ/ in favor of /ø/, /o/ in favor of /ɔ/ and /e/ in favor of /ɛ/. Vaissière (2011)'s findings were supported by Brkan (2018), De Foy (2019) and Kamiyama (2009). Consistent with the aforementioned views, Fougeron and Smith (1993) report that French vocalic inventory englobes 10 oral vowels (Arabic, e, ɛ, y, ø, æ, a, u, o, ɔ) and 3 nasal vowels (ɛ̃, ɑ̃, ɔ̃).

In fact, the phonetic evolution of the language as well as the low frequency of occurrence of the nasal anterior rounded vowel /œ̃/ and the posterior oral vowel /ɑ/ has resulted in their disappearance in favor of the nasal anterior unrounded /ɛ̃/ and the anterior oral vowel /a/ in France (De Foy, 2019). Brkan (2018) confirms that the neutralization of the contrast /a/ and /ɑ/ does not lead to confusions. However, the contrast between the nasal vowels /œ̃/ and /ɛ̃/ is preserved in Belgium and in Quebec where some minimal pairs are still retained like brun/brin, patte/pâte and l'un/lin, (Brkan, 2018). Yet, these minimal pairs are quite rare (De Foy, 2019).

Concerning the /e ɛ/, /ø œ / and /o ɔ/ contrasts, Kamiyama (2009) states that these vowels are partially neutralized in the final syllable, where according to the law of position of Delattre (1966), mid-close vowels tend to appear in open syllables and mid-open vowels tend to appear in closed syllables. Additionally, the minimal pairs for these vowels are very limited, with the exception of the /e/-/ɛ/ contrast in verbal endings (future /e/ vs. conditional /ɛ/). For that, some phoneticians attributed these archiphonemes /E/, /Œ/and /O/ to describe these sounds respectively: /e ɛ/, /ø œ/ and /o ɔ/ (Lauret, 2007). Lauret (2007) recommends that French L2 learners should be introduced to these archiphonemes before perfecting the mid-closed and mid-open vowels contrasts.

It is to be noted that the official Tunisian education program of French L2 does not teach students to make the difference between /œ̃ ɛ̃/, /ɑ a/, /œ ø/, /o ɔ/ and between /e ɛ/.

2.2.2.1.4. Segmental difficulties for Tunisian learners of French

According to Maume (1973), acquiring French for Arabic learners implies the acquisition of phonetic and phonological systems that are radically different from L1 since French belongs to romance languages and MSA (hence TA) to the Semitic languages. TA and French languages have numerous differences on the phonological level. There are some phonemes found in French that do not exist in TA and vice-versa.

On the consonantal level, TA is a consonantally rich language with 28 consonants whereas French has 17 consonants. Some specific places of articulation that exist in TA do not exist in French such as the uvular sounds [x, q], the pharyngeal sounds [ħ, ʕ] and the emphatic sounds [tˤ, dˤ, ðˤ, ṣ]. Additionally, TA has 13 consonants that do not exist in French and all the consonants in TA can be doubled or geminated (Kandeel & Marcy, 2019). On the other hand, there are two French consonants that do not exist in TA: [ʁ] and [ɥ]. These tow consonants do not present pronunciation difficulties for Tunisians as they can pronounce /ɥ/ by combining [n] and [i] and pronouncing [ʁ] as [r] (Baidoun, 2015).

On the vocalic level, French is vocalically rich language with 13 vowels whereas TA has only 3 vowels. This means that Tunisian learners of French L2 and Arabic learners in general must progress from a system with three vowels to a system comprising 13 vowels, which increases the probability of vocalic errors (Calaque, 1992). Languages with dense vocalic inventories require precise articulation, whereas languages with sparse vocalic systems exhibit significant intra-category variability according to their consonantal environment (Kandeel & Marcy, 2019; Nawafleh & Alrabadi, 2017). Indeed, the openness or height of the tongue is more developed in French than in Arabic. Added to that, nasality does not exist in Arabic. The phonemes that are found in French and that do not exist in TA are the following: / ɑ̃, ɔ̃, ɛ̃, e, ɛ, y, ø, æ, o, ɔ / (Mzoughi, 2015).

Tunisian learners of French usually replace the missing sound in TA by a comparable sound in the French phonological system. For example, the sound /p/ (although it exists) is sometimes replaced by /b/ (for example /pjano/ could be pronounced as [bjano]), the sound /v/ could be replaced by /f/ (for example /vɛst/ is variably pronounced as [fi:sta]), the sound /y/ is replaced by /i/ (for example /sirkylasjɔ̃/ becomes [sirkilasjɔ̃]) and /e/ could be incorrectly pronounced as /i/ (/ekɔl/ becomes [ikɔl]) (Ben Hamida, 2009; Maume, 1973; Mzoughi, 2015). In addition, Maume (1973) points out that Tunisians do not distinguish between [e] and [ɛ], [a] and [ɑ], [œ] and [ø] as the differences between these segments are not taught in schools. For example, the sound /i/ becomes [y] when it is near /y/ and at the same point of articulation (/kɔ̃fityr/ becomes [kɔ̃fytyr]), /e/ becomes /i/ (/telefɔn/ becomes [telifɔn]).

Surprisingly, only a few empirical studies investigated the phonological difficulties for Tunisian learners of French L2. Ammar (2018) examined the production of French vowels by 16 Tunisian children. She found that Tunisian children have difficulties in pronouncing new French vowels /ɔ y ø œ e ɛ ɑ/ and that they have no problems with the identical vowels /a i u/. This was further confirmed in a study by Ben Hamida (2009), who found that due to the interference between TA and French languages, Tunisian learners may have difficulty pronouncing the vowels [y, ɔ, e]. According to the researcher, the sound [y] could be pronounced as [i] ([*invɛʀsite] instead of [ynivɛʀsite]). Also, the sound [ɔ] could be pronounced as [u] ([*boko] instead of [bɔku]) and the sound [e] could be falsely pronounced as [i] as in the word [*ikɔl].

To summarize, it seems that Tunisian learners do not have problems with the identical vowels /a i u/ but rather with all the new vowels /ɔ y ø œ e ɛ ɑ/. However, we have noticed that none of the before-mentioned studies that examined segmental difficulties for Tunisian learners of French have investigated the nasal vowels /ɑ̃, ɔ̃, ɛ̃/. Furthermore, studies on the phonological production of Arabic learners of French from various nationalities revealed that not only the Tunisian learners have difficulties with the new French vowels / ɑ̃, ɔ̃, ɛ̃, e, ɛ, y, ø, æ, o, ɔ/ but also from other Arabic nationalities. These studies were conducted by Abusharifa (2016) (30 Palestinian participants), Baidoun (2015) (18 participants from Maroc, Libanon, Tunisia, Syria, and Yemen), de Vos (2017) (5 Sudanese participants) Nawafleh (2016) (10 Jordanian participants).

2.2.2.2. L2 Suprasegmentals

The suprasegmental features differ fundamentally from segmentals (Hunag & Jun, 2011). Suprasegmentals or prosody are two terms that could be used interchangeably (Isaacs & Trofimovich, 2016). Suprasegmentals are not limited to segments but also operate on syllables and larger units (words and phrases) (Ghanem & Kang, 2018). The acoustic features of suprasegmentals could be measured in terms of duration, intensity, and fundamental frequency (f0).

Suprasegmentals play a crucial role in L2 pronunciation perception (Baills, Alazard-Guiu, & Prieto 2022; Levis & Silpachai, 2022) as they offer the framework that connects individual sounds and transmits the linguistic and paralinguistic meanings, i.e., messages and intentions (Kallio, 2022). Prosody also proves learners' capacity to use correct word stress or the use of appropriate declarative and interrogative intentions in a sentence. Likewise, suprasegmentals indicate learners' ability to deliver speech without many pauses or corrections at an optimal tempo (Saito & Plonsky, 2019, p. 8). The accurate use of the suprasegmental features of L2 has been found of prime importance in ensuring intelligibility and comprehensibility (Hahn, 2004; Kang, 2010; Munro & Derwing, 1999). Some researchers even argued that suprasegmentals should be given the priority in pronunciation instruction where native listeners rely on stress pattern for example more than on the segmental features (Fraser, 2001; Tanner & Landon, 2009). Zielinski (2015) argued that segmental and the suprasegmental features of L2 are equally important where they should be viewed as part of an integrated and interactive system that influences each other. Nevertheless, despite the importance of suprasegmentals on L2 production, studies examining the acquisition of L2 suprasegmentals is still scarce (Parlak, 2018).

Suprasegmentals include stress, rhythm, and intonation (Saito & Saito, 2017). Since this dissertation only targets stress (see section 2.2.2.2.1), the other two suprasegmental features, intonation and rhythm, will only be briefly discussed.

2.2.2.2.1. Stress

Stress is defined as the relative prominence of a syllable in comparison to the neighbouring syllables of the same word as a result of being more audible. Articulatorily, stress is generated by an increased physical effort of the laryngeal muscles exerted through the articulatory organs to enhance subglottal air pressure and vibration of the vocal cords (Parlak, 2018; Reetz & Jongman, 2011). Acoustically, stress is also defined as the syllable in any given word or an utterance with the highest values of pitch (measured in Hz), length (measured in milliseconds) and intensity (measured in dB) (Ghanem & Kang, 2018). A stressed syllable is characterized by one or more of these acoustic features. The impact of these acoustic features on the morphological elements varies depending on the language (Fox, 2000).

There are three levels of stress: primary stress, secondary stress, and unstressed (Hayes, 1995). Primary stress involves the principal pitch prominence in the word. Primary stress is marked with a short-raised stroke [ˈ]. The secondary stress involves the subsidiary pitch prominence. It is marked with a short-lowered stroke [ˌ]. The unstressed level involves the non-prominent syllables that have no pitch change. As an illustration, in the word "phonetician" [ˌfoʊ.nə.ˈtɪ.ʃn̩], the first syllable receives a secondary stress, the second and fourth syllables are unstressed, and the third syllable has the primary stress (Reetz & Jongman, 2011).

Stress is essential for language processing and acquisition as it helps in the location of word boundaries and for word recognition (Skoruppa et al., 2013). In variable stress languages or stress-timed languages such as English and Spanish, stress is variable across words, and they may convey differences in meaning. Contrastingly, in fixed stress languages, stress always falls on the same position. For example, stress always falls on the last syllable in French, on the first syllable in Hungarian and Czech and on the penultimate syllable in Polish. When a stressed syllable in French is heard, then the listener will know that it is the end of the word whereas when a stressed syllable is perceived in Czech, the listeners know that it is the beginning of the word. Yet, in English, stress can occur on the antepenultimate, penultimate or final syllables, consequently, word delimitation in English cannot be applied.

On the sentence level, the alteration of stress creates rhythms. The position of strong stress varies depending on the stress patterns of the nearby words (Betti & Ulaiwi, 2018). In stress-timed languages, the linguistic rhythm is based on the isochrony of interstress intervals, i.e., the intervals between the syllables have the same duration (Reetz & Jongman, 2011). The unstressed syllables shorten to fit into the rhythm. Learners who have a syllable-timed language as an L1 usually find multisyllabic words as problematic when defining the stressed syllable. On the other hand, in the syllable-timed languages, all the syllables (stressed and unstressed) have approximately the same regular intervals and the same duration to pronounce. Function words are usually much more prone to such strong–weak alternations than content words.

In the context of this dissertation, the specific pronunciation proficiency for TA learners of French L2 will be investigated. Consequently, the suprasegmental features of TA and French will be presented in the following sections. Furthermore, an explanation of why stress was chosen to be studied will be provided. However, in reviewing the literature on the suprasegmental properties of Arabic in general and TA in specific, we were surprised by the modest number of studies compared to French. Indeed, Odisho (2005) asserted that the prosodic features of Arabic, particularly rhythm and intonation, are rarely covered since they are unrecognized as a key component of the language, with the primary attention being on segmental phonetics. Added to that, Shifflett (2011) stated that the Tunisian dialect is one of the less frequently investigated Arabic dialects. Accordingly, when addressing the suprasegmental features of Arabic in the following section, they also apply to TA.

2.2.2.2.2. Rhythm

Rhythm refers to the organization of timing in speech and expresses the relationship between the stressed and unstressed syllables and their pattern of combination. It refers also to the rate at which words, phrases and sentences are spoken in order to shape the overall intonation of speech. The acoustic correlates of speech rhythm are based on distinct phonetic durational units among syllables, the voiced and unvoiced intervals, and amplitude peak intervals (Ibrahim, Asadi, Kassem, & Dellwo, 2020).

Rhythm varies across languages. O'Connor (1973), for examples, states that distinguishing languages in terms of rhythm such as syllable-timed languages and stress-timed is crucial for uncovering their prosodic behaviour. Syllable-timed languages (such as Spanish, Italian, and French) induce a strong perception of equal prosodic saliency across syllables, where they assign almost equal intervals of time to syllables. In contrast, in stress-timed languages (such as English and Dutch), the stressed syllables are louder, longer, and higher than unstressed syllables, which are quieter, lower, and shorter, giving the impression of a Morse rhythmic effect (Prieto & Roseano, 2018).

2.2.2.2.3. Intonation

Intonation is described as "the melody of an utterance" (Prieto & Roseano, 2018, p. 217). It was defined by Gussenhoven (2004) as "the use of phonological tone for non-lexical purposes, or - to put it positively - for the expression of phrasal structure and discourse meaning" (p. 12).

Intonation is the variation of pitch (rising or falling, dipping or peaking) in the spoken language at the phrase or sentence level. Intonation is related to multiple acoustic features and perceptual phenomena that englobes three systematic variables: tonality (grouping utterances into prosodic phrases), tonicity (the prosodic prominence distribution), and tone (pitch contour around the prominences) (Liu & Reed, 2021). Intonation plays an important role in the meaning of an utterance as it is closely related to information structure, morphosyntactic structure, and the pragmatic functions (Levis & Wichmann, 2015). It could for example convey completeness, doubt, astonishment and so on.

2.2.2.2.4. Suprasegmental features of TA

2.2.2.2.4.1. Stress

According to Alqattan (2015, p. 105), Arabic stress pattern is a knowledge gap in the current literature. Arabic is a stress-timed language where stress falls on a particular location. Stress in Arabic is divided into four levels: primary, secondary, middle, and weak (Betti & Ulaiwi, 2018). Moreover, stress is used to show the construction of form and to express different meanings (Betti & Ulaiwi, 2018).

The position of stress is predictable in all Arabic dialects (Hellmuth, 2018) and is sensitive to syllable wight. Khalifa (2020) defined a light syllable as a simple peak of vowel preceded by a consonant (CV), whereas a heavy syllable is composed of consonant in onset and branching rime (CVC or CVV), and the super-heavy syllable comprises a heavy syllable plus a consonant (CVVC or CVCC). The super-heavy consonant is usually limited to the final position Khalifa (2020).

All dialects of Arabic follow the same pattern of stress placement: stress falls on the final super-heavy syllables (e.g., [da.'xalt] ‘I went in’) or falls on the penultimate (second to last) heavy syllable (e.g., ['mak.tab] ‘an office') (Hellmuth, 2018; Khalifa, 2020). If the word has no long vowels or final super-heavy structure and consists of two syllables, stress then falls on the antepenultimate (CV) (e.g., ['da.wa] ‘medicine’) (Mustafawi, 2018, p. 21). Likewise, in case the word has open syllables and short vowels (CVCVCV(C)), stress falls also on the antepenultimate (Mustafawi, 2018).

To recapitulate, Betti and Ulaiwi (2018) describe the following stress placement rules:

i) When the word consists of a sequence of syllables /cv/ the first syllable carries primary stress and other syllables carry weak stresses.

ii) When a word includes a long syllable /v:/, this syllable receives primary stress and other syllables receive weak stresses.

iii) When a word consists of two long syllables or more, the syllable which is near to the end of the word receives primary stress and the long syllable which is near the beginning of the word receives secondary stress. (p.85)

The same authors also summarize the general stress rules in Arabic (therefore in TA) as follows:

i). Stress is on the first syllable, no matter what that first syllable is like (strong or weak) e.g., 'ilaa (to, toward).

ii) Stress falls on the second syllable (the penult) if the syllable is strong (CVV) e.g., ju'huudun (efforts).

iii) If the second syllable of the word is weak (CV), then stress is on the third syllable (the antepenult) e.g., aa'Simatun (a capital). Stress falls on the syllable which includes a long vowel whether the syllable is in initial 'qaala (he said), medial yu'qaabil (he meets), or final position hu'maa (they both). of the word. In the absence of a long vowel in a word, stress is generally on the first syllable except when germination occurs in the middle of the word e.g., 'naHnu (we). (p.85)

On the phrasal level, in Arabic and TA stress falls on noun-adjective phrases. However, on the sentence level, stress varies according to the emphasis placed on a particular lexical term (Betti & Ulaiwi, 2018). This was confirmed by Hellmuth (2018) who stated that the sensitivity of stress placement varies across dialects. For example, /katabahu/ “he wrote it' is said to be produced with penultimate stress in Egypt, but with antepenultimate stress in Lebanon and Jordan (de Jong & Zawaydeh, 1999).

On the word level, stress placement could also vary according to dialects. As MSA never stresses the final syllable of the word even when a heavy syllable is in the word-final position (Mustafawi, 2018, p. 20), Hellmuth (2018) stated that the super-heavy syllables could be stressed in the final position in some dialects. Indeed, according to Ghazali (1973), stress falls on the word-final position when it is long like in the word [hnaa] ‘here’ and [mak.ˈtba] ‘library’. The difference between Arabic and TA in stress placement is due to the nature of the vocalic system of TA that tends to delete vowels from syllables (/qa.di:m/ ‘old’ in MSA and /qdi:m/ in TA).

Other stress features in TA were revealed by Bouchhioua (2009) who conducted a study with seven Tunisians. Bouchhioua (2008, 2009) found that the duration and intensity are not cues to stress in TA and that the only phonetic characteristics that are involved in lexical stress are F0, spectral balance and F1 lowering.

2.2.2.2.4.2. Rhythm

Arabic is a stress-timed language. Research examining Arabic rhythm is scarce compared to other languages (Droua-Hamdani, Selouani, Boudraa, & Cichocki, 2010). According to Ghazali, Hamdi, and Barkat, (2002) and Odisho, (2005), the perception of different rhythms in

Arabic is mainly correlated with stress placement, the shift in vowel quality and quantity, and syllable structure. The north African Arabic is characterized by a faster rhythm than middle eastern Arabic (Ghazali et al., 2002). In particular, TA is known for short vowel deletion in open syllables, resulting in a variety of consonant clusters and syllable forms with complex onsets and codas (Ghazali et al., 2002).

2.2.2.2.4.3. Intonation

As previously defined, intonation is the pitch variation. Hellmuth (2013) stated that there is a descriptive gap in the field of Arabic intonation as there has been little work done on the description of intonational phonology for the various Arabic dialects. Intonation in Arabic is used to indicate a conjunction of the syntactic, semantic, and attitudinal characteristics of an utterance (Odisho, 2005).

The majority of Arabic dialects share the same overall nuclear contours: rise, fall, and level (Bani Younes & Hellmuth, 2020). Chahal (2007) stated that all Arabic dialects use boundary tones to mark the intonational phrases and that pitch accents are correlated with stressed syllables. Bani Younes and Hellmuth (2020) added that the majority of Arabic dialects have a fall in the declaratives and a rise in the yes-no questions at the end of the utterances. However, the Maghreb dialects differ from the other dialects by the accentuation of content words where the Eastern dialects accentuate lexical words whereas Maghreb dialects do not (El Zarka, 2017).

2.2.2.2.5. Suprasegmental features of French

2.2.2.2.5.1. Stress

The stress system of French is not like any other Romance languages. According to Delais-Roussarie et al. (2015), the metrical patterns in French are based on the occurrence of two separate stress types allocated to the phrasal level: the mandatory phrase-final stress (final accent) and the optional phrase initial stress (initial accent).

The location of the primary rhythmic stress is fixed as it falls on the final full syllable of a prosodic phrase (Côté, 2022; Fagyal et al., 2006). French has no tonal segments nor metrical structure (Di Cristo, 1998). According to Walker (2001), unstressed syllables are given equal importance within the phrase as they are uttered within a regularity and balance that prevent vowel reduction and prominence of the syllable at the end of the phrase.

Stress in French is not used to differentiate between the meaning of words, i.e., does not have a distinctive function, but it can serve to determine the meaning of sentences. There are two broad types of stress in French: rhythmic and emphatic (Côté, 2022; Price, 2005). The emphatic stress is the "accent affectif" and it is defined by giving a greater degree of prominence as a way to express a kind of emotion or reaction. The rhythmic stress is characterized by a greater prominence as well to a given word to highlight its meaning.

Stressed syllables are prominent, and they are characterized by durational lengthening compared to the other unstressed syllables (Delais-Roussarie et al., 2015; Jun & Fougeron, 2000; Vaissière, 1983). This phenomenon is called final lengthening (Fagyal et al., 2006). They are also characterized by loudness, pitch, and quality (Roach, 1991). These characteristics do not have an equal importance as the strongest effects for stress in French are pitch and length.

2.2.2.2.5.2. Rhythm

French is a syllable-timed language and, as such, all syllables within the phrase are given equal significance as they lead to phrase-final stressed syllables (Walker, 2001, p. 177). A significant feature of French is that, regardless of how quickly or slowly the individual speaks each syllable regardless it is stressed or not, takes roughly the same amount of time, i.e., a constant rate of syllable succession (Price, 2005, p. 44)

According to Price (2005), there are three types of groups: the breath group, the sense group and the rhythmic group. The breath group relates to everything uttered between two breath intakes and it could be longer than a sentence. The sense group corresponds to the grammatical sub-divisions within the sentence. The rhythmic group is very important for French as it is the minimal unit of prosody. There are no clear rules to divide the sentences into rhythmic groups, but they are usually short, normally three or four syllables, and seldom more than seven syllables. The last syllable of the rhythmic group is pronounced with the most stress. Yet, Price (2005) provided some guidelines to divide the sentences into rhythmic groups:

(i) The rhythmic group may or may not be followed by a pause.

(ii) Wherever there is any kind of normal pause between breath groups and/or sense groups, there is a division between rhythmic groups (pauses for hesitation, as in Je vous ai . . . euh . . . sous estimé, are excluded from this definition).

(iii) The rhythmic group coincides with one or more sense groups.

(iv) Constructions such as the following can never be split, i.e. they must fall within the same rhythmic group:

(a) preposition + noun phrase, pronoun, infinitive, etc., e.g. pendant quelques instants; avant de partir;

(b) adjective + following noun, e.g. cette agréable surprise, une charmante petite fille (but adjectives following the noun may – but not necessarily – form a different rhythmic group, e.g. c'est un étudiant | très intelligent; il a des problèmes | insurmontables);

(c) subject and object pronouns and their verb, e.g. je vous les donne; envoyez-le-moi (but a noun subject or object may – but not necessarily – form a separate rhythmic group, e.g. mon fils aîné | arrive demain or ils écrivent | plusieurs lettres);

(d) auxiliary verb + past participle, e.g. nous avons fini; (mon frère) est parti; (mon frère) sera choqué. (p. 38)

According to Price (2005), the number of rhythmic groups in a particular sentence increases as the individual speaks more slowly.

2.2.2.2.5.3. Intonation

French is characterized by the unpredictability of intonational patterns (Price, 2005). The declarative sentences start with a relatively high pitch and ends with a lower one. If the declarative sentence contains more than two rhythmic groups; the intonation rises at the first group and then falls at the second one. Wh-questions and imperative sentences are also characterized with a falling intonation at the end of the utterances.

According to Fagyal, Douglas, and Jenkins (2006), each utterance contains an underlying metrical structure, which is a regular pattern of stressed and unstressed syllable interchange. This structure provides a 'docking sites' for two tonal events to be physically realized: the pitch accents and the boundary tones. An intonation occurs when moving from one tonal event to another.

2.2.2.2.6. Suprasegmental difficulties for Tunisian learners of French

According to Cheddad (2011, p. 56), the intonation and rhythmic systems of Arabic and French are nearly identical: they have a linguistic and expressive role. In this regard, intonation and rhythm will not be examined in this dissertation.

Concerning stress placement, there is evidence of L1 transfer effects of stress placement into L2 according to learners'L1 stress placement norms (Almbark, Bouchhioua, & Hellmuth, 2014). Indeed, Embarki, Abou Haidar, Zeroual, and Naboulsi (2016) confirmed that Arabic speakers mirror the stress placement system of their dialectal L1 on the production of French where instead of placing stress on the last syllable of the rhythmic group, they place a high pitch and increase intensity of the first syllable of the word.

Furthermore, according to Price (2005), stress in French is considered as relatively weak, as a consequence, L2 learners whose L1 belong to stress-timed languages (e.g., Arabic) tend to not only stress the correct syllable but overstress it. As French is a syllable-timed language and TA is stress-timed language, this dissertation will investigate only stress placement.

After examining the global and specific dimensions of L2 pronunciation, the following section deals with the effects of various task types and the main theories that are related to task types.

3. Task type effects and L2 pronunciation

3.1. Task types

Isaacs and Trofimovich (2016, p. 9) define the term assessment as "the process of information gathering (e.g., about an L2 learner's or test taker's ability), potentially from multiple and varied sources on the variable(s) of interest, including generating information about what learners can do to feed into the teaching cycle". According to these authors, assessments and tests do not have the same definitions as all tests could be assessments but not all assessments are tests. They define tests as:

> […] a particular type of assessment involving the elicitation of an L2 learner's or test taker's performance followed by inferences or decision making on the basis of that

performance, generally informed by a test score or a numerical indicator from a score report. (p.9)

Banerjee, Lestari and Rossi (2020) define task type as “[t]he activities or tasks that learners have to do in a test.” (p.78). The impact of task on performance is an important variable to consider in assessment situations (Harding, 2018). Indeed, Saito and Plonsky (2019, pp. 13-14) indicated that there is no single best assessment approach for evaluating L2 pronunciation proficiency. Yet, researchers have frequently relied on a single speaking task to assess linguistic dimensions, using picture description task to measure comprehensibility for example (Inceoglu, 2019). Nevertheless, Saito and Plonsky (2019) advocate the use of controlled and spontaneous tasks to capture learners’ L2 pronunciation knowledge under different speaking contexts. In fact, Bergeron and Trofimovich demonstrated that accentedness and comprehensibility can be differentiated in a picture description task but not in an interview task.

According to Saito (2015), controlled tasks such as word, sentence and paragraph readings are used to highlight certain features in L2 speech production such as segmental accuracy. Controlled tasks are designed to project the explicit and controlled declarative knowledge of the learner, i.e., when learners highly control and monitor their accurate use of the target pronunciation proficiency features by accessing metalinguistic knowledge (Saito & Plonsky, 2019, p. 14). The advantage of controlled tasks is that they ensure similar speech content from all the participants, which allows direct comparison.

In contrast, Spada and Tomita (2010) defined extemporaneous tasks as a type of free speech in which participants deliver their intended message with no room for conscious monitoring. Saito and Plonsky (2019) also defined the spontaneous pronunciation as “the degree of L2 learners’ access to explicit pronunciation knowledge in a timely manner while their primary focus is on meaning rather than form" (p.13). In other words, spontaneous tasks mirror learners' unconscious and unmonitored use of L2 pronunciation and how much they can access their internalized and automatized knowledge. Spontaneous tasks are representative of the real-world speech. Throughout these tasks, learners will pay simultaneous attention to grammatical, phonological, lexical and pragmatic aspects of the target language and will have less time to access their explicit articulatory knowledge. As an illustration, Saito and Brajot (2013) reported that L2 learners tend to make more pronunciation errors in free speech tasks than in word reading tasks. The most common forms of spontaneous tasks are picture naming, picture narrative, timed picture description tasks and interviews (Saito & Plonsky, 2019). However, the downside of using spontaneous tasks is that it is difficult to control the lexeme used by the speakers (Kamerhuber, Horvath, & Pustka, 2020).

Several theories have been proposed to explain the relationship between learners' performance and task complexity. Three main theories that will be presented in the following section: the Cognitive Load Theory, the Limited Attentional Capacity Model, and the Cognition Hypothesis.

3.2. Theories related to task complexity

3.2.1. Cognitive Load theory

With the less restrictive speech task type, a considerable proportion of input would take a flowing speech style (Hannah, Kim & Jung, 2022). Contrastingly, in the complex speaking tasks, speakers will have to deploy all the available linguistic resources (Crowther, 2018). Actually, tasks with different cognitive loads may result in varying performance. The heavier the cognitive load in a speaking task, the greater the number of mistakes in pronunciation (Morley,1994, as cited in Lan, 2014). The Cognitive Load theory (Sweller, 1988, 2011) posits that the human cognition process englobes the working memory and the long-term memory. The short-term memory has a limited capacity and can handle information effectively at once (Mestre, 2012). If the working memory is overloaded, the learner may not be able to process the information properly in the long-term memory, i.e., cognitive overload occurs (Mestre, 2012). In other terms, the learning outcomes are limited to the capabilities and limits of the human cognitive processing.

The cognitive load theory also suggests that task performers face three types of demands when participating in instructional activities. The first type is the intrinsic cognitive load and is associated with the inherent complexity of the task and is typically influenced by the learner's familiarity with the task and his mental ability. The second type is the extraneous cognitive load which is subject to teaching approaches and material design. It is the result of the load generated by a poorly designed content and does not contribute to the learning process. Finally, the third type is the germane cognitive load which is related to the elements that help information processing and might be adjusted by task designers. This type is beneficial in that it encourages further learning and for the storage of schemata in the long-term memory (Papi, Vasylets, & Ahmadian, 2022).

3.2.2. Limited Attentional Capacity Model

The Cognitive Load theory and the Limited Attentional Capacity Model (or Trade-off Hypothesis) share much in common (Fisch, 2017). Skehan and Foster's (2001) Limited Attentional Capacity Model stipulates that the increase in the cognitive task complexity will lead learners to pay attention on the content and the meaning of the task over language form which will result in trade-off effects, i.e., affecting the complexity, accuracy and fluency of the input (Michel, 2017). Learners cannot process multiple resources simultaneously in complex tasks.

The hypothesis also indicates that when a learner focuses on one aspect of pronunciation, this impedes the development of the other pronunciation constructs (Kuiken & Vedder, 2012).

3.2.3. Cognition Hypothesis

In contrast to the Cognitive Load theory, the Cognition Hypothesis (Robinson, 2001a, 2001b, 2005, 2011) proposes another prediction about the attentional demand of tasks. The hypothesis cites that increasing task complexity "has the potential to connect cognitive resources, such as attention and memory, with effort at conceptualization and the L2 means to express it" (2011, p. 14). In other terms, highly complex tasks push speakers to simultaneously achieve more structural complex and greater accurate production. Robinson (2001a, p. 29) defined task complexity as "the result of the attentional, memory, reasoning and other information processing demands imposed by the structure of the task.

The Cognition Hypothesis also seeks to provide a sequencing for tasks by suggesting that tasks should be sequenced from simple to complicated in terms of the attentional, memory, reasoning and cognitive loads, and that this sequencing will encourage learners to accomplish the tasks (Robinson, 2022).

Also, the Cognition Hypothesis distinguishes between two dimensions of task complexity: resource-directing and resource-dispersing. The resource-directing dimension of task complexity is related to the direct attention to the numerous dimensions of L2 that can be employed to perform them whereas the resource-dispersing dimension of task complexity is related to attentional and cognitive demands but does not direct attentional or memory resources to any components of L2 that can be used to complete the task (Robinson, 2022). Additionally, Robinson (2011) suggested that there are factors that decide the intrinsic cognitive demands of tasks: complexity (cognitive factors), difficulty (learner factor), and conditions (interactional factors).

3.3. Task effects

Recall that the use of different task types reflects learners' ability to adjust their speech to situational formality, i.e., their stylistic variation according to tasks (Labov, 1972). In other words, through the controlled tasks, learners will demonstrate their knowledge in a formal (in class) context, whereas in extemporaneous tasks learners will elaborate how much they can use their knowledge in non-formal contexts (outside class) (Saito & Plonsky, 2019). The reduced attention to speech corresponds to the use of a less conscious and spontaneous style while the increased attention to speech leads to a more explicit, analyzed and conscious style. Munro and Derwing (2015) advocated combining different methods to produce comparable and accurate data. Additionally, aside from altering learners' performance, using different task kinds alters listeners' assessments of comprehensibility and accentuation (Crowther, 2018; Inceoglu, 2019).

Yet, according to Nagle (2022, p. 276), although several studies have addressed the global and specific features of pronunciation using controlled and spontaneous speaking tasks, few studies have done so within a single task. Consequently, it would be important to fill in this research gap by using controlled and spontaneous measuring tasks with varying degrees of cognitive loads.

4. Conclusion

The first part of this chapter reviewed the theoretical and empirical frameworks regarding the age effect studies and SLA. Then, the second part examined the dimensions of the global and specific L2 pronunciation, particularly of TA and French. Finally, the third part introduced task types (i.e., spontaneous and controlled tasks) and their effects on learners' outcomes. The following chapter presents the methodology used to answer the research questions of this dissertation. To reiterate, the aim of this thesis is to fill in the gaps of the previous research by examining the effects of age (4 years old vs 8 years old) and exposure (1600h vs 1000h) on the pronunciation of Tunisian learners of L2 French in the school setting across four task types (ranging from controlled to spontaneous).

Chapter 3

Research Method

This chapter presents the methodology employed to investigate the proposed research questions. It begins by stating the characteristics of participants and listeners. This is followed by a presentation of the instruments designed and used to elicit data. Thereafter, the procedures adopted to carry out participants' recruitment and data collection are explained. Information about coding and the scoring procedures are discussed in the final section of this chapter.

1. Research design

In order to accomplish the objectives of the research, a quantitative correlational approach was implemented. Quantitative research is defined by Brock (2017) as the "methodology used to determine whether, and to what degree, a relationship exists between two or more variables within a population (or a sample)" (p. 2). In quantitative research, data are analyzed using statistics as a mathematical tool to uncover casual-like correlations (Phakiti, 2014). Thus, this approach will allow quantifying the variables to determine the effects of age of acquisition (AoA) and exposure to the TL on the global and specific pronunciation of the three groups across different task types.

2. Variables

There are two independent variables (IV) and five dependent variables (DV). The independent variables are two: (1) the number of hours of exposure to French, and (2) age of acquisition. These IVs were measured using a questionnaire (appendix B) that was distributed and filled by the parents of the child participants. The dependent variables were the levels of proficiency the pronunciation accuracy of segments, stress placement accuracy, comprehensibility, intelligibility and accentedness. Segmental pronunciation accuracy was measured through word list reading, sentence reading, picture identification and storytelling tasks (see forthcoming discussions for details). It was analysed through auditory perceptual analyses by the researcher and an expert in phonetics. Word stress accuracy placement was measured through word list reading, sentence reading, picture identification and storytelling tasks. This variable was analysed through acoustic analyses (PRAAT). Comprehensibility was measured through listeners' judgments in sentence reading and storytelling tasks. A 9-point Likert scale was used to evaluate the degree of comprehensibility. Accentedness was measured in the exact same way as comprehensibility. Finally, intelligibility was measured using the sentence reading and storytelling tasks by expert listeners who transcribed orthographically the auditory extracts and then the mean of correctly transcribed words is calculated.

Another variable that was examined was the effect of the tasks utilized in the research. It was the moderator variable that will help us understand the strength of the relationship between the dependent variables and the independent variables, in other words, whether AOA and exposure could have stronger or weaker effects on general and specific pronunciation through spontaneous and controlled tasks. Table 9 summarizes the variables and the instruments adopted in the current research.

Table 9

Variables and Instruments of Measurement

Variables	Instruments
IV: AOA	Questionnaire
IV: Number of hours of exposure to French L2	Questionnaire
DV: segment pronunciation accuracy	Word list reading Sentence reading Picture identification Story telling
DV: stress placement accuracy	Word list reading Sentence reading Picture identification Story telling
DV: comprehensibility	Sentence reading Story telling
DV: intelligibility	Sentence reading Story telling
DV: accentedness	Sentence reading Story telling

3. Participants

3.1. Speakers

Sixty-eight Tunisian learners of French L2 participated in this study. Participants were enrolled in twelve private and public schools in the region of Sfax, Tunisia. It is to be noted that the study was meant to be conducted in four regions in Tunisia: Sfax, Tunis, Sousse and Gabes.

However, due to the COVID-19 pandemic, it was decided to conduct the study in the researcher's hometown city Sfax for safety reasons.

Participants were classified into three groups according to their AOA and the amount of L2 exposure:

- Group 1 (G1) included 23 participants (11 girls 12 boys) that began learning French at the age of four and had approximately 1600 hours of exposure to French. All the participants were enrolled in five different private schools. Their mean age at testing is 12 years old (in the sixth grade).
- Group 2 (G2) included 24 students (13 boys 11 girls) that started learning L2 at the age of eight and had approximately 1000 hours of exposure to the target L2. All the participants were enrolled in four different public schools. Their mean age at testing is 12 years old (in the sixth grade).
- Group 3 (G3) was composed of 21 students (11 girls 10 boys) that started learning L2 at the age of four (same AOA as in G1) and have had 1000 hours of exposure to French (same amount of exposure as in G2). Participants were enrolled in five private schools. Their mean age at the time of the study was 8 years old (in the third grade).

To control and estimate the number of hours of extracurricular exposure to the target language, a questionnaire was administered to collect data concerning their language learning experience (e.g., starting age, frequency of contact with L2 speakers, frequency of use of French outside schools, time spent abroad). Participants were considered eligible to take part in the study if they meet the following criteria: firstly, they do not present any pronunciation anomalies, second, their parents are native speakers of Arabic; and last, they do not use French as the language of communication outside school. The participants had different levels within each group to ensure the generalizability of the results. Table 10 summarizes the characteristics of all participants.

Table 10

Characteristics of the Participants

	G1	**G2**	**G3**
Number	23	24	21
AOA	4 years	8 years	4 years
Age at testing	12 years old	12 years old	8 years old
Number of hours of exposure to French	1600	1000	1000
School-level	6th grade	6th grade	3rd grade
L1	Arabic	Arabic	Arabic
Number of hours of extracurricular exposure	similar	similar	similar

The oral productions of G1 and G3 were compared as they have the same starting age (4 years) but different amounts of exposure to French (1000h and 1600h). The oral productions of G2 and G3 were compared as they have the same number of hours of exposure (1000h) and different starting ages (4 years and 8 years). Participants' age at testing is 12 years and 8 years old.

Participants were asked to read a list of forty words, read four sentences, identify forty pictures, tell the story of *the tortoise and the hare* and describe a picture.

3.2. Listeners

Two experienced Tunisian teachers of French and one Tunisian student at the Department of Education at UQAM University (a francophone university) participated in the study as listeners or speech raters. The mean age of the French teachers is 60 years old, and the age of the third listener is 30 years old. The listeners were chosen to be native speakers of Tunisian because the purpose of this study was not to evaluate participants according to native-like level. The two teachers had an experience in teaching French L2 in Tunisian public schools for a period of 30 years, while the third listener has background knowledge in French phonetics.

The listeners had two main tasks: the first task consists in listening to 136 excerpts from the oral productions of speakers in test 2 and test 4 and transcribe them orthographically. The second tasks consisted in listening to 136 audio extracts from tests 2 and 4 and rate the degree of accentedness and comprehensibility using a 9-point Likert scale that was used in the studies of Munro and Derwing (1999) and Munro et al. (1996). Due to the large number of tasks, the rating process was over a period of three weeks.

4. Pilot study

A pilot study was conducted to test some preliminary aspects of the research to ensure data collection and management procedures could be feasibly carried out. The instruments were tested so we could find and ameliorate any flaws in the research design before conducting the study. The pilot phase allowed us as well to measure the total duration of the tests (approximately 15 minutes) and to identify which sounds were problematic.

According to the previous studies examining segmental difficulties for Tunisian learners of French (Ben Hamida, 2009; Maume, 1973; Mzoughi, 2015), Tunisians have difficulties with vowels that do not exist in their L1 vocal repertoire (/e, ɛ, o, ɔ, y, ø, œ, ɑ̃, ɔ̃, ɛ̃ /) and with the /p v/

consonants. To narrow down the number of segments to be studied, a pilot study was conducted with 20 students over a period of two months. The participants were 12-year-old students enrolled in the same school and they started learning French at the age of 9. They were asked to identify 54 pictures in a picture identification task. The 54 words comprised six words for each of the problematic vowels (/ɑ̃, ɔ̃, ɛ̃, ɔ, y, e, œ/) and consonants (/v, p/). We listened carefully to all the utterances and a qualitative analysis has been conducted. As a result, participants pronounced the consonants accurately and they managed to pronounce correctly the /ɔ̃/, /ɔ/, and /œ/ vowels. These results contrast with previous studies (Ammar, 2018; Ben Hamida, 2009). Participants had mainly difficulties with two nasal vowels /ɑ̃/ and /ɛ̃/ and two oral vowels /y/ and /e/. Consequently, we amended the original version of the tests and only four vowels (/ɑ̃/, /ɛ̃/, /y/ and /e/) were used when examining segmental pronunciation in this research.

5. Instruments

To provide answers to the research questions, speech data were collected using two controlled and two spontaneous speech tasks. These tasks were designed taking into consideration the young age of the participants and their academic level.

5.1. Test 1: Word list reading

The first task (appendix H) that was given to participants was a word list reading task, which is a controlled task that allows learners to strongly focus on their pronunciation. Based on the pilot study, students had difficulties with four vowels: two oral vowels /y, e/ and two nasal vowels /ɑ̃, ɛ̃/. In total, 40 words (ten words for each target sound) were presented in this task for participants to read.

The choice of words for this task was very limited as we had to select words that are easily identifiable by the participants. Also, taking into consideration the participants' young age and to reduce their stress level, the list of words was accompanied by pictures. The list of words was given to a third-year French teacher who confirmed the ability of 8-year-old students to read them. Furthermore, there are no words that had been borrowed into Tunisian to diminish the effect of L1. For example, the word ‘ambulance’, though it comprises two target sounds, was not presented to participants as it is pronounced in everyday use in TA as [ɑ̃bilans].

The following words were presented randomly to the participants:

- The [ɑ̃] sound: ampoule, pamplemousse, kangourou, tambour, température, ange, blanc, jambe, serpent, argent
- The [ɛ̃] sound: dindon, interdit, peinture, intelligent, important, main, raisin, faim, poussin, magicien
- The [y] sound: urgence, uniforme, allumette, tulipes, jumeaux, laitue, voiture, chaussure, légume, cube
- The [e] sound: échelle, élève, étoile, zéro, été, blé, idée, couper, dîner, fée

5.2. Test 2: Sentence reading

The second controlled task (appendix I) was a read-aloud test where participants were asked to read some sentences that comprised the target sounds. During this task, participants paid less attention to the target words than in the previous task. The sentences were created by us while taking into consideration the young age of the participants. All words in the sentences were taken from the French manual of the third-year students. A third-year teacher confirmed the ability of students to read the sentences with no difficulties. In these sentences, [ɛ̃] is repeated 10 times, [ɑ̃] is repeated 7 times, [y] is repeated 6 times and [e] is repeated 6 times. The target sentences are presented below:

- Dans le jardin de Martin se trouve des dindons, un poussin qui mange de la confiture, un lapin qui boit du lait et un chien qui joue avec sa poupée.

[dɑ̃ lə ʒardɛ̃ də martɛ̃ sə truv de dɛ̃dɔ̃ œ̃ pusɛ̃ ki mɑ̃ʒ də la kɔ̃fityr œ̃ lapɛ̃ ki bwa dy lɛ e œ̃ ʃjɛ̃ ki ʒu avɛk sa pupe]

- Le dimanche, Alain vient en vélo pour voir la tortue. Il dit : Lulu la tortue, au dîner, veux-tu des raisins, des caramels ou une laitue?

[lə dimɑ̃ʃ, alɛ̃ vjɛ̃ ɑ̃ velo pur vwar la tɔrty. Il di: lyly la tɔrty, o dine, və ty de rɛzɛ̃ de karamɛl u yn lɛty]

- Le dentiste avec l'infirmière au pantalon blanc donne un médicament aux enfants.

[lə dɑ̃tist avɛk lɛ̃firmjɛr o pɑ̃talɔ̃ blɑ̃ dɔn œ̃ medikamɑ̃ o zɑ̃fɑ̃]

- Le garçon intelligent qui porte des lunettes et écoute la musique va à l'école.

[lə garsɔ̃ ki pɔrt de lynɛt e ekut la myzik va a lekɔl]

The following words are the target words in the sentences used in the experiment:

- The [ɑ̃] sound: dentiste, pantalon, enfants, mange, dimanche, blanc, médicament
- The [ɛ̃] sound: dindon, infirmière, jardin, martin, poussin, lapin, chien, Alain, vient, raisins
- The [y] sound: Lulu (n=2, at the beginning and end of word), musique, confiture, tortue, laitue
- The [e] sound: vélo, laitue, médicament, école, poupée, dîner

5.3. Test 3: Picture identification task

The third task is a picture identification task (appendix J). This extemporaneous task was used to project learners' implicit spontaneous knowledge. 40 pictures that elicited objects that encompassed the target sounds were given to the participants. The items to be identified were inspired by the third-year French manual book. As in task 1, ten words for each target segment were presented. The following words were presented randomly to the participants to identify:

- The [ɑ̃] sound: pantalon, manteau, enfant, danser, dentiste, dent, maman, orange, viande, médicament
- The [ɛ̃] sound: ceinture, infirmière, princesse, cinquante, imprimante, pain, lapin, train, jardin, matin
- The [y] sound: musée, lunette, sucette, musique, écureuil, jus, tortue, autobus, confiture, sucre
- The [e] sound: vélo, lego, école, médecin, éléphant, poupée, boulanger, café, nez, clé

5.4. Test 4: Storytelling task

The fourth task is the storytelling task (appendix K). This spontaneous task was used to collect participants' spontaneous elicitations of two stories. During this task, participants focused more on the content of their speech rather than on their pronunciation.

The first story that participants were asked to produce was the story of 'the hare and the tortoise' by La Fontaine. The series of eight wordless pictures of the story portrays a hare who was making fun of the slow speed of a tortoise. The tortoise challenged the hare, and they started the race. During the race, the hare stopped to rest under a tree and fell asleep. The tortoise passed him and won the race. When the hare woke up, he found that the tortoise won the race and started crying. This story was chosen as it was familiar to all the participants. Some of the extracts from the ‘the hare and the tortoise’ fable were presented to the raters to evaluate intelligibility, comprehensibility and accentedness.

Another picture that embodies some friends in a forest having a picnic accompanied by some animals was also presented to participants to describe. This picture was used as it contains objects with target sounds (for example: confiture, enfants, jus). The words that comprised the target sounds were used to evaluate the accuracy of segmental pronunciation.

6. Procedure

This section identifies the procedures that the study underwent with regards to data collection, data processing and data analysis.

6.1. Data collection

The tests were carried out over a period of four months. As a first step, a certification of ethical acceptability for research involving human subjects was obtained from the University Human Research Ethics Committee, Concordia University (appendix C). As a following step, headmasters of schools in Sfax (Tunisia) were contacted by phone to ask for their permission to conduct the study in their school. The ten school directors that agreed to participate in our study signed a consent form (appendix D) and an invitation letter was sent to students’ parents (appendix E). The invitation letter comprises general information about the study and the criteria required to participate in the study. Parents who agreed to allow their child to participate in the research signed a consent form (appendix F) and filled the background language questionnaire that involves questions about their child’s extracurricular exposure to French. After reviewing the questionnaires, participants who did not meet the eligibility criteria were eliminated from the study. As a last step, participants were met individually during lunch hours. Meetings were held in quiet rooms under the supervision of a working member of the school in question. To lower the young participants’ stress level, I presented myself, the purpose of the meeting and reassured them that there were no right or wrong answers. After answering participants’ clarifications and questions, they signed a consent form (appendix G). It is to be emphasized that sanitary measures were taken, by keeping the safe distance, wearing a mask during the entire interview, and wiping the iPad with a sanitizer after each meeting. The total duration of the meetings was approximately 15 minutes.

The participants were asked to read a list of words with the target sounds (appendix H), read four sentences (appendix I), identify objects (appendix J), and narrate two stories (appendix K). All the instructions were presented in the native language of the participants (Tunisian Arabic) using simple vocabulary that is adapted to their age. The stimuli appeared on an iPad screen and all the oral productions were audio-recorded using an iPhone. Recording started before starting the first task and continued till the participant finished task four in order not to disturb the fluidity of the interview. In the case of a participant that did not know the word, the answer was presented on the screen solely with no audio help. At the end of the tests, a symbolic gift (a smiley pin) was given to the participants for accepting to be part of the study and for allotting their time during their lunch break.

6.2. Data processing

In order to ensure confidentiality of the participants’ personal data, all the responses to questionnaires and the audio files were kept on a secured computer and codes were attributed to each file. Recordings were classified into three groups (G1, G2, G3). All the responses of each participant were in a single audio file. Using Audacity as an audio editor, the long pauses, hesitations, and requests for clarifications were deleted from the recordings. After that, all recordings were segmented into four sub-files containing the responses for each task.

For task 1 (word list reading), each recording for each participant was divided into forty words using Audacity program. The total number of words was 2720. Then, the words were rearranged. Within each word, the 68 utterances were randomised and separated by a three second pause. Each sequence of words ranged in duration between 4 minutes and 4 minutes and 30s. Finally, we assessed the accuracy of the target segments of the 2720 words. The words were then presented to an expert in phonetics to evaluate the accuracy of the target segments. Any discrepancies were checked again by both parties, and they re-evaluated the target sound. The utterances in this task were also analyzed for word stress using Praat software (version 6.2.23; Boersma & Weenink, 2021). To this end, monosyllable words were eliminated (ange, blanc, jambe, main, faim, cube, blé, fée) because they cannot display stress patterns – all monosyllabic words are intrinsically stressed. Then, by using Praat, the words were segmented into syllables. The duration (in milliseconds, ms.) for each syllable in each word was extracted. A random sample of the evaluations was verified by a professor in phonetics from Concordia University.

For the sentence reading task (task 2), the words that contained the target sounds in this second task were extracted from each sentence (n=45) for each participant to be then evaluated by me and the expert. The words were rearranged and then presented in a random order for the expert. The total number of words that were evaluated by the expert and me was 3060 words. In the case of inconsistencies in evaluations, the word was relistened to and then re-evaluated. Additionally, for each participant, two sentences were presented to the raters for comprehensibility and accentedness ratings (mean length is 20 seconds). The excerpts were intentionally chosen to be relatively short as the study by Munro, Derwing, and Burgess (2010) demonstrated that listeners provide reliable judgments based on short samples. Two other sentences were presented for intelligibility evaluation (mean duration is 25 seconds). The excerpts that were selected for intelligibility evaluations for each participant were presented in a random way for raters in order not to create the sense of repetition. For example, for P1, the evaluators heard sentence 2 and 4; for P2, the evaluators transcribed sentence 3 and 1; etc. All the excerpts that were presented to the evaluators did not contain long pauses, hesitations nor requests for clarifications. The extracts in this task were also analyzed for the accuracy of stress placement. For that reason, the sentences were divided into phrases. The following syllables that appear in bold should be stressed by the participants.

- Dans le jardin de Mar**tin** / se trouvent des din**dons**, / un poussin qui mange de la confi**ture**, / un lapin qui boit du **lait** / et un **chien** / qui joue avec sa pou**pée**. /
- Le di**manche**, / Alain vient en vé**lo** / pour voir la tor**tue**. / Il **dit**: / Lulu la tor**tue**, / au **dîner**! Veux-tu des rai**sins**, / des cara**mels** / ou une lait**ue**? /
- Le den**tiste** / avec l'infirmière au pantalon **blanc** / donnent un médica**ment** / aux en**fants**. /

  Or Le den**tiste** / avec **l'infirmière** / au pantalon **blanc** / donnent un médica**ment** / aux en**fants**. /
- Le garçon qui porte des lu**nettes** / et écoute de la musi**que** / va à l'é**cole**.

All multisyllabic words in bold were extracted and analyzed for each participant for stress placement accuracy (Martin, dindon, confiture, poupée, dimanche, vélo, tortue, dîner, raisins, caramels, laitue, dentiste, infirmière, médicament, enfants, lunettes, musique, école). In addition, the words that contained the target sounds were analysed separately for stress placement accuracy ([ɑ̃] sound: dimanche, dentiste, enfant, médicament; [ɛ̃] sound: infirmière, martin, chien, raisins, dindon; [y] sound: musique, confiture, laitue, tortue; [e] sound: poupée, vélo, dîner, laitue, médicament, école).

For the third task (picture identification task), the same steps that were undertaken in the first task were followed: I divided each of the participant's' recording into forty words, rearranged the 2,720 recordings according to forty utterences. Each utterance contains 68 words in a random order divided by a three second pause. The accuracy of the segmental pronunciation was then evaluated by me and the expert. The extracted words were also measured for word stress accuracy using Praat software. The same steps in test 1 were applied. The following monosyllabic words were eliminated from the analysis: dent, pain, train, jus, nez, clé.

Finally, for the storytelling task, the words that contained the target sounds in the first and second parts of this task were extracted (total of 952 words). The researcher listened carefully to each extract for each participant and extracted the most repeated words. Table 11 presents a summary of the words that were the extracted and then analysed.

Table 11

List of Words Analyzed in the Storytelling Task

[ɑ̃] sound	[ɛ̃] sound	[y] sound	[e] sound
enfant enf**ant** pantalon sandwich mange	lapin chien	éc**u**reuil tortue confiture jus	vélo poupée **é**cureuil

Although we tried our best to encourage the participants to describe all the relevant details in the picture, because this was a spontaneous task, it was difficult for all the participants to say certain target words (e.g., "pantalon"). For this reason, some data is missing in the analysis.

Following the steps in tests 1 and 3, the output was rearranged into 14 words. Within each recording of the target words, the 68 utterances were separated by a three second pause. In the case a participant did not say the word in question, the word was replaced by “please move to the next participant”. For example, when organizing the word “pantalon”, when participant 15 did not say that word, the word was replaced by that sentence. The words were then presented to the expert to evaluate the accuracy of pronunciation of each target sound. The words were also evaluated for word stress accuracy. Approximately one minute from the ‘Hare and the tortoise’ story was analysed. By using Praat, the excerpts were converted into TextGrid format by using the annotate to TextGrid (silences) function that allow to divide each extract into phrases. Only the last words in each phrase were analysed. We reverified the division of the extracts into phrases by listening carefully to each excerpt and extracting the last words of each phrase. Monosyllabic words at the ends of phrases were eliminated. Each 1-minute extract contained approximately 11 words. Afterwards, these words were divided into syllables.

Additionally, 25 seconds were extracted from the ‘tortoise and the hare’ recordings to be presented to the listeners to transcribe them orthographically (for assessing intelligibility). For each participant, the excerpts were selected either from the beginning, middle or end of stories in order not to create the sense of redundancy for the evaluators. All the extracts ended at locations of natural pauses in the utterances. Another 25 seconds from the same story were extracted to be introduced to the raters to evaluate accentedness and comprehensibility. All the recordings were presented to the raters in a random order. 25 seconds from each extract were analysed using Praat for stress accuracy analysis.

At a later stage, the researcher met with the listeners individually (online to limit direct contact during the COVID-19 crisis) and gave them the instructions, defined the terms of “accentedness” and “comprehensibility”, and answered all their requests for clarifications. Afterwards, the raters were instructed to listen carefully to each utterance one time and then write exactly what they had heard. Raters were also asked to evaluate the accentedness and the comprehensibility using a 9-point Likert scale. This methodology to rate intelligibility, comprehensibility and accentedness is consistent with previous speech sample ratings by Derwing and Munro (1997), Munro and Derwing (1995a, 2020) and Munro et al. (2006). Raters were given two practice stimuli for orthographic transcription and rating that were taken from the audio recordings from the pilot study. They were also instructed to try to focus as much as possible on accentedness and comprehensibility and disregard the grammatical and lexical errors. We made sure that the evaluators’ speakers in their laptops operated correctly and that they fully understood the tasks.

6.3. Scoring procedure

6.3.1. Segments

To measure the segmental pronunciation, words from tasks 1, 2, 3 and 4 were evaluated by the researcher and an expert in phonetics using a 0 - 1 scale. This scale was used in the study conducted by Trofimovich and Isaacs (2012) to measure segmental, prosodic, and temporal features of pronunciation. 0 stands for incorrect pronunciation and 1 stands for correct pronunciation. Trofimovich, Kennedy, and Blanchet (2015, p. 102) defined the erroneous segmental pronunciation as the phonemic deletions (e.g., [kaf] instead of [kafe]), substitutions (e.g., [torti] instead of [torty]) and additions (e.g., [tortyr] instead of [torty]). The final score for segmental accuracy was calculated by dividing the total number of words that was correctly pronounced by the total number of words produced in each task (1, 2, 3, and 4) and for each learner separately and multiplied by 100. For example, for the sound /y/ for participant 1 in task1, P1 has had 4 mispronounced words, so the following formula was calculated:

$$\frac{10-4}{10} \times 100 = 60$$

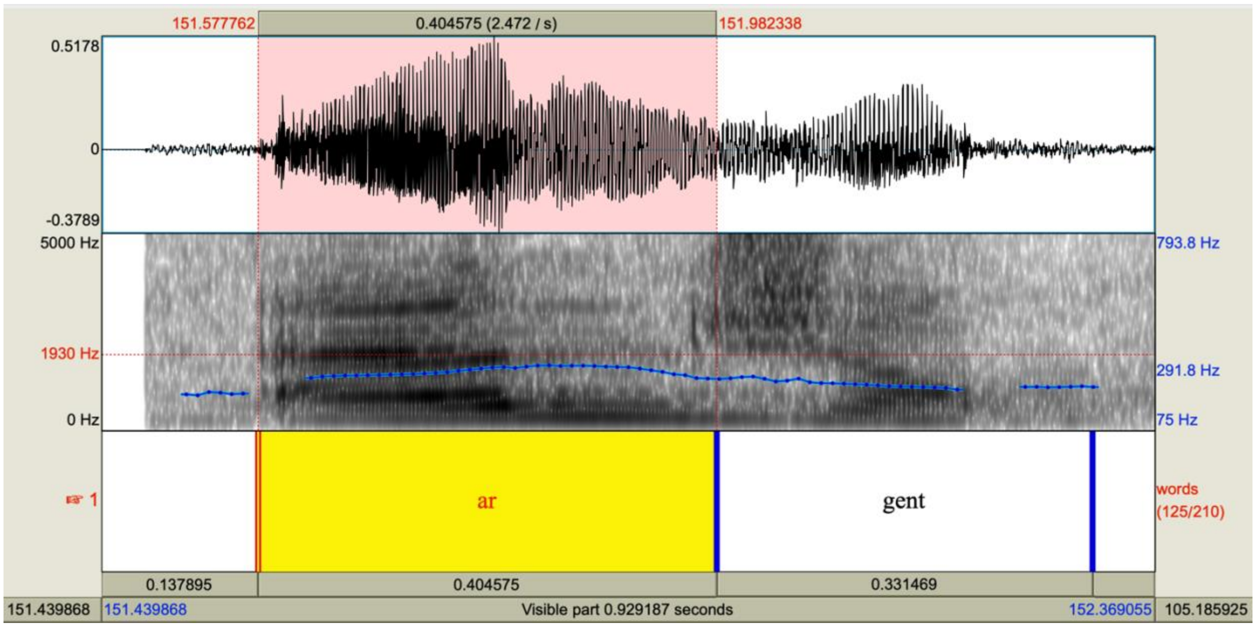
Afterwards, the mean score for every sound for each participant was calculated using SPSS in the four tasks.

6.3.2. Stress

To measure the accuracy of syllabic stress placement, words in tests 1 and 3 were evaluated by the researcher and then a sample was verified by a professor in phonetics. The duration was computed for each syllable in each word. If a longer duration was placed on the last syllable of the words, then "1" was assigned to that word, otherwise, an erroneous syllabic stress placement was assigned a "0". For example, participant 39 put the stress on the first syllable in the word "argent" where the duration of the first syllable is 40ms and the duration of the second syllable is 33ms. Consequently, participant 39 obtained "0" for stress placement in the word "argent". Figure 7 illustrates an example of the erroneous stress placement by participant 39.

Figure 7

Example of an Incorrect Stress Placement Accuracy of the Word "Argent" Using PRAAT



To measure the accuracy of phrasal stress placement in test 2, the final syllable of each rhythmic group was analysed. The duration was computed for each syllable in the final word of the rhythmic group (Martin, dindons, confiture, poupée, dimanche, vélo, tortue, raisins, caramels, laitue, dentiste , médicament, enfants, lunettes, musique, l'école). For every accurate phrasal stress placement, 1 was assigned to the participant, otherwise 0.

Finally, to measure phrasal stress placement in test 4, the duration of syllables of each extracted word was calculated. For every accurate phrasal stress placement, 1 was assigned to the participant, otherwise 0. As a final step, the percentage for syllabic stress accuracy was calculated by dividing the total number of words that have the correct stress placement by the total number of words produced for each learner separately and multiplied by 100. The mean score for every sound for each participant was calculated using SPSS in the four tasks.

6.3.3. Accentedness

To measure accentedness in tests 2 and 4, 25 seconds were extracted from the picture description for each L2 learner. Accentedness was assessed using a 9-point Likert scale used in the study of (Saito et al., 2016) where 1 = no accent, and 9 = heavily accented. To quantify accentedness, an overall score was calculated for each group.

6.3.4. Comprehensibility

To measure comprehensibility in tests 2 and 4, 25 seconds were extracted from the picture description task for each L2 learner. Listeners rated comprehensibility using the 9-point Likert scale that was used in the study of Derwing and Munro (1997) where l means ‘extremely easy to understand’ and 9 means ‘impossible to understand’. To quantify comprehensibility, an overall score was calculated for each group.

6.3.5. Intelligibility

To quantify the orthographical transcriptions, the methodology of Derwing and Munro (1997) was followed. The stimulus was played as many times as necessary by the researcher and then the utterances were transcribed. In a following step, for each transcription, the percentage of exact word matches with the original transcription by the participant was calculated. This procedure entails calculating the number of words correctly transcribed divided by the total number of words and then multiplied by 100. The following formula was used:

$$\text{Score in \%} = \frac{\text{number of correct transcribed words}}{\text{total number of words}} \times 100$$

Derwing and Munro (1997) categorized errors into trivial and nontrivial. Trivial errors are errors that do jeopardize intelligibility such as omissions of function words (e.g., transcribing ‘le tortue, instead of ‘le le tortue’) and regularizations (correcting minor errors such as transcribing ‘la tortue’ instead of ‘la tortie’). These errors were disregarded. According to

Derwing & Munro (1997), only nontrivial mistakes determine the score of intelligibility. Nontrivial errors encompass content word omissions, novel words and substitutions. Novel words are defined as the insertions of words that have no phonological resemblence to the original word in the extract and substitutions are defined as the substituions of ones that have similar phonetic or phonological features (Munro & Derwing, 1995a).

Nevertheless, we did not foresee that the methodology that was used by Derwing and Munro (1997), Munro and Derwing (1995a, 2020) for English learners could not be applied to French language, as it led our raters to transcribe exactly what they had heard, including the phonological mistakes. For that, it was decided to

- analyse solely content words and exclude function words such as articles (e.g., le, la), and fillers (e.g., um, aah)
- take into consideration non-trivial errors.
- in the case of a mispronounced word by the participant and the listeners wrote it as heard, it was calculated incorrect (e.g., P3 said 'lapon' and listener 1 wrote 'lapon).
- in the case of a mispronounced word by the participant and the listeners transcribed it as intended, it was calculated as correct (e.g., P3 said 'lapon' and listener 2 wrote 'lapin). An overall transcription score mean of listeners was calculated for each group. The following extracts are examples from task 2 (sentence reading) and task 4 (storytelling) that demonstrate how intelligibility was measured and calculated. Other examples are presented in appendix A.

- **Example 1 of an original transcript by participant 47 from test 2**

Le garçon intelligent qui qui pate des lu qui pate des lunettes et écoute la musique va à l'icole. Le dimanche Alain le dimanche Alain vient en vélo pour voir la tortue. Il dit : lulu la torti au diner veut tu veut ti des raisins des ca des caramels ou une laitue.

Note: participant 47 said "pate" instead of "porte". For that, this word was calculated as incorrect for listeners who wrote "pate".

Transcriptions of the listeners:

→Listener 1

Le garçon intelligent X qui pate des lunettes et écoute la musique va à l'école. Le dimanche Alain X X X vient en vélo pour voir la tortue. Il dit : lulu la tortue au diner veut tu X X des raisins des caramels ou une laitue.

X : is a missing repeated word that was not taken into consideration

X : is a missing word that was calculated as incorrect

Number of total transcribed words : 47

Number of errors: 5

% of correctly transcribed words = $\frac{47-6}{47} \times 100 = 87.23\%$

→ Listener 2

Le garçon intelligent qui qui pate des linettes et écoute la misique va à l'école. Le dimanche Alain vient X X X au vélo pour voir la tortue. Il dit : lili la tortie au diner veut tu X ti des raisins des des caramelles au des lities.

X : is a missing word that was calculated as incorrect

Number of total transcribed words: 48

Number of errors: 6

% of correctly transcribed words = $\frac{48-6}{48} \times 100 = 87.5\%$

→ Listener 3

Le garçon intelligent qui qui pate des qui pate des lunettes et écoute la musique va à l'école. Le dimanche alain le dimanche alain vient en vélo pour voir la tortue il dit lulu la torti au diner veux tu veu tuté raisins des des ca des caramels ou une laitue.

Number of total transcribed words : 52

Number of errors: 2

% of correctly transcribed words = $\frac{52-1}{52} \times 100 = 96.15\%$

- **Example 1 of an original transcript by participant 44 from test 4**

Dans la jardin le lépain et les animaux, lepain lepain dit à le méti le tortir moi moi et font je cou bo vite que tou. et font un course le pain court vit court vé vite

Note: the listeners reported a great difficulty to transcribe this extract in particular.

Transcriptions of the listeners:

→Listener 1

Dans le jardin le lépain et les animaux, lepain X dit à la X X torti moi X X X je cours plus vite que tu X font X course et le pain court vite X X X.

Number of total transcribed words: 38

Number of errors: 16

% of correctly transcribed words = $\frac{38 -16}{38} \times 100 = 57.89\%$

→Listener 2

Dans le jardin X lépain et les animaux, lepain X dite X X X un la torture X X et font X X X X X X X X X course, et lépain coure vet, court vete, vite.

Number of total transcribed words: 38

Number of errors: 26

% of correctly transcribed words = $\frac{38 - 26}{38} \times 100 = 31.57\%$

→Listener 3

dans le jardin les lépin et les animaux lépin lépin dit X le X X torture X X X X mange X cou bavette que tou et font X course et lepin coure X X vi vite.

Number of total transcribed words: 37

Number of errors: 23

% of correctly transcribed words = $\frac{37 - 23}{37} \times 100 = 37.83\%$

7. Statistical analysis

All statistical tests were conducted using simple or linear correlations using the Statistical Package for Social Sciences (SPSS) to answer the research questions (version 28.0). Numerous parametric and non-parametric tests were carried out to determine the relationships between independent variables. A total of 20536 ratings were collected and used in the analysis of results (5440 ratings for task 1, 6936 ratings for task 2, 5440 ratings for task 3 and 2720 for task 4). We started by transferring the data from Excel file to SPSS. Since the size of groups is small (G1 = 23, G2 = 24, G3 = 21), a Shapiro-Wilk test was used as our numerical means of assessing normality. The reason behind verifying normality is to determine the distribution of the percentage of each variable in order to choose an appropriate statistical method. According to the results of the normality tests, parametric (ANOVA) and non-parametric (Kruskal-Wallis) tests were administered to find any significant differences between the three groups for every variable. The inter-rater reliability was also tested by computing the intraclass correlations (Shrout & Fleiss, 1979). Post-hoc comparisons were carried out using the Mann–Whitney test with the Bonferroni correction. Chapter four presents the results obtained.

Chapter 4

Results

Chapter four reports the results of the statistical analysis of the data collected from the four tasks. Results are arranged into three main sections. The first one presents the main findings of the three groups for each variable arranged according to task type. The second section presents the results according to the dependent variables. The third section deals with task effects. A summary of findings is presented at the end of this section.

1. Effects of AOA and L2 exposure according to task type

1.1. Task 1: word list reading

1.1.1. Segment pronunciation accuracy

A Shapiro-Wilk test was performed for each group for the mean score of pronunciation accuracy of the four target sounds. Results (appendix L) showed that the distribution of data departed significantly from normality for most variables. For the mean score of correct pronunciation of [ɑ̃], results of Shapiro-Wilk test demonstrated that the null hypothesis is rejected, and that the data are not normally distributed for G 1 ($W(23) = 0.86$, p-value = .004) and for G2 ($W(24) = 0.91$, p-value = .032). However, the Shapiro-Wilk test results for G3 did not show evidence of non-normality ($W(21) = 0.95$, p-value = .337). The null hypothesis is not rejected, and the data are normally distributed for G3. Based on these outcomes, a non-parametric test was used.

The results of Shapiro-Wilk test for [ɛ̃] showed that the data are not normally distributed for G 1 and for G3 and are normally distributed for G2. Thus, a non-parametric test was used too. Concerning [y] sound, the distribution of data departed significantly from normality for the three groups. The null hypothesis is rejected, and a non-parametric test was used. The results of the [e] sound was similar to these of [ã] where there is a non-normal distribution for G 1 and for G2 and a normal distribution for G3. Nevertheless, the results of the normality test for the mean score of correct pronunciation of the four target sounds demonstrated a normal distribution of data for the three groups. For that, a parametric test was used to analyse this variable.

1.1.1.1. The [ã] sound

A descriptive statistical analysis was carried out to calculate and analyse the mean score of correct pronunciation of the [ã] sound in task 1. Results (provided in table 12) show that G1 obtained the highest mean score of correct pronunciation of [ã], followed by G2 and then G3.

Table 12

Mean Score of Correct Pronunciation of [ã] in Task 1

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [ã]	23	.50	1.00	.8565	.15023
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [ã]	24	.30	1.00	.6375	.24283
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [ã]	21	.20	1.00	.5857	.22866
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [ã] across the three groups in task 1, H (2, N = 68) = 16.16, $p< .001$, $\eta^2 = .24$ with a mean rank of 47.65 for G1, 29.67 for G2 and 25.62 for G3. To establish the differences among the three groups, Dunn's pairwise comparison tests (table 13) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a statistically significant difference between G1 and G2 (p = 0.005) and between G1 and G3 (p = 0.001). However, there is no statistically significant difference between G2 and G3 (p = 1).

Table 13

Pairwise Comparisons of Groups for Mean Score of Correct Pronunciation of [ã] in Task 1

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group3- Group2	4.048	5.851	.692	.489	1.000
Group3- Group1	22.033	5.910	3.728	<.001	.001
Group2- Group1	17.986	5.714	3.148	.002	.005

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.1.1.2. The [ɛ̃] sound

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct pronunciation of the sound [ɛ̃] in task 1. Results (provided in table 14) show that G1 obtained the highest mean score of correct pronunciation of [ɛ̃], followed by G3 and then by G2.

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [ɛ̃] across the three groups in task 1, H (2) = 21.63, $p < .001$, $\eta^2 = .32$ with a mean rank of 48.13 for G1, 22.06 for G2 and 33.79 for G3. To establish the difference among the three groups, Dunn's pairwise comparison tests (table 15) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 and between G1 and G3. Yet, there is no statistically significant difference between G2 and G3.

Table 14

Mean Score of Correct Pronunciation of [ɛ̃] in Task 1

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [ɛ̃]	23	.70	1.00	.9435	.08435
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [ɛ̃]	24	.10	1.00	.5625	.27317
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [ɛ̃]	21	.20	1.00	.7429	.26753
	Valid N (listwise)	21				

Table 15

Pairwise Comparisons of Groups for Mean Score of Correct Pronunciation of [ɛ̃] in Task 1

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-11.723	5.745	-2.040	.041	.124
Group2-Group1	26.068	5.611	4.646	<.001	.000
Group3-Group1	14.345	5.803	2.472	.013	.040

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.1.1.3. The [y] sound

A descriptive statistical analysis was carried out to calculate and analyse the mean score of correct pronunciation of the sound [y] in task 1. Results (provided in table 16) show that G1 obtained the highest mean score of correct pronunciation of [y], followed by G3 and then by G2.

Table 16

Mean Score of Correct Pronunciation of [y] in Task 1

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [y]	23	.80	1.00	.9087	.08482
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [y]	24	.00	1.00	.4875	.34429
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [y]	21	.10	1.00	.6762	.28090
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [y] across the three groups in task 1, H (2) = 19.55, $p < .001$, $\eta^2 = 29$ with a mean rank of 47.96 for G1, 22.94 for G2 and 32.98 for G3. To establish the difference between the groups, Dunn's pairwise comparison tests (table 17) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 ($p < .001$) and between G1 and G3 ($p = .033$). However, there is no statistically significant difference between G2 and G3 ($p = .254$).

Table 17

Pairwise Comparisons of Groups for Mean Score of Correct Pronunciation of [y] in Task 1

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-10.039	5.823	-1.724	.085	.254
Group2-Group1	25.019	5.686	4.400	<.001	.000
Group3-Group1	14.980	5.882	2.547	.011	.033

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.1.1.4. The [e] sound

A descriptive statistical analysis was carried out to calculate the mean score of correct pronunciation of the sound [e] in task 1. Results (provided in table 18) show that G1 obtained the highest mean score of correct pronunciation of [e], followed by G3 and then by G2.

Table 18

Mean Score of Correct Pronunciation of [e] in Task 1

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [e]	23	.64	1.00	.8972	.10367
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [e]	24	.36	1.00	.6629	.17472
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [e]	21	.36	1.00	.7662	.19193
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [e] across the three groups in task 1, H (2) = 18.95, $p < .001$, $\eta^2 = .28$ with a mean rank of 47.61 for G1, 22.94 for G2 and 33.36 for G3. To establish the difference between the groups, Dunn's pairwise comparison tests (table 19) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a statistically significant difference between G1 and G2 and between G1 and G3 but not between G2 and G3.

Table 19

Pairwise Comparisons of Groups for Mean Score of Correct Pronunciation of [e] in Task 1

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-10.420	5.820	-1.790	.073	.220
Group2-Group1	24.671	5.684	4.341	<.001	.000
Group3-Group1	14.252	5.879	2.424	.015	.046

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.1.1.5. Mean score of sounds

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation of the four target sounds in task 1. Results (provided in table 20) show that G1 obtained the highest mean score of correct pronunciation of the four sounds, followed by G3 and then G2.

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 for the mean score of correct pronunciation of the four sounds in task 1. Results revealed that there was a statistically significant difference between at least two groups (F (2, 65) = [24.534], $p < .001$, η^2=.43). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons (annex L) found that the mean value of the correct pronunciation of the four sounds was significantly different between G1 and G2 ($p < .001$, 95% C.I. = [.17, .36]) and between G1 and G3 ($p < .001$, 95% C.I. = [.09, .28]). Yet, there was no statistically significant difference between G2 and G3 ($p = .096$, 95% C.I. = [-.18, .01]).

Table 20

Mean Score of Correct Pronunciation of the Four Sounds for Each Group in Task 1

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of four sounds	23	.72	1.00	.8986	.07497
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of four sounds	24	.35	.93	.6314	.16697
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of four sounds	21	.47	1.00	.7153	.13929
	Valid N (listwise)	21				

1.1.1.6.Comparison between sounds

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct pronunciation of the four target sounds in task 1. Results (provided in table 21) show that participants obtained the highest mean score of correct pronunciation for the sound [e], followed by [ɛ̃], then by [ɑ̃] and finally [y].

A Friedman test was carried out to compare the mean score of segmental accuracy for the four tasks. There was found to be a significant difference between the methods, $\chi^2(3) = 13.99$, $p < 0.003$. Dunn-Bonferroni post hoc tests were carried out (table 22) and there were significant differences between all the pairs of sounds except for [ɑ̃] and [y], and [ɛ̃] and [e]. To determine the effect sizes, Kendall's W test was administered, W = .07 which indicates a small size effect.

Table 21

Mean Score of Correct Pronunciation of the Four Sounds in Task 1

	N	Minimum	Maximum	Mean	Std. Deviation
Mean of correct pronunciation of the [ɑ̃] sound	68	.20	1.00	.6956	.23905
Mean of correct pronunciation of the [ɛ̃] sound	68	.10	1.00	.7471	.27344
Mean of correct pronunciation of the [y] sound	68	.00	1.00	.6882	.31268
Mean of correct pronunciation of the [e] sound	68	.36	1.00	.7741	.18617
Valid N (listwise)	68				

Table 22

Pairwise Comparisons for the Mean Score of Correct Pronunciation of the Four Sounds in Task 1

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
[ɑ̃] sound- [y] sound	-.029	.221	-.133	.894	1.000
[ɑ̃] sound- [ɛ̃] sound	-.478	.221	-2.159	.031	.185
[ɑ̃] sound- [e] sound	-.640	.221	-2.889	.004	.023
[y] sound- [ɛ̃] sound	-.449	.221	-2.026	.043	.257
[y] sound- [e] sound	-.610	.221	-2.756	.006	.035
[ɛ̃] sound-[e] sound	.162	.221	.731	.465	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.1.2. Stress placement accuracy

A Shapiro-Wilk test was performed for each group for the mean score of correct stress placement for the four target sounds (appendix L). Results showed that the distribution data departed significantly from normality for most variables. For the mean score of correct stress placement of [ɑ̃], results of Shapiro-Wilk test demonstrate the null hypothesis is rejected and that the data are not normally distributed for G1 and for G2. Nevertheless, data are normally distributed for G3. The results of Shapiro-Wilk test for [ɛ̃] demonstrated similar results where the data are not normally distributed for G1 and for G2 and not normally distributed for G3. The results of Shapiro-Wilk test for the mean score of correct stress placement for [y] and [e] revealed a non-normal distribution of the data for the three groups. Based on these results, a non-parametric test was used. Yet, the results of the normality test for the mean score of correct stress placement for the mean score of the four target sounds demonstrated a normal distribution of data for the three groups. For that, a parametric test was used to analyse this variable.

1.1.2.1. The [ɑ̃] sound

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct stress placement for words containing the sound [ɑ̃] in task 1. Results (provided in table 23) show that there is a difference in the mean score where G1 obtained the highest mean score of correct pronunciation of [ɑ̃], followed by G3 and G2 where they obtained almost similar mean scores.

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct stress placement for the words containing the sound [ɑ̃] across the three groups in task 1, H (2) = 12.34, p = .002, η^2= .18 with a mean rank of 45.91 for G1, 29.48 for G2 and 27.74 for G3. To test the significance among the three groups, Dunn's pairwise comparison tests (table 24) were carried out using the adjusted Bonferroni correction. The results indicated a significant difference between G1 and G2 and between G1 and G3 but not between G2 and G3.

Table 23

Mean Score of Correct Stress Placement in Words Containing [ã] in Task 1

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of correct stress placement for [ã]	23	.57	1.00	.7826	.14201
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [ã]	24	.00	.86	.6012	.22678
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [ã]	21	.29	.86	.6054	.18016
	Valid N (listwise)	21				

Table 24

Pairwise Comparisons of Groups for the Mean Score of Stress Placement Accuracy in Words Containing the [ã] Sound in Task 1

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group3-Group2	1.741	5.746	.303	.762	1.000
Group3-Group1	18.175	5.804	3.132	.002	.005
Group2-Group1	16.434	5.611	2.929	.003	.010

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.1.2.2. The [ɛ̃] sound

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct stress placement of words containing the sound [ɛ̃] in task 1. Results (provided in table 25) show that G1 and G2 obtained very close mean values whereas G3 obtained a mean value significantly lower than the other groups.

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct stress placement of words containing the sound [ɛ̃] across the three groups in task 1, H (2) = 12.26, p = .002, η2= .18 with a mean rank of 40.3 for G1, 39.65 for G2 and 22.26 for G3. To test the significance among the three groups, Dunn's pairwise comparison tests (table 26) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a non-significant difference between G1 and G2 and a significant difference between G1 and G3 and G2 and G3.

Table 25

Mean Score of Correct Stress Placement in Words Containing [ɛ̃] in Task 1

Group		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Group1	Mean of correct stress placement for [ɛ̃]	23	.50	1.00	.7609	.15035
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [ɛ̃]	24	.38	1.00	.7552	.18237
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [ɛ̃]	21	.25	1.00	.5833	.18257
	Valid N (listwise)	21				

Table 26

Pairwise Comparisons of Groups for the Mean Score of Stress Placement Accuracy in Words Containing the [ɛ̃] Sound in Task 1

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group3-Group2	17.384	5.761	3.018	.003	.008
Group3-Group1	18.042	5.819	3.101	.002	.006
Group2-Group1	.659	5.626	.117	.907	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.1.2.3.The [y] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct stress placement of words containing the sound [y] in task 1. Results (provided in table 27) show that G1 obtained the highest mean score, followed by G2 and then G3.

Table 27

Mean Score of Correct Stress Placement in Words Containing [y] in Task 1

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of correct stress placement for [y]	23	.78	1.00	.9034	.08412
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [y]	24	.44	1.00	.8750	.12821
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [y]	21	.56	1.00	.7989	.11450
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct stress placement of the sound [y] across the three groups in task 1, H (2) = 10.29, p = .006, η^2= .15 with a mean rank of 40.83 for G1, 37.81 for G2 and 23.79 for G3. To test the significance among the groups, Dunn's pairwise comparison tests (table 28) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a non-statistically significant difference between G1 and G2 (p = 1) and a significant difference between G1 and G3 (p = .008) and between G2 and G3 (p = .036).

Table 28

Pairwise Comparisons of Groups for the Mean Score of Stress Placement Accuracy in Words Containing the [y] Sound in Task 1

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group3-Group2	14.027	5.586	2.511	.012	.036
Group3-Group1	17.040	5.643	3.020	.003	.008
Group2-Group1	3.014	5.455	.552	.581	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.1.2.4. The [e] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct stress placement of words containing the sound [e] in task 1. Results (provided in table 29) show that G1 and G2 obtained close mean scores followed by G3.

Table 29

Mean Score of Correct Stress Placement in Words Containing [e] in Task 1

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of correct stress placement for [e]	23	.50	1.00	.9239	.12351
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [e]	24	.63	1.00	.9062	.12901
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [e]	21	.38	1.00	.8393	.16366
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed that there is non-significant statistical difference in the mean score of correct stress placement of words containing the sound [e] across the three groups, H (2) = 4.56, p = .102, η2= .07 with a mean rank of 38.83 for G1, 36.38 for G2 and 27.62 for G3.

1.1.2.5. Mean score of sounds

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct stress placement of the four sounds in task 1. Results in table 30 show that G1obtained the highest mean score, followed by G2 and then by G3.

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 on the mean score of correct stress placement of the four sounds in task 1. Results revealed that there is a statistically significant difference between at least two groups (F (2, 65) = [9.45], $p < .001$, $\eta^2 = .23$). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons (annexe L) demonstrated that the mean value of the correct pronunciation of the four sounds was significantly different between G1 and G3 ($p < .001$, 95% C.I. = [.06, .21]) and between G2 and G3 ($p = .039$, 95% C.I. = [.003, .15]). Yet, there was no statistically significant difference between G1 and G2 ($p = .14$, 95% C.I. = [-.01, .13]).

Table 30

Mean Score of Correct Stress Placement of the Four Sounds in Task 1

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean	23	.65	.97	.8427	.07499
	Valid N (listwise)	23				
Group2	Mean	24	.43	.96	.7844	.12224
	Valid N (listwise)	24				
Group3	Mean	21	.52	.94	.7068	.10783
	Valid N (listwise)	21				

1.2. Task 2: Sentence reading

1.2.1. Segment pronunciation accuracy

A Shapiro-Wilk test was performed for each group for the mean score of pronunciation accuracy of the four target sounds. Results displayed in appendix L demonstrated that the distribution data departed significantly from normality for most variables. For the mean score of correct stress placement for [ɑ̃], results of the Shapiro-Wilk test demonstrated that the null hypothesis is rejected, and that the data are not normally distributed for the three groups. The same results were found for the [ɛ̃] and [e] sounds. For the results of [y] sound, the distribution of data departed significantly from normality for G1 and for G3, but data were normally distributed for G2 ($W(24) = 0.931$, $p = .105$). Finally, the results of the normality test for the mean score of the four target sounds demonstrated a non-normal distribution of data for G1 and

G2 but a normal distribution for G3. Based in these results, a non-parametric test was used for all the variables.

1.2.1.1. The [ɑ̃] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation of the sound [ɑ̃] in task 2. Results (provided in table 31) showed that G1 obtained the highest mean score of correct pronunciation of [ɑ̃], followed by G3 and then G2.

Table 31

Mean Score of Correct Pronunciation of [ɑ̃] in Task 2

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of correct stress placement for [ɑ̃]	23	.89	1.00	.9903	.03201
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [ɑ̃]	24	.56	1.00	.8657	.13097
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [ɑ̃]	21	.78	1.00	.9153	.09233
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [ɑ̃] across the three groups in task 2, H (2) = 16.67, $p < .001$, $\eta^2 = .25$ with a mean rank of 46.2 for G1, 25.79 for G2 and 31.64 for G3. To test the significance among the groups, Dunn's pairwise comparison tests (table 32) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 (*p* = .000) and between G1 and G3 (*p* = .018). However, there is no significant difference between G2 and G3 (*p* = .790).

Table 32

Pairwise Comparisons of Groups for the Mean Score of Correct Pronunciation of [ã] in Task 2

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2- Group3	-5.851	5.232	-1.118	.263	.790
Group2- Group1	20.404	5.109	3.994	<.001	.000
Group3- Group1	14.553	5.285	2.754	.006	.018

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.2.1.2. The [ɛ̃] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation for words containing the sound [ɛ̃] in task 2. Results (provided in table 33) show that there is a difference in the mean score where G1 obtained the highest mean score of correct pronunciation of [ɛ̃], followed by G3 and then G2.

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [ɛ̃] across the three groups in task 2, H (2) = 11.46, p = .003, η^2= .17 with a mean rank of 44.52 for G1, 25.92 for G2 and 33.33 for G3. To test the significance among the three groups, Dunn’s pairwise comparison tests (table 34) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 (*p* = .002) but there is no significant difference between G1 and G3 and between G2 and G3.

Table 33

Mean Score of Correct Pronunciation of [ɛ̃] in Task 2

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [ɛ̃]	23	.80	1.00	.9478	.07305
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [ɛ̃]	24	.20	1.00	.7542	.22646
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [ɛ̃]	21	.30	1.00	.8333	.20331
	Valid N (listwise)	21				

Table 34

Pairwise Comparisons of Groups for Mean Score of Correct Pronunciation of [ɛ̃] in Task 2

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-7.417	5.656	-1.311	.190	.569
Group2-Group1	18.605	5.524	3.368	<.001	.002
Group3-Group1	11.188	5.713	1.958	.050	.151

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.2.1.3. The [y] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct stress placement for words containing the sound [y] in task 2. Results (provided in table 35) show that G1 obtained the highest mean score of correct pronunciation of [y], followed by G3 and then by G2.

Table 35

Mean Score of Correct Pronunciation of [y] in Task 2

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [y]	23	.91	1.00	.9842	.03523
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [y]	24	.00	1.00	.5871	.29241
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [y]	21	.27	1.00	.7403	.23405
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [y] across the three groups in task 2, H (2) = 35.8, $p < .001$, $\eta^2 = .53$ with a mean rank of 53.28 for G1, 20.5 for G2 and 29.93 for G3. To test the significance among the groups, Dunn's pairwise comparison tests (table 36) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 and between G1 and G3. However, no significant difference was observed between G2 and G3.

Table 36

Pairwise Comparisons of Groups for Mean Score of Correct Pronunciation of [y] in Task 2

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-9.429	5.750	-1.640	.101	.303
Group2-Group1	32.783	5.615	5.838	<.001	.000
Group3-Group1	23.354	5.808	4.021	<.001	.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.2.1.4. The [e] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation for words containing the sound [e] in task 2. Results (provided in table 37) show that G1 and G3 obtained comparable mean score results of correct pronunciation of [e], followed by G2.

Table 37

Mean Score of Correct Pronunciation of [e] in Task 2

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [e]	23	.91	1.00	.9881	.03130
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [e]	24	.45	1.00	.8144	.17677
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [e]	21	.82	1.00	.9437	.06727
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [e] across the three groups in task 2, H (2) = 20.83, p < .001, η^2= .31 with a mean rank of 46.11 for G1, 22.4 for G2 and 35.62 for G3. To test the significance among the groups, Dunn's pairwise comparison tests (table 38) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 and between G2 and G3. However, there is no significant difference between G1 and G3.

Table 38

Pairwise Comparisons of Groups for the Mean Score of Correct Pronunciation of [e] in Task 2

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-13.223	5.335	-2.478	.013	.040
Group2-Group1	23.713	5.210	4.551	<.001	.000
Group3-Group1	10.490	5.389	1.946	.052	.155

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.2.1.5. Mean score of sounds

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation of the four target sounds in task 2. Results (provided in table 39) show that G1 obtained the highest mean score, followed by G3 and then by G2.

Table 39

Mean Score of Correct Pronunciation of the Four Sounds for Each Group in Task 2

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of four sounds	23	.90	1.00	.9776	.02906
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of four sounds	24	.38	.98	.7554	.18028
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of four sounds	21	.56	1.00	.8582	.10931
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the four sounds for each group in task 2, H (2) = 33.45, $p < .001$, $\eta^2 = .5$ with a mean rank of 53.13 for G1, 20.75 for G2 and 29.81 for G3. To test the significance among the three groups, Dunn's pairwise comparison tests (table 40) were carried out using the adjusted Bonferroni correction. The results indicated a significant difference between G1 and G2 and between G3 and G1. However, there is no significant difference between G2 and G3.

Table 40

Pairwise Comparisons of Groups for the Mean Score of Correct Pronunciation of the Four Sounds for Each Group in Task 2

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-9.060	5.886	-1.539	.124	.371
Group2-Group1	32.380	5.748	5.633	<.001	.000
Group3-Group1	23.321	5.946	3.922	<.001	.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.2.1.6. Comparison between sounds

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation of the four target sounds in task 1. Results (provided in table 41) show that participants obtained the highest mean score of correct pronunciation for the sound [ɑ̃], followed by [e], then by [ɛ̃] and finally [y].

Table 41

Mean Score of Correct Pronunciation of Each of the Four Sounds in Task 2

	N	Minimum	Maximum	Mean	Std. Deviation
Mean of correct pronunciation of the [ɑ̃] sound	68	.56	1.00	.9232	.10733
Mean of correct pronunciation of the [ɛ̃] sound	68	.20	1.00	.8441	.19576
Mean of correct pronunciation of the [y] sound	68	.00	1.00	.7687	.27224
Mean of correct pronunciation of the [e] sound	68	.45	1.00	.9131	.13460
Valid N (listwise)	68				

A Friedman test was carried out to compare the mean score of segmental accuracy for the four tasks. There was found to be a significant difference between the methods, $\chi^2(3) = 38.21$, $p < 0.001$. Dunn-Bonferroni post hoc tests were carried out (Table 42) and there were significant differences between [ɑ̃] and [y], between [e] and [y] and between [ɛ̃] and [e]. Yet, there was no statistically significant difference between [ɛ̃] and [y], between [ɑ̃] and [ɛ̃] and between [ɑ̃] and [e]. To determine the effect sizes, Kendall's W test was administered, W = .19 which indicates a small size effect.

Table 42

Pairwise Comparisons for the Mean Score of Correct Pronunciation of the Four Sounds in Task 2

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
[ɛ̃] sound- [y] sound	.382	.221	1.727	.084	.505
[ɑ̃] sound- [y] sound	.941	.221	4.251	<.001	.000
[e] sound- [y] sound	-.971	.221	-4.384	<.001	.000
[ɑ̃] sound- [ɛ̃] sound	.559	.221	2.524	.012	.070
[ɛ̃] sound- [e] sound	-.588	.221	-2.657	.008	.047
[ɑ̃] sound- [e] sound	-.029	.221	-.133	.894	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.

1.2.2. Stress placement accuracy

A Shapiro-Wilk test was performed for each group for the mean score of correct stress placement for words containing the four target sounds and all the words in task 2 (appendix L). Results showed that the distribution of data departed significantly from normality for most of the dependent variables. For the mean score of correct stress placement of [ɑ̃], [ɛ̃] and [e], results of Shapiro-Wilk test demonstrate that the null hypothesis is rejected, and that the data are not normally distributed. For [y], the data are not normally distributed for G1 and G2, but this is not the case for G3. For the mean score of correct stress placement of all words, the data are not normally distributed for G1 and G3 contrary to G2. Based on these outcomes, a non-parametric test was used for the aforementioned variables. Yet, for the mean score of correct stress placement of the four target sounds, the data are normally distributed for G1 and G2 but deviate from normality for G3. For that, a parametric test was used to analyse this variable.

1.2.2.1. The [ɑ̃] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct stress placement for words containing the sound [ɑ̃] in task 2. Results (provided in table 43) show that G3 obtained the highest mean score of correct pronunciation of [ɑ̃], followed by G1 and then by G2.

Table 43

Mean Score of Correct Stress Placement in Words Containing the Sound [ɑ̃] in Task 2

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of correct stress placement for [ɑ̃]	23	.25	1.00	.7717	.22504
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [ɑ̃]	24	.25	1.00	.6458	.25449
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [ɑ̃]	21	.50	1.00	.8095	.19211
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed that there is no statistically significant difference in the mean score of correct stress placement of the sound [ɑ̃] across the three groups in task 2, H (2) = 5.41, p = .067, η^2= .08 with a mean rank of 37.02 for G1, 27.48 for G2 and 39.76 for G3.

1.2.2.2. The [ɛ̃] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct stress placement for words containing the sound [ɛ̃] in task 2. Results (provided in table 44) show that G1 obtained the highest mean score followed by G3 and G2, which had close mean values.

Table 44

Mean Score of Correct Stress Placement in Words Containing the Sound [ɛ̃] in Task 2

Group		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Group1	Mean of correct stress placement for [ɛ̃]	23	.50	1.00	.6957	.16782
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [ɛ̃]	24	.25	1.00	.5104	.22697
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [ɛ̃]	21	.25	1.00	.5476	.18740
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct stress placement in words containing the sound [ɛ̃] across the three groups in task 2, H (2) = 11.04, p = .004, η^2= .16 with a mean rank of 44.72 for G1, 27.65 for G2 and 31.14 for G3. To establish the difference between the groups, Dunn's pairwise comparison tests (table 45) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 and between G1 and G3. But there is no statistically significant difference between G2 and G3 (p = 1).

Table 45

Pairwise Comparisons of Groups for the Mean Score of Stress Placement in Words Containing the [ɛ̃] Sound in Task 2

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-3.497	5.518	-.634	.526	1.000
Group2-Group1	17.072	5.388	3.168	.002	.005
Group3-Group1	13.575	5.574	2.436	.015	.045

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.2.2.3. The [y] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct stress placement for words containing the sound [y] in task 2. Results (provided in table 46) show that G1 obtained the highest mean score. G3 and G2 obtained close mean values.

Table 46

Mean Score of Correct Stress Placement in Words Containing the Sound [y] in Task 2

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of correct stress placement for [y]	23	.40	1.00	.6783	.15654
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [y]	24	.20	.80	.5167	.20359
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [y]	21	.00	1.00	.5238	.26440
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score for correct stress placement of words containing the sound [y] across the three groups in task 2, H (2) = 7.94, p = .019, η^2= .12 with a mean rank of 43.57 for G1, 29.54 for G2 and 30.14 for G3. To test the significance among the groups, Dunn's pairwise comparison tests (table 47) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a statistically significant difference between G1 and G2 and a no significant difference between G1 and G3 and G2 and G3.

Table 47

Pairwise Comparisons of Groups for the Mean Score of Stress Placement in Words Containing the [y] Sound in Task 2

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-.696	5.672	-.123	.902	1.000
Group2-Group1	14.024	5.539	2.532	.011	.034
Group3-Group1	13.327	5.729	2.326	.020	.060

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.2.2.4. The [e] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct stress placement for words containing the sound [e] in task 2. Results (provided in table 48) show that G1 and G3 obtained close mean scores, followed by G2.

Table 48

Mean Score of Correct Stress Placement in Words Containing the Sound [e] in Task 2

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of correct stress placement for [e]	23	.40	1.00	.7304	.16634
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [e]	24	.20	1.00	.5917	.23204
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [e]	21	.00	1.00	.7238	.27185
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically non-significant difference in the mean score of correct stress placement of the sound [e] across the three groups in task 2, H (2) = 5.68, p = .06, $\eta 2$= .08 with a mean rank of 37.13 for G1, 27.25 for G2 and 39.9 for G3.

1.2.2.5. Mean score of the four sounds

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct stress placement of the words containing the four target sounds in task 2. Results (provided in table 49) show that G1 obtained the highest mean score, followed by G3 and then by G2.

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 on the mean score of correct stress placement of words containing the four target sounds in task 2. Results revealed that there was a statistically significant difference between at least two groups (F (2, 65) = [7.11], p = .002, $\eta 2$=.18). The results of the post hoc comparisons using the Tukey's

HSD Test for multiple comparisons (annexe L) indicated that the mean value of correct pronunciation of the four sounds was significantly different between G1 and G2 (p = .001, 95% C.I. = [.06, .27]). However, there is no statistically significant difference between G1 and G3 statistically (p = .29, 95% C.I. = [-.04, .18]), and between G2 and G3 (p = .092, 95% C.I. = [-.20, .01]).

Table 49

Mean Score of Correct Stress Placement of Words Containing the Four Target Sounds in Task 2

Group		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Group1	Mean of correct stress placement of the four sounds	23	.48	.90	.7351	.12256
	Valid N (listwise)	23				
Group2	Mean of correct stress placement of the four sounds	24	.28	.96	.5724	.15413
	Valid N (listwise)	24				
Group3	Mean of correct stress placement of the four sounds	21	.15	.89	.6671	.16741
	Valid N (listwise)	21				

1.2.2.6. Mean score of all words

A descriptive statistical analysis was carried out to calculate the mean score for correct stress placement in all the words in task 2. Results (provided in table 50) show that G1 obtained the highest mean score, followed by G3 and then by G2.

Table 50

Mean Score of Correct Stress Placement in all the Words in Task 2

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of correct stress placement for all words	23	.47	.89	.7391	.11386
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for all words	24	.32	.95	.6031	.14052
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for all words	21	.21	.89	.6742	.14663
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct stress placement in all the words across the three groups in task 2, H (2) = 12.31, p = .002, η^2= .18 with a mean rank of 44.2 for G1, 24.19 for G2 and 35.67 for G3. To test the significance among the groups, Dunn's pairwise comparison tests (table 51) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a statistically significant difference between G1 and G2 and a non-significant difference between G1 and G3 and G2 and G3.

Table 51

Pairwise Comparisons of Groups for the Mean Score of Correct Stress Placement in all the Words in Task 2

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-11.479	5.866	-1.957	.050	.151
Group2-Group1	20.008	5.728	3.493	<.001	.001
Group3-Group1	8.529	5.925	1.440	.150	.450

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.2.3. Comprehensibility

To test inter-rater reliability and to ensure internal consistency in the comprehensibility ratings in task 2, a two-way mixed, consistency, average measure intraclass correlation (ICC) (Shrout & Fleiss, 1979) was calculated for each listener. High reliability among the listeners was observed as the ICC measures were r = .984 with a 95% confidence interval from .976 to .989 (F (67, 134) = 61.173, p = 0). A Shapiro-Wilk test was performed for each group for the mean score of comprehensibility. Results (appendix L) showed that the distribution of data departed significantly from normality for all groups. For that, a non-parametric test was employed. Afterwards, a descriptive statistical analysis was carried out to calculate and analyse the mean score of comprehensibility (table 52). On a scale ranging from 1 (extremely easy to understand) to 9 (impossible to understand), G2 obtained the highest mean score, followed by G3 and then by G1.

Table 52

Mean Score of Comprehensibility in Task 2

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of comprehensibility	23	1.00	6.33	2.3768	1.21980
	Valid N (listwise)	23				
Group2	Mean of comprehensibility	24	1.33	9.00	5.0556	2.57762
	Valid N (listwise)	24				
Group3	Mean of comprehensibility	21	2.00	8.00	4.8571	2.16685
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of comprehensibility across the three groups in task 2, H (2) = 22.42, p <.001, η^2= .33 with a mean rank of 18.67 for G1, 43.04 for G2 and 42.07 for G3. To test the significance among the three groups, Dunn's pairwise comparison tests (table 53) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a statistically significant difference between G1 and G2 and between G1 and G3. However, there is no significant difference between G3 and G2.

Table 53

Pairwise Comparisons of Groups for the Mean Score of Comprehensibility in Task 2

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group1-Group3	-23.398	5.951	-3.932	<.001	.000
Group1-Group2	-24.368	5.753	-4.235	<.001	.000
Group3-Group2	.970	5.892	.165	.869	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.2.4. Accentedness

To test interrater reliability and to ensure there were no significant inter-rater differences in the accentedness ratings in task 2, the average consistency using a two-way mixed intraclass correlation (ICC) (Shrout & Fleiss, 1979) was calculated for each listener. High reliability among the listeners was observed as the ICC measures were r = .975 with a 95% confidence interval from .963 to .984 (F (67, 134) = 40.179, p = 0). A Shapiro-Wilk test was performed for each group for the mean score of accentedness. Results (appendix L) showed that the distribution of data departed significantly from normality for all groups. For that, a non-parametric test was employed. Afterwards, a descriptive statistical analysis was carried out to calculate and analyse the mean score for accentedness (table 54). On a scale ranging from 1 (no accent) to 9 (strong accent), G2 obtained the highest mean score, followed by G3 and then by G1.

Table 54

Mean Score of Accentedness in Task 2

Group		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Group1	Mean score of accentedness	23	1.33	5.33	2.5797	.94931
	Valid N (listwise)	23				
Group2	Mean score of accentedness	24	2.00	8.67	5.1667	2.30521
	Valid N (listwise)	24				
Group3	Mean score of accentedness	21	2.33	7.67	5.0952	1.84735
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of accentedness across the three groups in task 2, H (2) = 24.79, p <.001, η^2= .37 with a mean rank of 17.85 for G1, 43.08 for G2 and 42.93 for G3. To test the significance among the groups, Dunn's pairwise comparison tests (table 55) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a statistically significant difference between G1 and G2 and between G1 and G3. However, there is no significant difference between G3 and G2.

Table 55

Pairwise Comparisons of Groups for the Mean Score of Accentedness in Task 2

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group1-Group3	-25.081	5.952	-4.214	<.001	.000
Group1-Group2	-25.236	5.754	-4.386	<.001	.000
Group3-Group2	.155	5.892	.026	.979	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.2.5. Intelligibility

To test interrater reliability and to ensure there were no significant inter-rater differences in the intelligibility ratings in task 2, the average consistency using a two-way mixed intraclass correlation (ICC) (Shrout & Fleiss, 1979) was calculated for each listener. High reliability among the listeners was observed as the ICC measures were r = .951 with a 95% confidence interval from .926 to .968 (F (67, 134) = 20.304, $p < .001$). A Shapiro-Wilk test was performed for each group for the mean score of accentedness. Results (appendix L) showed that the distribution of data departed significantly from normality for all groups. For that, a non-parametric test was employed. Afterwards, a descriptive statistical analysis was carried out to calculate and analyse the mean score for intelligibility. Results (provided in table 56) show that G1 obtained the highest mean score, followed by G3 and then G2.

Table 56

Mean Score of Intelligibility in Task 2

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of intelligibility	23	92.87	100.00	98.3465	2.24368
	Valid N (listwise)	23				
Group2	Mean of intelligibility	24	64.91	100.00	91.2167	11.21673
	Valid N (listwise)	24				
Group3	Mean of intelligibility	21	78.64	100.00	93.2319	6.31391
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of intelligibility across the three groups in task 2, H (2) = 12.75, p = .002, η^2= .19 with a mean rank of 46.28 for G1, 29.02 for G2 and 27.86 for G3. To test the significance among the three groups, Dunn's pairwise comparison tests (table 57) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 and between G1 and G3. However, there is no significant difference between G2 and G3.

Table 57

Pairwise Comparisons of Groups for the Mean Score of Intelligibility in Task 2

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group3-Group2	1.164	5.821	.200	.842	1.000
Group3-Group1	18.425	5.880	3.134	.002	.005
Group2-Group1	17.262	5.684	3.037	.002	.007

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.3.Task 3: Picture identification task

1.3.1. Segment pronunciation accuracy

A Shapiro-Wilk test was performed for each group for the mean score of pronunciation accuracy of the four target sounds. Results (appendix L) showed that the distribution of data departed significantly from normality for most of the variables. For the mean score of correct pronunciation of [ɑ̃] and [ɛ̃], the results of Shapiro-Wilk test demonstrated that the null hypothesis is rejected, and that the data are not normally distributed. Based on these outcomes, a non-parametric test was used for these two variables. For the results of the normality test of [y], the distribution of data departed significantly from normality for G1 and G2, but data were normally distributed for G3. A non-parametric test was used as well for [y]. Nevertheless, the results of the normality test for [e] and the mean score of the four target sounds demonstrated a normal distribution of data for almost all groups (for the exception of G1 in both variables). For that, a parametric test was used.

1.3.1.1. The [ɑ̃] sound

A descriptive statistical analysis was carried out to calculate the mean score of correct pronunciation of the sound [ɑ̃] in task 3. Results (provided in table 58) show that G1 obtained the highest mean score of correct pronunciation of [ɑ̃] and that G2 and then G3 obtained very similar mean score results.

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [ɑ̃] across the three groups in task 3, $H\,(2, N = 68) = 19.74$, $p < .001$, $\eta^2 = .29$ with a mean rank of 48.61 for G1, 27.08 for G2 and 27.52 for G3. To test the significance among the groups, Dunn's pairwise comparison tests were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 and between G1 and G3. However, there is no statistically significant difference between G2 and G3.

Table 58

Mean Score of Correct Pronunciation of the [ã] Sound in Task 3

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [ã]	23	.93	1.00	.9826	.02993
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [ã]	24	.47	1.00	.8944	.11449
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [ã]	21	.67	1.00	.8984	.09338
	Valid N (listwise)	21				

1.3.1.2. The [ɛ̃] sound

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct pronunciation of the sound [ɛ̃] in task 3. Results (provided in table 59) show that G1 obtained the highest mean score of correct pronunciation of [ɛ̃], followed by G3 and then G2.

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [ɛ̃] across the three groups in task 3, H (2) = 17.17, $p < .001$, $\eta 2= .26$ with a mean rank of 41.93 for G1, 21.56 for G2 and 41.14 for G3. To test the significance among the groups, Dunn's pairwise comparison tests (table 60) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 and between G2 and G3. However, there is no significant statistical difference between G1 and G3.

Table 59

Mean Score of Correct Pronunciation of the [ɛ̃] Sound in Task 3

Age Group		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Group1	Mean score of correct pronunciation of [ɛ̃]	23	.82	1.00	.9328	.06261
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [ɛ̃]	24	.64	1.00	.8220	.08254
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [ɛ̃]	21	.45	1.00	.8961	.16833
	Valid N (listwise)	21				

Table 60

Pairwise Comparisons of Groups for Mean Score of Correct Pronunciation of [ɛ̃] in Task 3

Sample 1-Sample 2	**Test Statistic**	**Std. Error**	**Std. Test Statistic**	**Sig.**	**Adj. Sig.[a]**
Group2-Group3	-19.580	5.686	-3.444	<.001	.002
Group2-Group1	20.372	5.552	3.669	<.001	.001
Group3-Group1	.792	5.743	.138	.890	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.3.1.3. The [y] sound

A descriptive statistical analysis was carried out to calculate the mean score of correct pronunciation of the sound [y] in task 3. Results (provided in table 61) show that G1 obtained the highest mean score of correct pronunciation of [y], followed by G3 and then G2.

Table 61

Mean Score of Correct Pronunciation of the [y] Sound in Task 3

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [y]	23	.73	1.00	.9289	.08649
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [y]	24	.09	1.00	.6174	.29486
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [y]	21	.45	1.00	.7446	.17632
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [y] across the three groups in task 3, H (2) = 20.93, p < .001, $\eta 2$ = .31 with a mean rank of 49.24 for G1, 24.25 for G2 and 30.07 for G3. To establish the difference between the groups, Dunn's pairwise comparison tests (table 62) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 and between G1 and G3. However, there is no statistically significant difference between G2 and G3.

Table 62

Pairwise Comparisons of Groups for Mean Score of Correct Pronunciation of [y] in Task 3

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-5.821	5.817	-1.001	.317	.951
Group2-Group1	24.989	5.680	4.399	<.001	.000
Group3-Group1	19.168	5.875	3.262	.001	.003

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.3.1.4. The [e] sound

A descriptive statistical analysis was carried out to calculate and analyse the mean score of correct pronunciation of the sound [e] in task 3. Results (provided in table 63) show that G1 obtained the highest mean score of correct pronunciation of [e], whereas G3 and G2 obtained very close mean score results.

Table 63

Mean Score of Correct Pronunciation of the [e] Sound in Task 3

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [e]	23	.80	1.00	.9623	.06301
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [e]	24	.67	1.00	.8639	.10674
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [e]	21	.60	1.00	.8603	.10730
	Valid N (listwise)	21				

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 on the mean score of pronunciation accuracy of [e] in task 3. Results revealed that there was a statistically significant difference between at least two groups ($F(2, 65) = [8.559]$, $p < .001$, $\eta^2=.21$). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons indicated that the mean value of the correct pronunciation of [e] was significantly different between G1 and G2 ($p = .002$, 95% C.I. = [.03, .16]) and between G1 and G3 ($p = .002$, 95% C.I. = [.03, .17]). Yet, there was no statistically significant difference between G2 and G3 ($p = .991$, 95% C.I. = [-.06, .07]).

1.3.1.5. Mean score of sounds

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct pronunciation of the four target sounds in task 3. Results (provided in table 64) show that G1 obtained the highest mean score of correct pronunciation of the four sounds, followed by G3 and then by G2.

Table 64

Mean Score of Correct Pronunciation of the Four Sounds in Task 3

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of four sounds	23	.88	1.00	.9516	.03311
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of four sounds	24	.62	.95	.7994	.10275
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of four sounds	21	.64	.96	.8499	.08687
	Valid N (listwise)	21				

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 on the mean score of correct pronunciation of the four sounds in task 3. Results revealed that there is a statistically significant difference between at least two groups ($F(2, 65) = [21.81]$, $p < .001$, $\eta^2=.40$). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons found that the mean value of the correct pronunciation of the four sounds was significantly different between G1 and G2 ($p < .001$, 95% C.I. = [.09, .21]) and between G1 and G3 ($p < .001$, 95% C.I. = [.04, .16]). Yet, there was no statistically significant difference between G2 and G3 ($p = .097$, 95% C.I. = [-.007, .11]).

1.3.1.6. Comparison between sounds

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct pronunciation of the four target sounds in task 3. Results (provided in table 65) show that participants obtained the highest mean score of correct pronunciation for the sound [ɑ̃], followed by [e], then by [ɛ̃] and finally [y].

Table 65

Mean Score of Correct Pronunciation of the Four Sounds in Task 3

	N	Minimum	Maximum	Mean	Std. Deviation
Mean of correct pronunciation of the [ɑ̃] sound	68	.47	1.00	.9255	.09535
Mean of correct pronunciation of the [ɛ̃] sound	68	.45	1.00	.8824	.11968
Mean of correct pronunciation of the [y] sound	68	.09	1.00	.7620	.24233
Mean of correct pronunciation of the [e] sound	68	.60	1.00	.8961	.10455
Valid N (listwise)	68				

A Friedman test was carried out to compare the mean score of segmental accuracy for the four sounds. There was found to be a significant difference between the methods, χ^2 (3) = 40.37, $p < 0.001$. Dunn-Bonferroni post hoc tests were carried out (table 66) and there were significant differences between all the sounds except for [ɑ̃] and [e], and [ɛ̃] and [e]. To determine the effect sizes, Kendall's W test was administered, W = 0.2 which indicates a small size effect.

Table 66

Pairwise Comparisons for the Mean Score of Correct Pronunciation of the Four Sounds in Task 3

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
[ɛ̃] sound- [y] sound	.544	.221	2.458	.014	.084
[e] sound- [y] sound	-.890	.221	-4.018	<.001	.000
[ɑ̃] sound- [y] sound	-1.243	.221	-5.613	<.001	.000
[e] sound- [ɛ̃] sound	-.346	.221	-1.561	.119	.711
[ɑ̃] sound- [ɛ̃] sound	-.699	.221	-3.155	.002	.010
[ɑ̃] sound- [e] sound	.353	.221	1.594	.111	.665

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.3.2. Stress placement accuracy

A Shapiro-Wilk test was performed for each group for the mean score of correct stress placement for the four target sounds in task 3 (appendix L). Results showed that the distribution data departed significantly from normality for most variables. For the mean score of correct stress placement of words containing the sound [ɑ̃], results of Shapiro-Wilk test demonstrated that the null hypothesis is rejected, and that the data are not normally distributed for G1 and for G2 but are normally distributed for G3. For the results of Shapiro-Wilk test of [ɛ̃], the data are not normally distributed for all the groups. For the results of the normality test of the mean score of correct stress placement of words containing the sound [y] the data are not normally distribution for the three groups. The results of Shapiro-Wilk test for [e] demonstrated that data are not normally distributed for G1 and G3, but the null hypothesis is retained for G2. Finally, the results of the normality test for the mean score of correct stress placement for the mean score of the four target sounds demonstrated a normal distribution of data for the three groups.

1.3.2.1. The [ɑ̃] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct stress placement of words containing the sound [ɑ̃] in task 3. Results (provided in table 67) show that G1 obtained the highest mean score of correct stress placement of [ɑ̃], followed by G3 and then by G2.

Table 67

Mean Score of Correct Stress Placement in Words Containing the Sound [ɑ̃] in Task 3

Group		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Group1	Mean of correct stress placement for [ɑ̃]	23	.38	.88	.6793	.14993
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [ɑ̃]	24	.38	.86	.6071	.12077
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [ɑ̃]	21	.38	.88	.6429	.13279
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct stress placement of words containing the sound [ɑ̃] across the three groups in task 3, $H(2) = 12.34$, $p = .002$, $\eta^2 = .18$ with a mean rank of 45.91 for G1, 29.48 for G2 and 27.74 for G3. To establish the difference between the groups, Dunn’s pairwise comparison tests (table 68) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 and between G1 and G3. However, no significant difference was observed between G2 and G3.

Table 68

Pairwise Comparisons of Groups for the Mean Score of Stress Placement in Words Containing the [ɑ̃] Sound in Task 3

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group3-Group2	1.741	5.746	.303	.762	1.000
Group3-Group1	18.175	5.804	3.132	.002	.005
Group2-Group1	16.434	5.611	2.929	.003	.010

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.3.2.2. The [ɛ̃] sound

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct stress placement for words containing the sound [ɛ̃] in task 3. Results (provided in table 69) show that G1 and G3 obtained close mean scores followed by G2.

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct stress placement in words containing the sound [ɛ̃] across the three groups in task 3, $H(2) = 12.26$, $p = .002$, $\eta^2 = .18$ with a mean rank of 40.3 for G1, 39.65 for G2 and 22.26 for G3. To establish the difference between the groups, Dunn's pairwise comparison tests (table 70) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a non-significant difference between G1 and G2 and a significant difference between G1 and G3 and between G2 and G3.

Table 69

Mean Score of Correct Stress Placement in Words Containing the Sound [ɛ̃] in Task 3

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of correct stress placement for [ɛ̃]	23	.50	1.00	.7717	.13406
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [ɛ̃]	24	.50	1.00	.7188	.13417
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [ɛ̃]	21	.50	.88	.7321	.11378
	Valid N (listwise)	21				

Table 70

Pairwise Comparisons of Groups for the Mean Score of Stress Placement in Words Containing the [ɛ̃] Sound in Task 3

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group3-Group2	17.384	5.761	3.018	.003	.008
Group3-Group1	18.042	5.819	3.101	.002	.006
Group2-Group1	.659	5.626	.117	.907	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.3.2.3. The [y] sound

A descriptive statistical analysis was carried out to calculate the mean score of correct stress placement for words containing the sound [y] in task 3. Results (provided in table 71) show that G1 obtained the highest mean score, followed by G2 and then by G3.

Table 71

Mean Score of Correct Stress Placement in Words Containing the Sound [y] in Task 3

Group		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Group1	Mean of correct stress placement for [y]	23	.63	1.00	.8424	.10804
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [y]	24	.50	1.00	.7917	.14590
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [y]	21	.38	1.00	.7202	.14200
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct stress placement for words containing the sound [y] across the three groups in task 3, H (2) = 10.29, p = .006, η^2= .15 with a mean rank of 40.83 for G1, 37.81 for G2 and 23.79 for G3. To test the significance among the groups, Dunn's pairwise comparison tests (table 72) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a no statistically significant difference between G1 and G2 (p = 1) and a significant difference between G1 and G3 (p = .008) and G2 and G3 (p = .036).

Table 72

Pairwise Comparisons of Groups for the Mean Score of Stress Placement in Words Containing the [y] Sound in Task 3

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group3-Group2	14.027	5.586	2.511	.012	.036
Group3-Group1	17.040	5.643	3.020	.003	.008
Group2-Group1	3.014	5.455	.552	.581	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.3.2.4. The [e] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct stress placement for words containing the sound [e] in task 3. Results (provided in table 73) show that G1 and G3 obtained close mean scores followed by G2.

Table 73

Mean Score of Correct Stress Placement in Words Containing the Sound [e] in Task 3

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of correct stress placement for [e]	23	.50	1.00	.7500	.13056
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for [e]	24	.38	1.00	.6875	.19153
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for [e]	21	.38	1.00	.7202	.14200
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed that there is no significant statistical difference in the mean score of correct stress placement of words containing the sound [e] across the three groups in task 3, H (2) = 4.56, p = .102, η^2= .07 with a mean rank of 38.83 for G1, 36.38 for G2 and 27.62 for G3.

1.3.2.5. Mean score of sounds

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct stress placement of words containing the four target sounds in task 3. Results in table 74 show that G1 obtained the highest mean score, followed by G3 and then by G2.

Table 74

Mean Score of Correct Stress Placement Words Containing the Four Target Sounds in Task 3

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of correct stress placement for four sounds	23	.59	.88	.7609	.08354
	Valid N (listwise)	23				
Group2	Mean of correct stress placement for four sounds	24	.50	.87	.7013	.08592
	Valid N (listwise)	24				
Group3	Mean of correct stress placement for four sounds	21	.44	.91	.7039	.09962
	Valid N (listwise)	21				

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 on the mean score of correct stress placement of words containing the four target sounds in task 3. Results revealed that there was a statistically significant difference between at least two groups (F (2, 65) = [9.45], p < .001, η^2=.23). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons found that the mean value of the correct stress placement of the four sounds was significantly different between G1 and G3 ($p < .001$, 95% C.I. = [.06, .21]) and between G2 and G3 ($p = .039$, 95% C.I. = [.003, .15]). Yet, there was no statistically significant difference between G1 and G2 ($p = .14$, 95% C.I. = [-.01, .13]).

1.4. Task 4: Storytelling task

1.4.1. Segment pronunciation accuracy

A Shapiro-Wilk test was performed for each group for the mean score of pronunciation accuracy of the four target sounds in task 4. Results displayed in appendix L demonstrated that the distribution data departed significantly from normality for most variables. For the mean score of correct stress placement for [ɑ̃], [ɛ̃] and [e] results of the Shapiro-Wilk test demonstrated that the null hypothesis is rejected, and that the data are not normally distributed for the three groups. For the results of [y] sound, the distribution of data departed significantly from normality for G1 and G3 but were normally distributed for G2. Finally, the results of the normality test for the mean score of the four target sounds demonstrated a non-normal distribution of data for G1 and G2 but a normal distribution for G3. Based on these results, a non-parametric test was used for all the variables.

1.4.1.1. Results for the [ɑ̃] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation of the sound [ɑ̃] in task 4. Results (provided in table 75) showed that G1 obtained the highest mean score of correct pronunciation of [ɑ̃], followed by G3 and then G2.

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [ɑ̃] across the three groups in task 4, $H(2) = 8.18$, $p = 0.17$, $\eta^2 = .12$ with a mean rank of 42.02 for G1, 29.27 for G2 and 32.24 for G3. To test the significance among the groups, Dunn's pairwise comparison tests (table 76) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2 only. There is no significant difference between G2 and G3 and

Table 75

Mean Score of Correct Pronunciation of [ã] in Task 4 between G1 and G3.

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [ã]	23	.80	1.00	.9826	.05762
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [ã]	24	.00	1.00	.8229	.26580
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [ã]	21	.33	1.00	.8929	.16985
	Valid N (listwise)	21				

Table 76

Pairwise Comparisons of Groups for the Mean Score of Correct Pronunciation of [ã] in Task 4

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2- Group3	-2.967	4.748	-.625	.532	1.000
Group2- Group1	12.751	4.637	2.750	.006	.018
Group3- Group1	9.784	4.796	2.040	.041	.124

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.4.1.2. The [ɛ̃] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation for words containing the sound [ɛ̃] in task 4. Results (provided in table 77) show that there is a difference in the mean score where G3 obtained the highest mean score of correct pronunciation of [ɛ̃], followed by G1 and then G2.

Table 77

Mean Score of Correct Pronunciation of [ɛ̃] in Task 4

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [ɛ̃]	23	.50	1.00	.8913	.21087
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [ɛ̃]	24	.00	1.00	.8750	.26580
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [ɛ̃]	21	.50	1.00	.9048	.20119
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a no statistically significant difference in the mean score of correct pronunciation of the sound [ɛ̃] across the three groups in task 4, $H\,(2) = .06$, $p = .970$, $\eta^2 < .001$.

1.4.1.3. The [y] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation for words containing the sound [y] in task 4. Results (provided in table 78) show that G1 obtained the highest mean score of correct pronunciation of [y], followed by G3 and then by G2.

Table 78

Mean Score of Correct Pronunciation of [y] in Task 4

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [y]	23	.67	1.00	.9348	.12801
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [y]	24	.00	1.00	.6076	.30044
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [y]	21	.25	1.00	.7976	.23801
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [y] across the three groups in task 4, H (2) = 17.78, $p < .001$, $\eta^2 = .27$ with a mean rank of 45.63 for G1, 22.92 for G2 and 35.55 for G3. To test the significance among the groups, Dunn's pairwise comparison tests (table 79) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2. However, no significant difference was observed between G2 and G3 and between G1 and G3.

Table 79

Pairwise Comparisons of Groups for Mean Score of Correct Pronunciation of [y] in Task 4

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-12.631	5.531	-2.284	.022	.067
Group2-Group1	22.714	5.402	4.205	<.001	.000
Group3-Group1	10.083	5.587	1.805	.071	.213

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.4.1.4. The [e] sound

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct pronunciation for words containing the sound [e] in task 4. Results (provided in table 80) show that G1 obtained the highest mean score results of correct pronunciation of [e], followed G2 and then by G3.

Table 80

Mean Score of Correct Pronunciation of [e] in Task 4

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [e]	23	.67	1.00	.9710	.09603
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [e]	24	.33	1.00	.8750	.21563
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [e]	21	.33	1.00	.8492	.21019
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed that there was no statistically significant difference in the mean score of correct pronunciation of the sound [e] across the three groups in task 4, $H(2) = 5.5$, $p= .064$, $\eta^2= .08$ with a mean rank of 40.22 for G1, 32.92 for G2 and 30.05 for G3.

1.4.1.5. Mean score of sounds

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation of the four target sounds in task 4. Results (provided in table 81) show that G1 obtained the highest mean score, followed by G3 and then by G2.

Table 81

Mean Score of Correct Pronunciation of the Four Sounds in Task 4

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of four sounds	23	.71	1.00	.9449	.07754
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of four sounds	24	.35	1.00	.7951	.16698
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of four sounds	21	.63	1.00	.8611	.10601
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the four target sounds across the three groups in task 4, $H\ (2) = 16.74$, $p < .001$, $\eta^2 = .25$ with a mean rank of 47.52 for G1, 24.67 for G2 and 31.48 for G3. To test the significance among the three groups, Dunn's pairwise comparison tests (table 82) were carried out using the adjusted Bonferroni correction. The results indicated a significant difference between G1 and G2 and between G3 and G1. However, there is no significant difference between G2 and G3.

Table 82

Pairwise Comparisons of Groups for the Mean Score of Correct Pronunciation of the Four Sounds in Task 4

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-6.810	5.849	-1.164	.244	.733
Group2-Group1	22.855	5.712	4.001	<.001	.000
Group3-Group1	16.046	5.908	2.716	.007	.020

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.4.1.6. Comparison between sounds

A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct pronunciation of the four target sounds in task 1. Results (provided in table 83) show that participants obtained the highest mean score of correct pronunciation for the sound [e], followed by [ɑ̃], then by [ɛ̃] and finally [y].

A Friedman test (table 84) was carried out to compare the mean score of segmental accuracy for the four tasks. There was found to be a significant difference between the methods, $\chi^2(3) = 18.96, p < 0.001$. Dunn-Bonferroni post hoc tests were carried out and there were no statistically significant differences between [y] and [e] and between [y] and [ɑ̃]. Yet, there was a significant difference between [y] and [ɛ̃]. Also, there was no significant difference between [e] and [ɑ̃] and between [ɑ̃] and [ɛ̃]. To determine the effect sizes, Kendall’s W test was administrated, W = .09 which indicates a small size effect.

Table 83

Mean Score of Correct Pronunciation of the Four Sounds in Task 4

	N	Minimum	Maximum	Mean	Std. Deviation
Mean of correct pronunciation of the [ɑ̃] sound	68	.00	1.00	.8985	.19606
Mean of correct pronunciation of the [ɛ̃] sound	68	.00	1.00	.8897	.22602
Mean of correct pronunciation of the [y] sound	68	.00	1.00	.7770	.26876
Mean of correct pronunciation of the [e] sound	68	.33	1.00	.8995	.18693
Valid N (listwise)	68				

Table 84

Pairwise Comparisons for the Mean Score of Correct Pronunciation of the Four Sounds in Task 4

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
[ɑ̃] sound- [y] sound	.559	.221	2.524	.012	.070
[e] sound- [y] sound	-.581	.221	-2.624	.009	.052
[ɛ̃] sound- [y] sound	.625	.221	2.823	.005	.029
[ɑ̃] sound- [e] sound	-.022	.221	-.100	.921	1.000
[ɑ̃] sound- [ɛ̃] sound	-.066	.221	-.299	.765	1.000
[ɛ̃] sound- [e] sound	.044	.221	.199	.842	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

1.4.2. Stress placement accuracy

A Shapiro-Wilk test was performed for each group for the mean score of correct stress placement in task 4 (appendix L). Results showed that data are normally distributed for G2 and G3 whereas the distribution of data departed significantly from normality for G1. For that, a parametric test was used to analyse this variable. A descriptive statistical analysis was carried out to calculate and analyse the mean score for correct stress placement in task 4. Results (provided in table 85) show that G3 obtained the highest mean score, followed by G1 and then by G2.

Table 85

Mean Score of Correct Stress Placement in Task 4

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of correct stress placement	23	50.00	100.00	85.7748	16.20742
	Valid N (listwise)	23				
Group2	Mean of correct stress placement	24	36.36	100.00	75.7229	15.75418
	Valid N (listwise)	24				
Group3	Mean of correct stress placement	21	45.45	100.00	73.0329	17.68267
	Valid N (listwise)	21				

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 on the mean score of correct stress placement of words containing the four target sounds in task 4. Results revealed that there was a statistically significant difference between at least two groups (F (2, 65) = [3.713], p = .03, η^2=.10). The results of post hoc comparisons using the Tukey's

HSD Test for multiple comparisons (annexe L) demonstrated that the mean value of the correct stress placement was significantly different between G1 and G3 (p = .034, 95% C.I. = [.78, 24.70]). However, there is no statistically significant difference between G1 and G2 (p = .101, 95% C.I.=[-1.51, 21.61]), and between G2 and G3 (p = .849, 95% C.I.= [-9.15, 14.53]).

1.4.3. Comprehensibility

To test inter-rater reliability and to ensure internal consistency in the comprehensibility ratings in task 4, a two-way mixed, consistency, average measure intraclass correlation (ICC) (Shrout & Fleiss, 1979) was calculated for each listener. High reliability among the listeners was observed as the ICC measures were r = .962 with a 95% confidence interval from .943 to .975 (F (67, 134) = 26.178, p = 0). A Shapiro-Wilk test was performed for each group for the mean score of comprehensibility. Results (table 86) showed that the distribution of data is normality distributed for all groups. For that, a parametric test was employed. A descriptive statistical analysis was carried out to calculate and analyse the mean score comprehensibility (table 85). On a scale ranging from 1 (extremely easy to understand) to 9 (impossible to understand), G2 obtained the highest mean score, followed by G3 and then by G1.

Table 86

Mean Score of Comprehensibility in Task 4

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of comprehensibility	23	2.00	8.00	4.5797	1.79003
	Valid N (listwise)	23				
Group2	Mean of comprehensibility	24	3.33	9.00	6.8472	1.30024
	Valid N (listwise)	24				
Group3	Mean of comprehensibility	21	2.33	8.67	5.3333	1.99165
	Valid N (listwise)	21				

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 on the mean score of comprehensibility in task 4. Results revealed that there was a statistically significant difference between at least two groups (F (2, 65) = [10.81], $p < .001$, η^2=.25). The results of post hoc comparisons using the Tukey's HSD Test for multiple comparisons (annexe L) demonstrated that the mean value of the correct stress placement was significantly different between G1 and G2 ($p < .001$, 95% C.I. = [-3.46, -1.08]) and between G2 and G3 ($p = .011$, 95% C.I. = [.29, 2.73). However, G1 and 3 are not significantly different ($p = .314$, 95% C.I. = [-1.99, .48]).

1.4.4. Accentedness

To test interrater reliability and to ensure there were no significant inter-rater differences in the accentedness ratings in task 4, the average consistency using a two-way mixed intraclass correlation (ICC) (Shrout & Fleiss, 1979) was calculated for each listener. High reliability among the listeners was observed as the ICC measures were r = .949 with a 95% confidence interval from .924 to .967 (F(67, 134) = 19.777, $p = 0$). A Shapiro-Wilk test was performed for each group for the mean score of accentedness. Results (appendix L) showed that the data are normally distributed for all groups. For that, a parametric test was employed. A descriptive statistical analysis was carried out to calculate and analyse the mean score for accentedness (table 87). On a scale ranging from 1 (no accent) to 9 (strong accent), G2 obtained the highest mean score, followed by G3 and then by G1.

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 on the mean score of accentedness in task 4. Results revealed that there was a statistically significant difference between at least two groups (F(2, 65) = [59.051], $p < .001$, η^2=.31). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons (annexe L) found that the mean value of the correct stress placement was significantly different between G1 and G2 ($p < .001$, 95% C.I. = [-3.23, -1.26]), between G2 and G3 ($p = .027$, 95% C.I. = [.103, 2.12) and between G1 and G3 ($p = .027$, 95% C.I. = [-2.15, -.11]).

Table 87

Mean Score of Accentedness in Task 4

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of accentedness	23	2.00	5.67	3.5217	1.06755
	Valid N (listwise)	23				
Group2	Mean score of accentedness	24	2.67	8.67	5.7639	1.65205
	Valid N (listwise)	24				
Group3	Mean score of accentedness	21	2.00	7.33	4.6508	1.43556
	Valid N (listwise)	21				

1.4.5. Intelligibility

To test interrater reliability and to ensure there were no significant inter-rater differences in the intelligibility ratings in task 4, the average consistency using a two-way mixed intraclass correlation (ICC) (Shrout & Fleiss, 1979) was calculated for each listener. High reliability among the listeners was observed as the ICC measures were r = .919 with a 95% confidence interval from .879 to .947 ($F(67, 134) = 12.332$, $p < .001$). A Shapiro-Wilk test was performed for each group for the mean score of intelligibility. Results (appendix L) showed that the distribution of data departed significantly from normality for G1 and G2. For that, a non-parametric test was employed. A descriptive statistical analysis was carried out to calculate and analyse the mean score for intelligibility. Results (provided in table 88) show that G1 obtained the highest mean score, followed by G3 and then by G2.

Table 88

Mean Score of Intelligibility in Task 4

Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean of intelligibility	23	79.37	100.00	93.6246	5.02974
	Valid N (listwise)	23				
Group2	Mean of intelligibility	24	42.46	100.00	84.9747	12.16554
	Valid N (listwise)	24				
Group3	Mean of intelligibility	21	84.25	100.00	91.1073	4.42772
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score intelligibility across the three groups in task 4, H (2) = 12.14, p = .002, η^2= .18 with a mean rank of 44.78 for G1, 24.69 for G2 and 34.45 for G3. To test the significance among the three groups, Dunn's pairwise comparison tests (table 89) were carried out using the adjusted Bonferroni correction. The results of these tests indicated a significant difference between G1 and G2. However, there is no significant difference between G2 and G3 and between G1 and G3.

Table 89

Pairwise Comparisons of Groups for the Mean Score of Intelligibility in Task 4

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-9.765	5.906	-1.653	.098	.295
Group2-Group1	20.095	5.768	3.484	<.001	.001
Group3-Group1	10.330	5.966	1.731	.083	.250

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

2. Effects of AOA and L2 exposure according to pronunciation constructs

2.1. Segment Pronunciation Accuracy

A Shapiro-Wilk test was performed for each group for the mean score of segmental accuracy across the four tasks. The Results (appendix L) confirm that the distribution of data departed significantly from normality for the mean score of [ɑ̃] and [ɛ̃]. For that, a non-parametric test was used. However, for the [y], [e] and the mean score of the four sounds, the null hypothesis is retained, and a parametric test was adopted.

2.1.1. The [ɑ̃] sound

A descriptive statistical analysis was carried out to calculate the mean score of correct pronunciation of the [ɑ̃] sound across the four tasks. Results (provided in table 90) show that G1 obtained the highest mean score of correct pronunciation of [ɑ̃], followed by G3 and then by G2.

Table 90

Mean Score of Correct Pronunciation of [ɑ̃] Across the Four Tasks

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [ɑ̃]	23	.85	1.00	.9530	.04429
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [ɑ̃]	24	.51	.97	.8052	.12117
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [ɑ̃]	21	.63	1.00	.8231	.10369
	Valid N (listwise)	21				

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [ɑ̃] across the three groups across the four tasks, H (2) = 27.48, $p < .001$, $\eta^2 = .41$ with a mean rank of 52.04 for G1, 24.92 for G2 and 26.24 for G3. To test the significance among the three groups, Dunn's pairwise comparison tests (table 91) were carried out using the adjusted Bonferroni correction. The results of tests indicated a statistically significant difference between G1 and G2 and between G1 and G3. However, there is no statistically significant difference between G2 and G3.

Table 91

Pairwise Comparisons of Groups for Mean Score of Correct Pronunciation of [ɑ̃] Across the Four Tasks

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group3- Group2	-1.321	5.901	-.224	.823	1.000
Group2- Group1	27.127	5.763	4.707	<.001	.000
Group3- Group1	25.805	5.961	4.329	<.001	.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

2.1.2. The [ɛ̃] sound

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation of the sound [ɛ̃] across the four tasks. Results (provided in table 92) show that G1 obtained the highest mean score of correct pronunciation of [ɛ̃], followed by G3 and then by G2.

A Kruskal-Wallis H test revealed a statistically significant difference in the mean score of correct pronunciation of the sound [ɛ̃] across the three groups across the four tasks, $H(2) = 19.20$, $p < .001$, $\eta^2 = .28$ with a mean rank of 46.67 for G1, 21.54 for G2 and 35.98 for G3. To test the significance among the three groups, Dunn's pairwise comparison tests (table 93) were carried out using the adjusted Bonferroni correction. The results of tests indicated a significant difference between G1 and G2 and between G2 and G3. Yet, there is no statistically significant difference between G1 and G3.

Table 92

Mean Score of Correct Pronunciation of [ɛ̃] Across the Four Tasks

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [ɛ̃]	23	.80	1.00	.9289	.06137
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [ɛ̃]	24	.38	.98	.7534	.12547
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [ɛ̃]	21	.59	1.00	.8443	.14676
	Valid N (listwise)	21				

Table 93

Pairwise Comparisons of Groups for Mean Score of Correct Pronunciation of [ɛ̃] in Task 1

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Group2-Group3	-14.435	5.900	-2.447	.014	.043
Group2-Group1	25.132	5.762	4.362	<.001	.000
Group3-Group1	10.698	5.960	1.795	.073	.218

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

2.1.3. The [y] sound

A descriptive statistical analysis was carried out to calculate the mean score of correct pronunciation of the sound [y] across the four tasks. Results (provided in table 94) show that G1 obtained the highest mean score of correct pronunciation of [y], followed by G3 and then by G2.

Table 94

Mean Score of Correct Pronunciation of [y] Across the Four Tasks

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [y]	23	.84	1.00	.9391	.04510
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [y]	24	.14	.98	.5749	.24082
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [y]	21	.45	.98	.7397	.17475
	Valid N (listwise)	21				

A one-way ANOVA test was performed to compare the effect of AOA and exposure to L2 on the mean score of correct pronunciation of the sound [y] across the four tasks. Results revealed that there was a statistically significant difference between at least two groups (F (2, 65) = [25.494], $p < .001$, η^2=.44). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons found that the mean value of the correct pronunciation of [y] was significantly different between G1 and G2 ($p < .001$, 95% C.I. = [.24, .49]), between G1 and G3 ($p < .001$, 95% C.I. = [.07, .33]), and between G2 and G3 ($p = .007$, 95% C.I. = [-.29, -.04]).

2.1.4. The [e] sound

A descriptive statistical analysis was carried out to calculate the mean score of correct pronunciation of the sound [e] across the four tasks. Results (provided in table 95) show that G1 obtained the highest mean score of correct pronunciation of [e], followed by G3 and then by G2.

Table 95

Mean Score of Correct Pronunciation of [e] Across the Four Tasks

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct pronunciation of [e]	23	.83	1.00	.9547	.04596
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of [e]	24	.62	.98	.8040	.10765
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of [e]	21	.66	.98	.8549	.08587
	Valid N (listwise)	21				

A one-way ANOVA test was performed to compare the effect of AOA and exposure to L2 on for the mean score of correct pronunciation of the four sounds across the four tasks. Results revealed that there was a statistically significant difference between at least two groups ($F(2, 65) = [19.345]$, $p < .001$, $\eta^2 = .37$). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons found that the mean value of the correct pronunciation of the four sounds was significantly different between G1 and G2 ($p < .001$, 95% C.I. = [.09, .21]) and between G1 and G3 ($p < .001$, 95% C.I. = [.04, .16]). Yet, there was no statistically significant difference between G2 and G3 ($p = .115$, 95% C.I. = [-.11, .01]).

2.1.5. Mean score of sounds

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation of the four target sounds across the four tasks. Results (provided in table 96) show that G1 obtained the highest mean score of correct pronunciation of the four sounds, followed by G3 and then G2.

Table 96

Mean Score of Correct Pronunciation of the Four Sounds According to Groups Across the Four Tasks

Age Group		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Group1	Mean score of correct pronunciation of four sounds	23	.85	.99	.9432	.03702
	Valid N (listwise)	23				
Group2	Mean score of correct pronunciation of four sounds	24	.51	.91	.7453	.11740
	Valid N (listwise)	24				
Group3	Mean score of correct pronunciation of four sounds	21	.66	.98	.8211	.08800
	Valid N (listwise)	21				

A one-way ANOVA test was performed to compare the effect of AOA and exposure to L2 on the mean score of correct pronunciation of the four sounds across the four tasks. Results revealed that there was a statistically significant difference between at least two groups (F (2, 65) = [30.183], $p < .001$, η^2=.48). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons found that the mean value of the correct pronunciation of the four sounds was significantly different between G1 and G2 ($p < .001$, 95% C.I. = [.14, .26]), between G1 and G3 ($p < .001$, 95% C.I. = [.06, .19]) and between G2 and G3 ($p = .014$, 95% C.I. = [.13, .14]).

2.1.6. Comparison between sounds

A descriptive statistical analysis was carried out to calculate the mean score for correct pronunciation of the four target sounds across the four tasks. Results (provided in table 97) show that participants obtained the highest mean score of correct pronunciation for the sound [e], followed by [ɑ̃], then by [ɛ̃] and finally [y].

Table 97

Mean Score of Correct Pronunciation of the Four Sounds Across the Four Tasks

	N	Minimum	Maximum	Mean	Std. Deviation
Mean of correct pronunciation of the [ɑ̃] sound	68	.51	1.00	.8607	.11562
Mean of correct pronunciation of the [ɛ̃] sound	68	.38	1.00	.8408	.13591
Mean of correct pronunciation of the [y] sound	68	.14	1.00	.7490	.23018
Mean of correct pronunciation of the [e] sound	68	.62	1.00	.8707	.10471
Valid N (listwise)	68				

A Friedman test (table 98) was carried out to compare the mean score of segmental accuracy for the four tasks. There was found to be a significant difference between the methods, χ^2 (3) = 30.68, $p < 0.001$. Dunn-Bonferroni post hoc tests were carried out and there were significant differences between the sound [y] and all the other sounds. Yet, there was statistically significant difference between [e] and [ɑ̃] and [ɛ̃] and between[ɑ̃] and [ɛ̃]. To determine the effect sizes, Kendall’s W test was administered, W = .15 which indicates a small size effect.

Table 98

Pairwise Comparisons for the Mean Score of Correct Pronunciation of the Four Sounds Across the Four Tasks

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
[ɛ̃] sound- [y] sound	.706	.221	3.188	.001	.009
[ɑ̃] sound- [y] sound	.971	.221	4.384	<.001	.000
[e] sound- [y] sound	-1.118	.221	-5.048	<.001	.000
[ɑ̃] sound- [ɛ̃] sound	.265	.221	1.196	.232	1.000
[e] sound- [ɛ̃] sound	-.412	.221	-1.860	.063	.377
[e] sound- [ɑ̃] sound	-.147	.221	-.664	.507	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

2.2. Stress placement accuracy

A Shapiro-Wilk test was performed for each group for the mean score of stress placement accuracy across the four tasks. The Results (appendix L) confirm that the distribution of data was normally distributed, thus, the null hypothesis is retained, and a parametric test was adopted. Afterwards, a descriptive statistical analysis was carried out to calculate and analyse the mean score of correct stress placement across the four tasks. Results (provided in table 99) show that G1 obtained the highest mean score of stress placement, followed by G2 and then by G3.

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 on the mean score of correct stress placement across the four tasks. Results revealed that there was a statistically significant difference between at least two groups (F (2, 65) = [10.039], $p < .001$, η^2=.24). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons found that the mean value of the correct stress placement across the four tasks was significantly different between G1 and G2 (p = .001, 95% C.I. = [.03, .15]) and between G1 and G3 ($p < .001$, 95% C.I.

= [.04, .15]). Yet, no statistically significant difference was found between G2 and G3 (p = .945, 95% C.I. = [-.05, .065]).

Table 99

Mean Score of Correct Stress Placement Across the Four Tasks

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of correct stress placement across the four tasks	23	.68	.91	.8001	.06602
	Valid N (listwise)	23				
Group2	Mean score of correct stress placement across the four tasks	24	.56	.89	.7115	.08052
	Valid N (listwise)	24				
Group3	Mean score of correct stress placement across the four tasks	21	.54	.88	.7038	.09384
	Valid N (listwise)	21				

2.3.Comprehensibility

A Shapiro-Wilk test was performed for each group for the mean score of comprehensibility in tasks two and four. The Results (table 100) confirm that data was normally distributed, thus, the null hypothesis is retained, and a parametric test was adopted. Afterwards, a descriptive statistical analysis was carried out to calculate and analyse the mean score of comprehensibility in tasks two and four. On a scale ranging from 1 (extremely easy to understand) to 9 (impossible to understand), G2 obtained the highest mean score, followed by G3 and then by G1.

Table 100

Mean Score of Comprehensibility in Tasks Two and Four

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of comprehensibility in tasks two and four	23	1.50	6.00	3.4783	1.09557
	Valid N (listwise)	23				
Group2	Mean score of comprehensibility in tasks two and four	24	2.33	9.00	5.9514	1.63483
	Valid N (listwise)	24				
Group3	Mean score of comprehensibility in tasks two and four	21	2.33	8.17	5.0952	1.65795
	Valid N (listwise)	21				

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 on the mean score of comprehensibility in tasks two and four. Results revealed that there was a statistically significant difference between at least two groups (F (2, 65) = [16.757], $p < .001$, $\eta^2=.34$). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons found that the mean value of comprehensibility in tasks two and four was significantly different between G1 and G2 ($p < .001$, 95% C.I. = [-3.51, -1.44]) and between G1 and G3 ($p = .002$, 95% C.I. = [-2.69, -0.54]). Yet, no statistically significant difference was found between G2 and G3 ($p = .138$, 95% C.I. = [-0.21, 1.92]).

2.4. Accentedness

A Shapiro-Wilk test was performed for each group for the mean score of accentedness in tasks two and four. The Results (appendix L) confirm that data was normally distributed, thus, the null hypothesis is retained, and a parametric test was adopted. Afterwards, a descriptive statistical analysis (table 101) was carried out to calculate and analyse the mean score of accentedness in tasks two and four. On a scale ranging from 1 (no accent) to 9 (strong accent), G2 obtained the highest mean score, followed by G3 and then by G1.

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 on the mean score of correct stress placement in tasks two and four. Results revealed that there was a statistically significant difference between at least two groups (F (2, 65) = [19.852], $p < .001$, η^2=.38). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons found that the mean value of accentedness in tasks two and four was significantly different between G1 and G2 ($p < .001$, 95% C.I. = [-3.37, -1.46]) and between G1 and G3 ($p < .001$, 95% C.I. = [-2.81, -0.84]). Yet, no statistically significant difference was found between G2 and G3 ($p = .318$, 95% C.I. = [-0.38, 1.57]).

Table 101

Mean Score of Accentedness in Tasks Two and Four

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of accentedness in tasks two and four	23	2.00	5.00	3.0507	.78244
	Valid N (listwise)	23				
Group2	Mean score of accentedness in tasks two and four	24	2.67	8.33	5.4653	1.74454
	Valid N (listwise)	24				
Group3	Mean score of accentedness in tasks two and four	21	2.33	7.17	4.8730	1.35801
	Valid N (listwise)	21				

2.5. Intelligibility

A Shapiro-Wilk test was performed for each group for the mean score of intelligibility in tasks two and four. The Results (appendix L) confirm that the distribution of data was normally distributed, thus, the null hypothesis is retained, and a parametric test was adopted. Afterwards, a descriptive statistical analysis was carried out to calculate and analyse the mean score of correct stress placement across tasks two and four. Results (provided in table 102) show that G1 obtained the highest mean score of stress placement, followed by G3 and then by G2.

A one-way ANOVA was performed to compare the effect of AOA and exposure to L2 on the mean score of correct stress placement across tasks two and four. Results revealed that there was a statistically significant difference between at least two groups (F (2, 65) = [8.109], p<.001, η^2=.20). Post hoc comparisons using the Tukey's HSD Test for multiple comparisons found that the mean value of the correct stress placement across the four tasks was significantly different between G1 and G2 ($p < .001$, 95% C.I. = [3.19, 12.59]). Yet, no statistically significant difference found between G1 and G3 ($p = .152$, 95% C.I. = [-1.05, 8.68]) and between G2 and G3 ($p = .113$, 95% C.I. = [-8.89, 0.74]).

Table 102

The Mean Score of Intelligibility in Tasks Two and Four

Age Group		N	Minimum	Maximum	Mean	Std. Deviation
Group1	Mean score of intelligibility in tasks two and four	23	88.44	99.62	95.9856	2.65124
	Valid N (listwise)	23				
Group2	Mean score of intelligibility in tasks two and four	24	55.49	100.00	88.0957	10.25136
	Valid N (listwise)	24				
Group3	Mean score of intelligibility in tasks two and four	21	83.49	99.61	92.1696	4.24866
	Valid N (listwise)	21				

3. Task effect on L2 pronunciation constructs

3.1. Segments

A Shapiro-Wilk test was performed for the mean score of segmental accuracy across the four tasks. Histograms showed that the residuals were skewed. The Results (appendix L) confirm that the distribution of data departed significantly from normality for the mean score of segments in the fours tasks. The null hypothesis is rejected, so the assumptions for carrying out a repeated measures ANOVA are not met, and the Friedman test is more appropriate. Afterwards, a descriptive statistical analysis was carried out to calculate and analyse the mean score of segmental accuracy across the four tasks. Results (provided in table 103) showed that participants obtained the lowest mean score in task 1 and that they obtained very similar results in tasks 2, 3 and 4.

Table 103

Mean Score of Segmental Accuracy Across the Four Tasks

Task		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Task1	Mean score of segmental accuracy in task 1	68	.35	1.00	.7477	.17378
Task2	Mean score of segmental accuracy in task 2	68	.38	1.00	.8623	.15385
Task3	Mean score of segmental accuracy in task 3	68	.62	1.00	.8665	.10210
Task4	Mean score of segmental accuracy in task 4	68	.35	1.00	.8662	.13728
	Valid N (listwise)	68				

A Friedman test (table 104) was carried out to compare the mean score of segmental accuracy for the four tasks. There was found to be a significant difference between the methods, $\chi^2(3) = 58.446$, $p < 0.001$. Dunn-Bonferroni post hoc tests were carried out and there were significant differences between task 1 and tasks 2, 3 and 4. However, there were no significant difference between the other tasks. To determine the effect sizes, Kendall's W test was administered, W = .286 which indicates a small size effect.

Table 104

Pairwise Comparisons for the Mean Score of Segmental Accuracy Across the Four Tasks

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Task 1-Task 3	-1.022	.221	-4.616	<.001	.000
Task 1-Task 2	-1.434	.221	-6.476	<.001	.000
Task 1-Task 4	-1.456	.221	-6.576	<.001	.000
Task 3-Task 2	.412	.221	1.860	.063	.377
Task 3-Task 4	-.434	.221	-1.959	.050	.300
Task 2-Task 4	-.022	.221	-.100	.921	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

3.2. Stress

A Shapiro-Wilk test was performed for the mean score of stress placement accuracy across the four tasks. The histograms showed that the residuals were skewed. The Results (appendix L) confirm that the distribution of data departed significantly from normality for the mean score of segments in the fours tasks. The null hypothesis is rejected, so the assumptions for carrying out a repeated measures ANOVA are not met, and the Friedman test is more appropriate. Afterwards, a descriptive statistical analysis was carried out to calculate and analyze the mean score of segmental accuracy across the four tasks. Results (provided in table 105) showed that participants obtained the highest mean score in task 4 followed by task 1 and then by task 3 and finally by task 2.

A Friedman test was carried out to compare the mean score of segmental accuracy for the four tasks. There was found to be a significant difference between the tasks, $\chi^2(3) = 30.476$, $p < 0.001$. Dunn-Bonferroni post hoc tests were carried out (table 106) and there were significant differences between task one and task two and three and between task four and task two and three. However, there was no statistically significant difference between task two and three and between task one and task four. To determine the effect sizes, Kendall's W test was administrated, W = .149 which indicates a small size effect.

Table 105

Mean Score of Stress Placement Accuracy Across the Four Tasks

Task		N	Minimum	Maximum	Mean	Std. Deviation
Task1	Mean score of segmental accuracy in task 1	68	.43	.97	.7801	.11612
Task2	Mean score of segmental accuracy in task 2	68	.21	.95	.6711	.14389
Task3	Mean score of segmental accuracy in task 3	68	.44	.91	.7222	.09254
Task4	Mean score of segmental accuracy in task 4	68	.36	1.00	.7829	.17177
	Valid N (listwise)	68				

Table 106

Pairwise Comparisons for the Mean Score of Stress Placement Accuracy Across the Four Tasks

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Task 2-Task 3	-.250	.221	-1.129	.259	1.000
Task 2-Task 1	.882	.221	3.985	<.001	.000
Task 2-Task 4	-1.044	.221	-4.716	<.001	.000
Task 3-Task 1	.632	.221	2.856	.004	.026
Task 3-Task 4	-.794	.221	-3.587	<.001	.002
Task 1-Task 4	-.162	.221	-.731	.465	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

3.3. Results for specific pronunciation targets

A Shapiro-Wilk test was performed for the mean score of specific pronunciation targets (segmental pronunciation accuracy and stress placement) across the four tasks. The histograms showed that the residuals were skewed. The Results (appendix L) confirm that the distribution of data departed significantly from normality in tasks one, two, and three. The null hypothesis is rejected, so the assumptions for carrying out a repeated measures ANOVA are not met, and the Friedman test is more appropriate. Afterwards, a descriptive statistical analysis was carried out to calculate and analyse the mean score of the specific pronunciation across the four tasks. Results (provided in table 107) showed that participants obtained the highest mean score in task 4 followed by task 3 and then by task 2 and finally by task 1.

Table 107

Mean Score of Specific Pronunciation Across the Four Tasks

Task		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Task1	Mean score of specific pronunciation in Task 1	68	.45	.97	.7639	.12194
Task2	Mean score of specific pronunciation in task 2	68	.45	.95	.7667	.12396
Task3	Mean score of specific pronunciation in task 3	68	.54	.93	.7944	.08377
Task4	Mean score of specific pronunciation in task 4	68	.57	1.00	.8245	.10976
	Valid N (listwise)	68				

A Friedman test was carried out to compare the mean score of segmental accuracy for the four tasks. There was found to be a significant difference between the tasks, $\chi^2(3) = 22.394$,

p <0.001. Dunn-Bonferroni post hoc tests were carried out and there was a significant difference between task 4 and tasks 1 and 2 (table 108). However, there was no significant difference between tasks one and two, one and three, three and four and between tasks two and three. To determine the effect sizes, Kendall's W test was administered, $W = .33$ which indicates a moderate size effect.

Table 108

Pairwise Comparisons for the Mean Score of Specific Pronunciation Across the Four Tasks

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Task 1-Task 2	-.044	.221	-.199	.842	1.000
Task 1-Task 3	-.382	.221	-1.727	.084	.505
Task 1-Task 4	-.926	.221	-4.185	<.001	.000
Task 2-Task 3	-.338	.221	-1.528	.127	.760
Task 2-Task 4	-.882	.221	-3.985	<.001	.000
Task 3-Task 4	-.544	.221	-2.458	.014	.084

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

3.4. Comprehensibility

A Shapiro-Wilk test was performed for the mean score of comprehensibility across the four tasks. Histograms showed that the residuals were skewed. The Results (appendix L) confirm that the distribution of data departed significantly from normality for the mean score of segments in tasks two and four. The null hypothesis is rejected, so the assumptions for carrying out a repeated measures ANOVA are not met, and the Wilcoxon signed rank test is more appropriate. Afterwards, a descriptive statistical analysis was carried out to calculate and analyse the mean score comprehensibility (table 109). On a scale ranging from 1 (extremely easy to understand) to

9 (impossible to understand), participants obtained the lowest mean score in task 2 and the highest mean score in task 4.

Table 109

Mean Score of Segmental Accuracy in Tasks Two and Four

Task		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Task2	Mean score of segmental accuracy in task 2	68	1.00	9.00	4.0882	2.38682
Task4	Mean score of segmental accuracy in task 4	68	2.00	9.00	5.6127	1.93734
	Valid N (listwise)	68				

A Wilcoxon signed rank test was carried out to compare the mean score of comprehensibility in tasks two and four. Results showed that there was a significant difference ($Z = 4.342$, $p < 0.001$, η^2=.53) between the mean scores in tasks two and task four. The median score for task 2 was 3.0 compared to 6.33 for task 4. Therefore, participants had better results in task 2 than in task 4.

3.5. Accentedness

A Shapiro-Wilk test was performed for the mean score of accentedness across the four tasks. Histograms showed that the residuals were skewed. The Results (appendix L) confirm that the distribution of data departed significantly from normality for the mean score of accentedness in tasks two and four. The null hypothesis is rejected, so the assumptions for carrying out a repeated measures ANOVA are not met, and the Wilcoxon signed rank test is more appropriate. Afterwards, a descriptive statistical analysis was carried out to calculate and analyse the mean score accentedness (table 110). On a scale ranging from 1 (no accent) to 9 (strong accent), there was a slight difference between the two tests and participants obtained the lowest mean score in task 2 and the highest mean score in task 4.

Table 110

Mean Score of Accentedness in Tasks Two and Four

Task		N	Minimum	Maximum	Mean	Std. Deviation
Task2	Mean score of segmental accuracy in task 2	68	1.33	8.67	4.2696	2.14967
Task4	Mean score of segmental accuracy in task 4	68	2.00	8.67	4.6618	1.67559
	Valid N (listwise)	68				

A Wilcoxon signed rank test was carried out to compare the mean score of accentedness in tasks two and four. Results showed that there was no significant difference ($Z = 1.420$, $p =.156$, $\eta^2=.17$) between the mean scores in tasks two and task four. Therefore, participants had similar results in tasks 2 and 4.

3.6. Intelligibility

A Shapiro-Wilk test was performed for the mean score of intelligibility across tasks two and four. The histograms showed that the residuals were skewed. The Results (appendix L) confirm that the distribution of data departed significantly from normality for the mean score of intelligibility in tasks two and four. The null hypothesis is rejected, so the assumptions for carrying out a repeated measures ANOVA are not met, and the Wilcoxon signed rank test is more appropriate. Afterwards, a descriptive statistical analysis was carried out to calculate and analyse the mean score intelligibility (table 111). Participants obtained the highest mean score in task 2 and the lowest mean score in task 4.

Table 111

Mean Score of Segmental Accuracy in Tasks Two and Four

Task		N	Minimum	Maximum	Mean	Std. Deviation
Task2	Mean score of segmental accuracy in task 2	68	64.91	100.00	94.2506	8.13172
Task4	Mean score of segmental accuracy in task 4	68	42.46	100.00	89.7943	8.88046
	Valid N (listwise)	68				

A Wilcoxon signed rank test was carried out to compare the mean score of intelligibility in tasks two and four. Results showed that there was a significant difference ($Z = 4.398$, $p<0.001$, $\eta^2=.53$) between the mean scores in tasks two and task four. The median score for task 2 was 96.48 compared to 91.39 for task 4. Therefore, participants had better results in task 2 than in task 4.

3.7. Results for global pronunciation

A Shapiro-Wilk test was performed for the mean score of the three general pronunciation contracts (comprehensibility, accentedness and intelligibility) in tasks two and four. The histograms showed that the residuals were skewed. The Results (appendix L) confirm that the distribution of data departed significantly from normality for the mean score of the general pronunciation in tasks two and four. The null hypothesis is rejected, so the assumptions for carrying out a repeated measures ANOVA are not met, and the Wilcoxon signed rank test is more appropriate. Afterwards, a descriptive statistical analysis was carried out to calculate and analyse the mean score the global pronunciation in tasks two and four (table 112). Participants obtained the highest mean score in task 2 and the lowest mean score in task 4.

Table 112

Mean Score of Segmental Accuracy in Tasks Two and Four

Task		N	Minimum	Maximum	Mean	Std. Deviation
Task2	Mean score of segmental accuracy in task 2	68	27.41	37.89	34.2028	2.04389
Task4	Mean score of segmental accuracy in task 4	68	20.04	37.89	33.3563	2.67755
	Valid N (listwise)	68				

A Wilcoxon signed rank test (table 113) was carried out to compare the mean score of intelligibility in tasks two and four. Results showed that there was a significant difference ($Z = -2.664$, $p < 0.008$, $\eta^2=.32$) between the mean scores in tasks two and task four. Among the 68 participants, 42 had higher mean score in task 2 than in task 4. Therefore, participants had better results in task 2 than in task 4 for the general pronunciation.

Table 113

Wilcoxon Signed Ranks Test for General pronunciation in Tasks Two and Four

		N	Mean Rank	Sum of Ranks
MeanT2 - MeanT4	Negative Ranks	26[a]	28.35	737.00
	Positive Ranks	42[b]	38.31	1609.00
	Ties	0[c]		
	Total	68		

a. MeanT2 < MeanT4
b. MeanT2 > MeanT4
c. MeanT2 = MeanT4

4. Summary of results

4.1. AOA and L2 exposure effects on L2 pronunciation constructs

To summarize, the main goal of this dissertation is to examine the AOA (age of acquisition) effect and exposure to L2 on the acquisition of the general and specific pronunciation and to examine the task effect as well. To this end, three groups with different starting ages and different amounts of exposure to L2 (G1: AOA 4 years, 1600h exposure L2 / G2: AOA 8 years, 1000h exposure L2 / G3: AOA 4 years, 1000h exposure L2) have undertaken four different tasks (two controlled and two spontaneous). To examine the effects of L2 exposure, the results of G1 and G3 were compared. To study the effects of AOA, the results of G2 and G3 were compared. Table 114 summarizes the main findings in each task where the ranking of participants was presented according to the mean score and the significant difference between the groups was presented according to the post-hoc statistical analyses.

To summarize the findings, G1 obtained the best mean score on almost all the pronunciation constructs whereas there was no statistically significant difference between G2 and G3 in almost all the pronunciation features targeted by the research. These results indicate that with the same starting ages (4 years old), G1 with more L2 exposure (1600h) produced more target-like forms than G3 (100h) and that there was no significant difference between G2 and G3 who have had the same amount of L2 exposure (1000h) and different AOAs (4 years vs 8 years).

Table 114

Summary of Results According to Task Type

Task	Ranking according to mean score	Significant difference
Task 1: word list reading		
- Segments		
• [ɑ̃]	G1 > G2 = G3	No difference between G2 and G3
• [ɛ̃]	G1 > G3 = G2	No difference between G2 and G3
• [y]	G1 > G3 = G2	No difference between G2 and G3
• [e]	G1 > G3 = G2	No difference between G2 and G3
• Mean	G1 > G3 = G2 G1 (M = .90) > G3 (M = .72) = G2 (M =.63) (p < .001)	No difference between G2 and G3
- Stress		
• [ɑ̃]	G1 > G3 = G2	No difference between G2 and G3
• [ɛ̃]	G1= G2 > G3	No difference between G1 and G2
• [y]	G1= G2 > G3	No difference between G1 and G2
• [e]	G1= G2 = G3	No difference between G1, G2 and G3
• Mean	G1= G2 > G3 G1 (M = .84) = G2 (M =.78) > G3 (M = .71) (p < .001)	No difference between G1 and G2
Task 2: sentence reading		
- Segments		
• [ɑ̃]	G1 > G3 = G2	No difference between G2 and G3
• [ɛ̃]	G1 = G3 = G2	No difference between G1/G3 and G2 / G3
• [y]	G1 > G3 = G2	
• [e]	G1 = G3 > G2	No difference between G2 and G3
• Mean	G1 > G3 = G2 G1 (M = .98) > G3 (M = .86) = G2 (M =.76) (p < .001)	No difference between G1 and G3 No difference between G2 and G3

- Stress		
• [ã]	G1 = G3 = G2	No difference between G1, G2 and G3
• [ɛ̃]	G1 > G3 = G2	No difference between G2 and G3
• [y]	G1 = G3 = G2	No difference between G1/G3 and G2/G3
• [e]	G1 = G3 = G2	No difference between G1, G2 and G3
• Mean Sounds	G1 = G3 = G2	No difference between G1/G3 and G2/G3
• Mean Words	G1 = G3 = G2	No difference between G1/G3 and G2/G3
	G1 (*M* = .74) > G3 (*M* = .67) = G2 (*M*= .60) (*p* = .002)	
- Comprehensibility	G1 > G3 = G2 G1 (*M* = 5.06) > G3 (*M* = 4.86)> G2 (*M*=2.38) (*p* <.001)	No difference between G2 and G3
- Accentedness	G2 = G3 > G1 G2 (*M*=.5.17)= G3 (*M* = 5.10)> G1 (*M* = 2.58) (*p* <.001)	No difference between G2 and G3
- Intelligibility	G1 > G3 = G2 G1 (*M* = 98.35) > G3 (*M* = 93.23) = G2 (*M*=91.22) (*p* = .002)	No difference between G2 and G3
Task 3: picture identification		
- Segments		
• [ã]	G1 > G3 = G2	No difference between G2 and G3
• [ɛ̃]	G1 = G3 > G2	No difference between G1 and G3
• [y]	G1 > G3 = G2	No difference between G2 and G3
• [e]	G1 > G3 = G2	No difference between G2 and G3
• Mean	G1 > G3 = G2	No difference between G2 and G3
	G1 (*M* = .95) > G3 (*M* = .85) = G2 (*M*=.80) (*p* < .001)	
- Stress		
• [ã]	G1 > G3 = G2	No difference between G2 and G3
• [ɛ̃]	G1 = G2 > G3	No difference between G1 and G2

• [y]	G1 = G2 > G3	No difference between G1 and G2
• [e]	G1 = G2 = G3	No difference between G1, G2 and G3
• Mean	G1 = G2 > G3 G1 (*M* = .76) = G2 (*M*=.70) > G3 (*M* = .70) ($p < .001$)	No difference between G1 and G2
Task 4: storytelling		
- Segments		
• [ɑ̃]	G1 = G3 = G2	No difference between G1/G3 and G2/G3
• [ɛ̃]	G1 = G3 = G2	No difference between G1, G2 and G3
• [y]	G1 = G3 = G2	No difference between G1/G3 and G2/G3
• [e]	G1 = G3 = G2	No difference between G1, G2 and G3
• Mean	G1 > G3 = G2 G1 (*M* = .94) > G3 (*M* = .86) = G2 (*M*=.80) ($p = .007$)	No difference between G2 and G3
- Stress	G1 = G2 = G3 G1 (*M* = 85.78) = G2 (*M* = 75.72) = G3 (*M* = 73.03) ($p = .03$)	No difference between G1/G2 and G2/G3
- Comprehensibility	G2 > G3 = G1 G2 (6.85) > G3 (5.33) = G1 (*M* = 4.58) ($p < .001$)	No difference between G1 and G3
- Accentedness	G2 > G3 > G1 G2 (5.76)> G3 (4.65)> G1 (*M* = 3.52) ($p < .001$)	All groups are significantly different
- Intelligibility	G1 = G3 = G2 G1 (*M* = 93.63) = G3 (91.11) = G2 (84.98) ($p = .02$)	No difference between G2/G3 and G1/G3

Table 115 summarizes the main findings for each L2 pronunciation target across the four tasks. The ranking of participants is presented according to the mean score and the significant difference between the groups is illustrated according to the post-hoc statistical analyses.

Table 115

Summary of Results according to L2 pronunciation constructs *Across the Four Tasks*

Dependent Variable	Ranking according to mean score	Significant difference
- Segments • [ɑ̃] • [ɛ̃] • [y] • [e] • Mean	G1 > G3 = G2 G1 = G3 > G2 G1 > G3 > G2 G1 > G3 = G2 G1 > G3 > G2 G1 (*M* = .94) > G3 (*M* = .82) > G2 (*M* = .75) ($p < .001$)	No difference between G2 and G3 No difference between G1 and G3 All groups are significantly different No difference between G2 and G3 All groups are significantly different (p =.014)
Stress	G1 > G2 = G3 G1(*M* = .80) > G2 (*M* =.71) = G3 (*M* = .70) ($p < .001$)	No difference between G2 and G3 (p = .95)
Comprehensibility	G2 = G3 > G1 5.9514 = 5.0952 > 3.4783	No difference between G2 and G3
Accentedness	G2 = G3 > G1 5.4653 = 4.8730 > 3.0507	No difference between G2 and G3
Intelligibility	G1 = G3 = G2 95.9856=92.1696=88.0957	No difference between G2/G3 and G1/G3

Table 116 summarizes the main findings of the comparison between the target sound in order to determine which L2 segments were the most difficult for the participants. The ranking of participants is presented according to the mean score and the significant difference between the groups is illustrated according to the post-hoc statistical analyses.

Table 116

Summary of Results for the Comparison of Segments

Dependent Variable	Ranking according to mean score	Significant difference
Task 1	[e] > [ɛ̃] > [ɑ̃] > [y] .7741>.7471>.6956>.6882	No difference between [ɑ̃] and [y], and [ɛ̃] and [e].
Task 2	[ɑ̃] > [e] > [ɛ̃] > [y] .9232 >.9131>.8441>.7687	No difference between [ɛ̃] and [y], and [ɑ̃] and [e].
Task 3	[ɑ̃] > [e] > [ɛ̃] > [y] .9255> .8961>.8824 >.7620	No difference between [ɑ̃] and [e], and [ɛ̃] and [e].
Task 4	[e] > [ɑ̃] > [ɛ̃] > [y] .8995>.8985>.8897 >.7770	There is a difference between [y] and all the other sounds
Across four tasks	[e] > [ɑ̃] > [ɛ̃] > [y] .8707>.8607 >.8408 >.7490	There is a difference between [y] and all the other sounds

The findings for the comparison between the target sounds indicate that the participants obtained the highest mean values for [e] followed by [ɑ̃], then by [ɛ̃] and finally by [y] across the four tasks. This concludes that the Tunisian learners of French L2 face the most difficulties with the sound [ɛ̃] and mostly with the sound [y].

4.2. Task effects on L2 pronunciation constructs

The second goal of this dissertation is to examine task effects on the different dimensions of L2 pronunciation. To achieve this goal, participants were given four different tasks: two controlled tasks and two spontaneous ones. The mean results for each task were compared for each dependent variable. Table 117 presents the main results.

Table 117

Summary of Results for Task Effects

DV	Results (lowest to highest)	Significant difference
Segments	T1 < T2 < T4 <T3	No significant difference between T2/T3, T2/T4, T3/T4
Stress	T2 < T3 < T1 <T4	No significant difference between T2/T3, T1/T4
Specific pronunciation	T1 < T2 < T3 <T4	No difference between T1, T2/ T1, T3/ T2, T3 There is a difference between : T1, T4/ T2, T4/ T3, T4
Comprehensibility	T2 <T4	Better results in T2 than in T4
Accentedness	T2 <T4	No significant difference
Intelligibility	T4 <T2	Better results in T2 than in T4
General pronunciation	T4 <T2	Better results in T2 than in T4

To summarize, participants obtained the lowest results for segmental pronunciation accuracy in T1 (the most controlled task) and that there was no statistically significant difference between T2, T3 and T4. For stress placement accuracy, participants obtained the highest mean score in T4 (the most spontaneous task). For comprehensibility and intelligibility, participants had better results in T2 (controlled) than in T4 (spontaneous). Finally, for accentedness, there was no statistically significant difference between the two tasks. To reiterate, participants had better results in the controlled tasks more than in the spontaneous ones.

Chapter 5

Discussion

This dissertation sought to examine whether AOA or L2 exposure has a greater effect on the general and specific L2 pronunciation proficiency for Tunisian learners of French L2 in the school setting. An additional goal was to explore the effects of tasks (controlled vs. spontaneous) on the participants' L2 pronunciation outcomes. In this chapter, a summary of the findings reported in the previous chapter is presented and discussed in order to offer insights to the three research questions:

1. Will an earlier starting age result in a better global and specific French L2 pronunciation proficiency across different task types in the school setting?
2. Will a longer exposure to French L2 result in a better global and specific L2 pronunciation proficiency across different task types in the school setting?
3. Will learners have better results in the controlled or in the spontaneous tasks?

This chapter summarizes and discusses the major findings. It is organized into three main parts: part one will discuss the findings about AOA and L2 exposure effects in each task and across the four tasks. Part 2 will explain the results about the task effects. Finally, part 3 will present a summary of the results.

1. Research objective 1: Analysing the effects of AOA and L2 exposure on pronunciation

The first goal of this dissertation is to investigate the effects of AOA and L2 exposure on the acquisition of French pronunciation by comparing the specific and general pronunciation constructs of three Tunisian groups ($N = 68$) that had different starting ages and different amounts of exposure to L2 (G1: AOA 4 years, 1600h exposure L2 / G2: AOA 8 years, 1000h exposure L2 / G3: AOA 4 years, 1000h exposure L2).

Previous studies that investigated the age effects and pronunciation in the school setting have shown inconclusive results that range from demonstrating an advantage for late starters to no effects of age. Fullana (2006) found that older learners had better results in segmental production accuracy. Likewise, Fullana (2006) and Fullana and MacKay (2003) demonstrated that older participants obtained better ratings in accentedness. Garcia Lecumberri and Gallardo (2003) also found that older learners were more intelligible and had less perceived foreign accent. However, Fullana and Mora (2007) and Mora and Fullana (2007) found that neither age nor exposure to TL had a significant effect on participants' segmental production. This result was further confirmed by the study of MacKay and Fullana (2007) who examined intelligibility and accentedness. These findings of the previous studies are partially in line with our findings. Our results will be discussed in detail for each pronunciation construct in the following sections.

1.1. Segmental Pronunciation Accuracy

Participants were tested for the segmental pronunciation accuracy in the word reading task, sentence reading task, picture identification task and storytelling task. Recall that the four target segments of this study (/ɑ̃, ɛ̃, y, e/) were chosen based on a pilot study, in which we found that they were the most problematic phonemes for Tunisian learners of French. Words that contained the target sounds were extracted in four tasks and then evaluated by the researcher and a phonetician by assigning "1" to correct pronunciation and "0" to incorrect pronunciation.

Results of segmental pronunciation accuracy of the mean score for the four vowels have shown that in tasks 1, 2, 3 and 4, G2 (older learners) was not significantly different from G3 (younger learners). For the L2 exposure effect, there was a significant statistical difference between G1 and G3 in all tasks. This is further confirmed when calculating the mean score of correct segmental pronunciation accuracy across the four tasks and all the groups are significantly different: G1 obtained the best mean score results, followed by G3 and then by G2. This indicates that younger learners (G3) had better pronunciation accuracy results than the older learners (G3) and that the groups of learners with more L2 exposure (G1) had better pronunciation accuracy results than those with less amount of L2 exposure (G3).

Our findings are inconsistent with the results of Fullana and Mora (2007) and Mora and Fullana (2007). Fullana and Mora (2007) administered acoustic analyses for the final-word obstruents pronunciation accuracy of 48 Catalan-Spanish learners of English L3 by using a delayed sentence repetition task. They found that neither AOA nor exposure to TL had significant effects for segmental pronunciation. These results corroborate those found in Mora and Fullana (2007) who examined the perception and production of two vowel contrasts among 48 Catalan-Spanish learners of English L3 by using a delayed sentence repetition task.

A possible explanation for our findings is the fact that the participants in our study had larger amounts of exposure to L2 than in the aforementioned studies, where the participants were tested after at most 600 hours of classroom exposure. Another possible explanation to our findings is that the results reported in Fullana and Mora (2007) and Mora and Fullana (2007) emerged from one single task (i.e., delayed sentence repetition task) whereas our results derive from four different task types.

In addition, when looking closely at the results for each vowel pronunciation accuracy, only task 1 yielded consistent findings where there was no statistically significant difference between G2 and G3. Otherwise, each sound in the other tasks has a different result. This finding could be explained by the effects of the task type, where learners demonstrated greater performance when the task required more attention on form. Consequently, G1 with more L2 exposure used their more developed knowledge to pronounce the target sounds in a more accurate way.

Additionally, despite the fact that our study did not originally seek to investigate the segmental difficulties for Tunisian learners of French, the limited number of empirical research that investigated segmental difficulties for Tunisian learners of French piqued our interest to further investigate this issue. There is also compelling evidence that errors in higher functional load segmental features are associated with greater loss of comprehensibility and intelligibility (Munro & Derwing, 2006; Suzukida & Saito, 2019). Investigating the precise set of problematic segments may guide instructors to explicitly teach these segments in class. For that, a qualitative analysis of the segmental mistakes was administered in order to determine the segmental difficulties for the Tunisian learners of French.

As described in the methodology section, the segments that were chosen to be studied in this dissertation were based on a pilot study as they were expected to be the most challenging sounds for Tunisian learners. However, it was found that the participants had mainly difficulties with the [ɛ̃] and [y] sounds. Additionally, a qualitative analysis was conducted where each word has been carefully listened to in tasks 1 and 3. Each pronunciation mistake was annotated and whenever there was hesitation, an expert in phonetics was consulted. The most common mistakes in tasks 1 and 3 are summarized in appendix M. As a summary, the most common segmental mistakes for the sound [ɑ̃] in task 1 is that it was pronounced as [am] and [ɑn] and as [ɑn] in task 3. For [ɛ̃], it was mistakenly pronounced as [in] and [en] in task 1 and as [e] in task 3. For [y], this phoneme was pronounced by a large number of participants as [i] in task 1 and as [i] and [u] in task 3. Finally, the sound [e] was pronounced as [i] and [ej] in tasks 1 and 3. These mistakes can be explained by the fact that the participants did not receive any focused and explicit pronunciation instruction in class. Besides, as the Contrastive Analysis Hypothesis suggests, the closer a phone in the TL to the L1 category, the easier it is to perceive and to produce. Added to that, according to Saito (2019), the school setting could facilitate only the development of easy aspects of L2 speech.

1.2. Stress Placement Accuracy

For stress placement accuracy, participants were evaluated acoustically via PRAAT software in the word reading task, sentence reading task, picture identification task and storytelling task. The duration of the last syllable of each word in tasks one and three was measured. In tasks two and four, only the final syllables of words in each phrase were evaluated. Concerning the stress placement mistakes, a zero was assigned when the duration of the last syllable of the word/ phrase is shorter than the other syllables. Word stress placement for Tunisian dialect is rule-governed (Almbark et al., 2014) whereas in French, the final syllable is always stressed.

Results of stress placement accuracy demonstrated that in tasks 1 and 3, G2 (older learners) obtained better results than G3 (younger learners). In tasks 2 and 4, there was no significant difference between the two groups. When calculating the mean score of correct stress placement across the four tasks, there was no statistically significant difference between G2 and G3. This result suggests that there are no significant effects of AOA on stress placement accuracy of L2. A possible justification for the fact that older learners had better results in the tasks that involved the pronunciation of isolated words (tasks 1 and 3) is that although we made sure that the selected words in these tasks are known by all the participants, some younger learners in group 3 faced some difficulties in recognizing certain words, probably due to anxiety. Yet, we have noticed that older learners were more confident when achieving these two tasks.

Indeed, according to Zheng and Cheng (2018), cognitive test anxiety is a significant negative predictor of L2 achievement which could explain the obtained results.

For the L2 exposure effect, there was a significant statistical difference between G1 and G3 in all tasks (for the exception of task 2) where G1 obtained the highest mean score in all tasks. When calculating the mean score of correct stress placement across the four tasks, there was a significant difference between G1 and G3. Put differently, L2 exposure had substantial effects on stress placement accuracy of L2 where the group of learners with more exposure to L2 had better results than those with less amounts of exposure to French. This could be explained by the fact that the group with more exposure placed stress more accurately than the other group. This means that an extended exposure is needed to better master this pronunciation feature.

When it comes to previous studies, to our knowledge, there are no studies that examined the effects of AOA and L2 exposure on stress placement accuracy in the school setting. Although there is a considerable number of studies that have investigated the effects of age and stress placement accuracy in naturalistic settings, the results of these findings cannot be extrapolated to instructed learning contexts. The main reason is due to the differences between the two environments, as explained in chapter 2 (the quantity and the quality of input).

1.3. Comprehensibility

For comprehensibility, learners were assessed by three non-native raters in the sentence reading task and storytelling task. A nine-point Likert scale was used where 1 stands for “extremely easy to understand” and 9 stands for “impossible to understand”. Results of comprehensibility ratings in task 2 (sentence reading) have shown that there was no statistically significant difference between G2 and G3. However, in task 4 (storytelling), G3 performed better than G2. When calculating the mean score of comprehensibility ratings across tasks two and four, there was no statistically significant difference between G2 and G3. These results indicate that there were no significant effects of AOA on comprehensibility of L2.

A probable explanation for these results is the potential effect of the biological age at the time of testing on participant's outcomes. In fact, the age of the participants in G1 and G2 is 12 years at the time of testing, whereas the age at testing in G3 is 8 years. This could suggest that older learners wanted to demonstrate their sophisticated cognitive development abilities and used more advanced and complex vocabulary to narrate the stories in task 4. As a result, this caused them to produce numerous grammatical and syntactic mistakes. Contrarily, the younger learners used short and simple sentences. As such, the grammatical and syntactic mistakes could have influenced raters' evaluation of comprehensibility. For that, the younger learner obtained better comprehensibility rating than the older ones in task 4.

When it comes to the mean score of the two tasks, the older and younger learners obtained similar comprehensibility ratings. This is probably due to the fact that, in task 4, the younger learners obtained slightly better results than the older learners, which did not exercise a noticeable impact on the overall results in the two tasks.

For L2 exposure effects, G1 was more comprehensible than G3 in task 2. However, there was no significant statistical difference between G1 and G3 in task 4. When calculating the mean score of comprehensibility ratings across tasks two and four, G1 obtained better mean results than G3. This indicates that the amount of L2 exposure has an effect on L2 comprehensibility. The findings in task 4, in which the groups with 1600h of exposure had similar results with the groups of participants with 1000h, could be explained by the fact that the participants in G3 used simple and short sentences that did not contain a lot of grammatical mistakes, whereas the participants in G1 produced stories in a more complex and structured manner with little grammatical mistakes. Consequently, the raters found that these two groups were comprehensible.

When it comes to previous studies, and just like stress placement accuracy, there are no studies that examined the effects of AOA and L2 exposure and comprehensibility in the school setting.

1.4. Accentedness

Concerning accentedness, learners were assessed by three non-native raters in the sentence reading and storytelling tasks. A nine-point Likert scale was used where 1 stands for “no accent” and 9 stands for “strong accent”. Results of accentedness ratings in task 2 have shown that there was no statistically significant difference between G2 and G3. However, in task 4, there was a significant difference between G2 and G3, where G3 had better accent ratings than G2. Although both groups (G2 and G3) had approximately the same amount of L2 exposure (1000h), the results in task 4 could be explained by the fact that the younger participants were enrolled in private schools, whereas the older learners were enrolled in public schools where the quality of French input and teaching in general could differ.

The fact that there were different results in tasks 2 and 4 can be justified by the task complexity effect, where task 2 (sentence reading) requires more attention to the form and students paid careful attention to their pronunciation and read the sentences slowly. While task 4 (storytelling) necessitates attention on meaning rather on form and participants narrated the stories without paying attention to their pronunciation.

When calculating the mean score of accentedness ratings across tasks 2 and 4, there was no statistically significant difference between G2 and G3. These results indicate that there were no significant effects of AOA on the accentedness of L2. This means that the results in task 4 did not have a great impact on the overall results and that the two groups have approximately the same level of accentedness.

For the amount of L2 exposure, there was a significant difference between G1 and G3 where G1 obtained better accentedness ratings in task 2 and task 4 and in the mean score of accentedness across the four tasks. These results indicate that there is a substantial effect of the amount of L2 exposure on accentedness. This could be justified by the fact that learners with 1600h of L2 exposure benefited greatly from their accumulated exposure to the TL and, in addition, produced more target-like levels of accentedness when compared to the group of participants that had fewer amounts of L2 exposure (1000h).

Concerning the previous empirical findings, our results are not consistent with the findings of Fullana (2006), Fullana and MacKay (2003), and Garcia Lecumberri and Gallardo (2003) who found an advantage for late starters. To recap, Fullana (2006) examined the accentedness of 323 Spanish learners of English L3 that had three different AOAs: 4, 8 and 11 years. They were tested in three-time points: after 200h, 416h and 726h. To elicit data, Fullana (2006) used an imitation task. The target segments were evaluated in terms of accentedness using a Likert scale. Concerning the study of Fullana and MacKay (2003), they examined the accentedness of 135 Spanish/ Catalan speakers of English with different starting ages and different degrees of exposure to TL. Participants completed an imitation task and raters evaluated the pronunciation of two vowels in terms of accentedness using a Likert scale. For the study of Garcia Lecumberri and Gallardo (2003), the researchers studied the accentedness of 60 Spanish learners of English L3 that had three different AOAs: 4, 8 and 11 years. To collect data, researchers used a storytelling task. Accentedness was evaluated by using a 9-point Likert scale.

In these previous studies, researchers found that learners who started L3 acquisition at a younger age obtained lower accent ratings compared to those who started L3 acquisition at the age of eleven. These results differ from our findings because the researchers evaluated only a specific set of segments in terms of accentedness. However, our methodology was adopted from the studies of Derwing and Munro (1997) and Munro and Derwing (1999). The accentedness was assessed by raters who listened to 30 second extracts from two different tasks (sentence reading and storytelling). Another probable explanation for the differences in the findings is that all the participants in the studies of Fullana (2006), Fullana and MacKay (2003), and Garcia Lecumberri and Gallardo (2003) are Spanish /Catalan learners of English L3. Consequently, the difference in the L1s, TLs and the geographical location of the studies could possibly clarify the discrepancy in the outcomes.

On the other hand, our findings are consistent with the results of MacKay and Fullana (2007) and Mora and Fullana (2007) who concluded that early and late learners had similar levels of accentedness. To recap, MacKay and Fullana (2007) examined the accentedness of two vowels of 135 Spanish /Catalan learners of English L3 through an imitation task. Mora and Fullana (2007) examined the production of two sets of vowel contrasts of 49 Spanish /Catalan learners of English L3 through a delayed sentence repetition task. MacKay and Fullana (2007) and Mora and Fullana (2007) explained their results by the fact that their participants are still in the early stages of FL learning.

1.5. Intelligibility

Finally, for intelligibility, the three non-native raters transcribed orthographically extracts from the sentence reading task and the storytelling task. Then, the percentage of correct transcribed words was calculated for each participant across the two tasks. Results of intelligibility in task 2 and task 4 indicated that there was no statistically significant difference between G2 and G3. When calculating the mean score of comprehensibility ratings across tasks two and four, there was also no statistically significant difference between G2 and G3. These results indicate that there were no significant effects of AOA on intelligibility of L2 and that regardless of the comprehensibility and accentedness ratings, older and younger participants succeeded in conveying their intended meaning.

For L2 exposure effects, G1 was more intelligible than G3 in task 2. However, there was no significant statistical difference between G1 and G3 in task 4. When calculating the mean score of comprehensibility ratings across tasks two and four, there was no significant difference between G1 and G3. This indicates that the amount of L2 exposure has no effect on L2 intelligibility. The discrepancies in the results could be explained by the task complexity effect where task 2 requires more attention on form and task four requires less attention.

According to Munro and Derwing (2019), intelligibility is the most important dimension of pronunciation as it underpins the success of the spoken communication. Despite the importance of intelligibility, Garcia Lecumberri and Gallardo's (2003) research was the only study that investigated age effects and intelligibility, but against protocols in the field, they used a 9-point Likert scale, not transcription or comprehension questions. The authors found that late learners who started L3 learning at an older starting age had better intelligibility ratings than early starters. These findings do not align with our findings as there were no effects of AOA neither of L2 exposure on intelligibility.

A possible explanation to our results that differ from Garcia Lecumberri and Gallardo's (2003) findings is the method used to evaluate intelligibility. In our research, the methodology used (orthographic transcriptions) was adopted from the studies of Derwing and Munro (1997), Munro and Derwing (1999) and Munro et al. (2006). This methodology differs from the one used in the study of Garcia Lecumberri and Gallardo (2003), who used a Likert scale. The use of a Likert scale was avoided in our research to eliminate the confusion among the raters between the terms of comprehensibility (ease of understanding) and intelligibility (the actual understanding of the intended message).

Also, as mentioned in the methodology section, the raters transcribed exactly what they had heard, including the phonological mistakes. As a result, the methodology adopted from the studies of Derwing and Munro (1997), Munro and Derwing (1999) and Munro et al. (2006) was modified. While analyzing the results, it was noticed that there was a slight difference between rater 1 and raters 2 and 3. Rater 1 used a top-down approach where he used his background information to predict the meaning of the utterance. To illustrate an example, when a participant said the word "torture", rater 1 wrote the "tortue" as he understood the meaning. Contrarily, raters 2 and 3 used a bottom-up approach where they focused on the sounds instead of the meaning of the intended message. It should be noted that the inter-rater correlation demonstrated a good harmony in the results between the three raters in task 2 ($F\,(67, 134) = 20.304$, $p < .001$) and in task 4 ($F(67, 134) = 19.777$, $p < .001$).

1.6. General discussion

Our results suggest that AOA does not have an impact on L2 pronunciation, but rather the amount of L2 exposure. These results do not align with the previous findings of Muñoz (2006b) and Singleton and Pfenninger (2017), who found that older learners learn faster and more efficiently than their younger counterparts in the limited-input setting that characterizes classroom-based instruction. Added to that, our findings do not provide evidence on "the younger, the better" belief but rather "the more, the better". Also, our results indicate that age effect is not the sole predictor on the learners' L2 outcome. Consequently, our findings support the results of Dai Weidong (2008) and Muñoz (2010, 2019) who demonstrated that starting age for foreign language learning is unimportant and that the quantity of input is the most important factor and even a much relevant factor than age.

In fact, Flege (2009), Moyer (2014), Muñoz (2010, 2019) and Pfenninger and Singleton (2017a) are adamant about the fact that the age factor should be viewed as a 'macrovariable' that is associated with a myriad of factors and could not be described as the main sole determinant factor that could determine learner's ultimate achievement in L2. Age is associated with both internal and external variables and cannot be studied in isolation (such as motivation and aptitude). Certainly, the extent to which language learners may benefit from language instruction differs substantially between individuals, owing to the fact that each learner has his unique processing qualities ranging from motivation, aptitude, affect and awareness (Saito, 2019). Flege (2009) asserts that the state of learners' cognitive development as well as the state of development of L1 proficiency, the phonetic inventory of L1 and the quality and quantity of L2 input are influencing factors that affect learners' L2 outcomes.

White and Turner (2005) confirmed the importance of the quality and quantity of L2 input by conducting a study that compared the oral production of two ESL classes in different regions in Montreal. The first group (n = 73) of students is enrolled in an intensive school whereas the second group (n = 79) is enrolled in non-intensive school settings. Results demonstrated that students who were enrolled in the intensive programme and had 400 hours of instruction had better results than those who were enrolled in the regular school and had 108 hours of instruction of English. In fact, young learners demonstrate a slower rate in acquiring an additional language than older learners (Larson-Hall, 2008; Nikolov & Mihaljević Djigunović, 2006). Indeed, according to Jaekel et al., (2017, p. 635), young learners require more exposure to TL in order to absorb the underlying rule/pattern without being aware of them. In the long run, some longitudinal studies reported that depending on the amount of TL exposure, there is a substantial progress in proficiency for early start programs at the end of the elementary level (Hopp et al., 2019; Unsworth et al., 2015).

However, although an early start could lead to a more exposure to TL, which leads to a better pronunciation as our results demonstrated, we underline that our findings should not be overgeneralized. The first reason concerns the limited number of the participants. Secondly, surveys reported that early learners can be overwhelmed by the early introduction of the foreign languages which contributes to a cognitive overload, thus, a loss of motivation, negative attitudes and failure to meet the minimum learning objectives in the TL (Lambelet & Berthele, 2015). Thirdly, the socio-economic difference between the participants can have an impact on the results, where the students in G1 and G2 belong to private schools, while G3 are learners who come from public schools. Lambelet and Berthele (2015) indicated in their study that the socioeconomic background, the psychological disposition of the learner and the aspects of the family life are factors that interact with AOA. For these reasons, the decision of lowering the starting age for learning foreign languages in schools should be taken with a great precaution.

To conclude, the findings in this dissertation allow us to answer the first two research questions:

- **Research question 1:** Will an earlier starting age result in a better global and specific French L2 pronunciation proficiency across different task types in the school setting?

  Answer: No, an earlier stating age did not result in a better global and specific pronunciation of French L2 pronunciation proficiency across the four tasks in the school setting where there was no statistically significant difference between G2 (AOA 8 years and 1000h exposure to L2) and G3 (AOA 4 years, 1000h exposure to L2).

- **Research question 2:** Will a longer exposure to French L2 result in a better global and specific L2 pronunciation proficiency across different task types in the school setting?

  Answer: Yes. A longer exposure to French L2 result in a better global and specific L2 pronunciation proficiency across the four tasks in the school setting where there was a statistically significant difference between G1 (AOA 4 years, 1600h exposure to L2) and G3 (AOA 4 years, 1000h exposure to L2).

2. Research objective 2: Analyzing the effects of task type on L2 pronunciation with regard to AOA and L2 exposure

The second goal of this dissertation is to investigate the effects of task type on French L2 pronunciation performance by comparing the pronunciation outcomes of the three Tunisian groups that had different AOA and different amounts of exposure to L2 across four task conditions: from a controlled task (word-list reading) to less controlled task (sentence reading), to less spontaneous task (picture identification) to spontaneous task (storytelling). Results of task effects will be discussed according to each pronunciation construct.

2.1. Segments

As a brief summary, participants were tested for the segmental pronunciation accuracy of four vowels (/ɑ̃, ɛ̃, y, e/) in the word reading task, sentence reading task, picture identification task and storytelling task. Results indicated that only task 1 was significantly different from the other three tasks whereas there is a significant difference between tasks 2, 3 and 4. Participants in task 1 obtained the lowest mean results. This indicates that participants obtained the lowest mean result in the most controlled task, while they achieved comparable results in the three other tasks. The results suggested that the cognitive demand of these tasks was not sufficiently different from one another.

Our findings are inconsistent with Robinson's cognition hypothesis (2007), which states that when tasks are cognitively challenging, learners will foster more accurate and complex L2 dimensions. Our results also contradict the earlier empirical findings that stipulate that when learners' performance is evoked through formal/ controlled tasks, the pronunciation is closer to target-like forms than when they are elicited in the spontaneous tasks (Plonsky & Saito, 2019). To cite an example, Rau, Chang, and Tarone (2009) reported that in picture description task, the participants mispronounced the sound /θ/ more frequently than in word and sentence reading tasks.

For task effects regarding AOA, older and younger learners obtained comparable results in each of the four tasks. This indicates that task type did not have an impact on L2 segmental pronunciation accuracy regarding starting ages. Likewise, for the task effects regarding L2 exposure effect, learners who had more exposure had better results than those with less exposure amounts across the four tasks. This suggests that task type did not have an effect on L2 segmental pronunciation accuracy regarding the amounts of L2 exposure.

When delving into details, and as previously mentioned, results were only consistent in task 1 by showing that there is no significant statistical difference between the older and younger learners and that learners with more exposure performed better than learners with less exposure to L2 for pronunciation accuracy of the four target sounds. When it comes to the other tasks, there was no significant difference between older and younger learners, no significant difference between learners with more and less amount of L2 exposure, and that all the groups were significantly different. A possible explanation for our results is the fact that task 1 is the first task that the participants started with. Despite our efforts to calm the participants, we have noticed that all of them with no exception were very anxious at the beginning of the task. We believe that their anxiety affected their performance in task 1.

2.2. Stress

Recall that participants were tested for the stress placement accuracy of four vowels (/ɑ̃, ɛ̃, y, e/) in the following four tasks: word reading, sentence reading, picture identification, and storytelling. Results indicated that there was no significant difference between tasks 2 and 3 and no significant difference was found between tasks 1 and 4 as well. Participants in tasks 1 and 4 obtained better mean results for stress placement accuracy than in tasks 2 and 3. In other terms, participants obtained the highest results in the most controlled and the most spontaneous tasks and the lowest results in the less controlled and less spontaneous tasks. This entails that our results are not consistent, and that task type does not have a clear impact on stress placement accuracy.

Regarding AOA, results of stress placement accuracy demonstrated that older learners obtained better results than the younger ones in tasks 1 and 3 whereas in tasks 2 and 4, and that there was no significant difference between the two groups. Again, this indicates that task type did not exercise a clear effect on stress placement accuracy of the TL. For the L2 exposure effect, learners with more exposure to L2 obtained better results than those with less exposure amounts in all the tasks and in the mean score across the four tasks except for task 2.

Just as in the segmental pronunciation accuracy, our findings are inconsistent with Robinson’s cognition hypothesis (2007) and with the findings of the empirical research. A possible justification for these results is the fact that our tasks were not accurately elaborated enough to get clear results for the stress placement construct. Indeed, the young age of the participants (8 years) and the restricted time of the meetings with the participants (15 min) required us to elaborate simple and short tasks which may led to the obtained results.

2.3. Comprehensibility

As a reminder, participants' L2 production was tested for comprehensibility in the sentence reading task, and storytelling task. These two tasks varied in the cognitive demand where sentence reading task is considered as simpler than the spontaneous storytelling task. Results indicated that participants obtained the lowest mean score in task 2 and the highest mean score in task 4. On the nine-point Likert scale was used where 1 stands for "extremely easy to understand" and 9 stands for "impossible to understand", these results indicate that participants obtained better results in task 2 (controlled) than in task 4 (spontaneous). This indicates that task type had an impact on comprehensibility. These results go in line with Skehan and Foster's (2001) Limited Attentional Capacity Model that states that complex tasks cause learners to disperse their attention over different performance demands which causes a trade-off between the form and the meaning. Therefore, learners have poorer language performance in more complex tasks.

Regarding AOA, results of comprehensibility ratings demonstrated that older and younger learners had comparable results in task 2. However, in task 4, young learners were more comprehensible than the older ones. These findings indicate that task type has indeed an effect on L2 comprehensibility regarding AOA. In fact, younger learners had better ratings than the older learners in the spontaneous tasks whereas both groups had similar results in task 2. This could be explained by the fact that task 2 is a much less complex task that does not necessitate a lot of cognitive effort. However, as previously mentioned, the biological age at the time of testing of the participants imposed an effect on learners' outcomes where the younger learners used simple and short sentences and the older learners utilized more complex and lengthy sentences. This cognitive effort of the older participants has led the unconscious and unmonitored L2 pronunciation forms to surface and consequently, to a less comprehensible production. As stated by Saito and Plonsky (2019), learners in spontaneous tasks focus on the grammatical, phonological, lexical and pragmatic aspects of the language rather than paying attention to their articulatory knowledge.

For the L2 exposure effect, learners with more exposure to L2 obtained better rating results than those with less exposure amounts in task 2. Yet, there was no significant difference between the two groups of participants in task 4. This confirms that task type has indeed an effect on L2 comprehensibility regarding L2 exposure amounts. In fact, task 2 is a controlled task that entices learners to elicit the explicit and conscious knowledge of L2 pronunciation forms. For that, learners with more hours of L2 exposure benefited from their extended exposure to TL and used their conscious and more developed knowledge to produce a more comprehensible output compared to the learners with less hours of L2 exposure. Concerning the results in task 4, the non-statistically significant difference between the two groups could be explained by the fact that participants with more L2 exposure produced an elaborated and structured speech which resulted in a comprehensible speech. However, participants with less L2 exposure, when faced with a highly demanding cognitive effort regarding their age at the time of testing (8 years old), used a much simpler vocabulary, syntax and grammar to narrate their story. As a consequence, listeners rated their speech production as comprehensible as the other groups.

2.4. Accentedness

Concerning accentedness, learners were assessed in the sentence reading and the storytelling tasks. Results indicate that there was no significant difference between tasks 2 and 4. This means that task type did not have an impact on L2 accentedness, thus contradicting Skehan and Foster's (2001) Limited Attentional Capacity Model.

Concerning AOA, results of accentedness ratings in task 2 have shown that there was no statistically significant difference between older and younger learners. However, in task 4, younger learners had better accent ratings than the older ones. These results suggest that there was an effect of task type regarding AOA. As explained earlier, these findings could be explained by the fact that younger learners produced simpler utterances in the cognitive demanding tasks.

Regarding the amount of L2 exposure, learners with more hours of L2 exposure obtained better accentedness ratings in task 2 and task 4 than the groups of participants with less hours of L2 exposure. These results further confirm that task type did not have an impact on L2 accentedness regarding the amount of L2 exposure and that regardless of the task type, learners with more hours of L2 exposure succeeded in producing a less accented speech than the other groups. This could be justified by the fact that learner with more exposure mastered better the target accent consciously in task 2 and unconsciously in task 4.

2.5. Intelligibility

Finally, for intelligibility, learners were assessed based on the transcriptions of extracts from the sentence reading and storytelling tasks. Results indicate that participants obtained the highest mean score in task 2 and the lowest mean score in task 4. This suggests that task type had an impact on intelligibility. These results go in line with Skehan and Foster's (2001) Limited Attentional Capacity Model.

Concerning AOA, the results of intelligibility in tasks 2 and 4 indicated that there was no statistically significant difference between the older and younger learners. This signifies that task type did not have an effect on intelligibility regarding AOA. In other terms, regardless of task type, older and younger learners had similar intelligibility levels.

For the amount of L2 exposure, learners with more hours of L2 exposure obtained better intelligibility results in task 2 than the groups of participants with less hours of L2 exposure. However, there was a significant difference between the two groups in task 4. These results indicate that task type had an impact on L2 intelligibility regarding the amount of L2 exposure. The inconsistencies in the results could be explained by the fact that in the controlled tasks, learners with more exposure were able to fully monitor their use of the target forms which resulted in better results. Yet, in the spontaneous tasks, and in accordance with Skehan and Foster's (2001) Limited Attentional Capacity Model, they did not have the same performance as in task 2.

2.6. Specific and global pronunciation

Results of the statistical analyses for the specific pronunciation constructs indicated that task 4 is statistically different from tasks 1 and 2, while there was a significant statistical difference between all the other tasks. This concludes that there is no clear effect of task type on the specific constructs of pronunciation (segments and stress).

Regarding the global pronunciation constructs (comprehensibility, accentedness and intelligibility), the statistical analyses indicated that participants performed better in task 2 than in task 4. This conforms to Skehan and Foster's (2001) Limited Attentional Capacity Model.

The discrepancies in the results between the specific and the global pronunciation constructs suggest that task type has indeed an effect on the different dimensions of L2 pronunciation and that task type must be taken into consideration in the pronunciation studies.

2.7. General discussion

According to Saito and Plonsky (2019), researchers apply different measures to evaluate pronunciation without taking into consideration the impact of the choice of the task type on the participants' performance. Besides, Bergeron and Trofimovich (2017) report that prior empirical research investigated comprehensibility and accentedness using one methodological task: picture narrative. They added that most research examining task effects on L2 production targeted mostly grammar and vocabulary in terms of complexity and accuracy. However, there is limited research that has been done on the impact of task effect on L2 speech dimensions, with the majority of previous efforts focused on fluency.

For that, AOA and L2 exposure were investigated under four different task types. To summarize, the results for the task type effect on L2 pronunciation constructs yielded inconclusive results regarding AOA and L2 exposure. In fact, learners had better results in the spontaneous tasks for the global pronunciation constructs. Yet, there was no consistent results for the task type effects the specific on pronunciation constructs. Regarding task type effects, AOA and L2 exposure on the different dimensions of L2 pronunciation, we conclude that there are indeed task type effects. We highly suggest that AOA and exposure studies apply different task types that lure the conscious and unconscious capacities of learners.

To conclude, the findings in this dissertation allow us to answer the third research question:

- **Research question 3:** Will learners have better results in the controlled or in the spontaneous tasks or vice-versa?

Answer: Learners had mixed results for task type effects on the specific pronunciation constructs whereas they had better results in the controlled tasks in the general pronunciation constructs.

Chapter 6

Conclusion

This section presents a number of concluding remarks about the investigation that was undertaken to determine the effects of age, L2 exposure and task types on the acquisition of pronunciation of French L2 by Tunisian learners in the school setting. In the final part, the study limitations, contribution to knowledge and directions for future research are presented.

1. Summary of the major findings

The aim of this dissertation is to contribute to the field of instructed second language learning by demonstrating that AOA is not a key factor to a better L2 French pronunciation for Tunisian learners, but rather the quantity of L2 exposure. As a summary, it was found that for the pronunciation accuracy of almost all the four segments, stress placement accuracy, comprehensibility and accentedness, there was no statistically significant difference between the younger with AOA 4 years and the older learners with AOA 8 years. Regarding intelligibility, there was no statistically significant difference between the younger and the older learners and between learners with different amount of L2 exposure (1000h and 16000h).

In fact, the early introduction of FL learning in elementary schools can result in an increase in the number of hours of L2 instruction/ exposure and allows young learners to naturally grasp the implicit learning. The increased exposure to L2 at early age can have a positive impact on the academic and intercultural achievement, the phonological development (Jaekel et al., 2017, p. 633) and can also lead to the development of a more coherent linguistic system (Larson-Hall, 2008). Furthermore, the early introduction of L2 can result in fewer issues with language anxiety for younger learners, increased levels of motivation and positive attitudes towards other languages and cultures, all of which might lead to better future employment

(Jaekel et al., 2017; Nikolov & Mihaljevic Djigunovic, 2011; Muñoz & Spada, 2018). In conclusion, at the national policy level, successful early foreign language programs require appropriate time provision and curriculum that are founded on a thorough understanding of early language learning processes and establish realistic goals.

The second major finding of this dissertation is that it demonstrated that task type is an important factor that must not be overlooked where our results suggest that under four task conditions (from more controlled task, less controlled task, less spontaneous task to very spontaneous task), participants did not obtain better results in the controlled tasks in comparison to the spontaneous tasks but rather we obtained inconclusive results. There was no statistically significant difference between all the tasks except for task 1 (word list reading) where participants obtained the lowest mean results in task 1 for the segmental pronunciation accuracy. For stress placement accuracy, there was no significant difference between tasks 2 and 3 and between tasks 1 and 4 in which participants had the lowest results in the tasks 2 and 3 and the highest results in tasks 1 and 4. Concerning comprehensibility and accentedness, learners achieved better results in task 4 than in task 2 which contrast the findings for intelligibility where the participants got better results in task 2 than in task 4. For that, we highly suggest AOA studies to adopt different task types as they yield to different outcomes.

Finally, although it was not originally part of the objectives, is that it contributed to knowledge about which segments are the most challenging for the Tunisian learners of French L2. These sounds are the /ɛ̃/ and /y/ sounds. Therefore, we recommend a focus on the teaching of these segments in the official curriculum in Tunisia, particularly if they contribute to mutual intelligibility (Celce-Murcia et al, 2010).

2. Limitations

Several limitations should be acknowledged in terms of the methodology. These limitations are as follows:

- The limited number of participants: The findings of this research are based on a relatively small number of participants (N = 68). In fact, at the earliest stages of this dissertation in 2017/2018, it was planned to collect data on the national level in Tunisia, from the north (Tunis), the east coast (Mahdia and Sfax) and from the South (Gabes) in order to ensure the generalizability of the findings. However, due to COVID-19 pandemic, there has been several travelling restrictions between cities, curfews, and a general health jeopardy. Thus, in accordance with our thesis directors, we decided to collect data only in our native city, Sfax. Besides, several headmasters understandably declined our request to conduct the data collect in their schools for fear of their students' health. These conditions limited considerably the number of desired participants.
- The target segments: Although this study did not originally aim to investigate the segmental difficulties for Tunisian learners of French, we were surprised to find that empirical studies examining this issue are limited. The findings of these studies have shown that Tunisian learners face difficulties with almost all sounds in French that do not exist in the TA segmental inventory. As a consequence, a pilot study has been conducted to limit the number of segments to be studies. The choice of the four target segments was based on that pilot study but a thorough study is needed t to determine exactly which segments present difficulties for Tunisian learners of French L2.

- The quality and quantity of input: An important variable that must not be overlooked and that we were not able to control is the quality and quantity of input that participants have received. The dissertation was carried out in twelve different schools. As a result, it was difficult to examine teachers' oral proficiency and the quality of input that participants have received from the age of four. Besides, although a questionnaire was administered to collect data about the participants' extracurricular exposure to French, we were limited to data that the parents of the participants filled out. Therefore, we tried to control this variable by selecting participants that had minimal contact with the French outside the classroom.

3. Contribution to knowledge

Despite the limitations, this dissertation undertakes to contribute to knowledge methodologically and practically with regard to acquisition of L2 pronunciation, age and exposure effects and task effects. These contributions are the following:

- Try to contribute to knowledge by addressing a gap in the literature regarding the age effects studies and phonology in the school setting by investigating the global and specific dimensions of pronunciation.
- Attempt to provide further knowledge about the age effects issue in a new geographical space (Tunisia) with a target language other than English (French).
- Sought to give further evidence that the "younger, the better" belief is incorrect.
- Aimed to offer insights into the extent to which the types of tasks have an impact on learners' outcomes by examining the effects of age and exposure to L2 under four different task conditions.

- Endeavored to present support for stakeholders that successful early language programs require an adequate and an important amount of time.
- Intended to inform educators about which segments that learners struggle with, what the common pronunciation mistakes are and which linguistic features that they should focus on in the classroom.

4. Recommendations for future research

Based on our findings, we recommend introducing foreign languages early, with as many hours of L2 input as possible. However, this recommendation is based on a small-scale study that examined one linguistic dimension: pronunciation. For that, further investigations are suggested to address the present study's limitations. We suggest the following recommendations for future research:

- Replicating the study with a large-scale population.
- Performing longitudinal research to investigate the long-term effects of Language exposure on L2 pronunciation. Addressing this recommendation would require a longitudinal design with points of measurement that are sensitive to student development, at the end of each academic level, for example.
- This study examined solely the oral production level of the participants and did not take into consideration the perception level, it is recommended to examine the perception and the production of the L2 pronunciation in relation to AOA and L2 exposure since L2 learning takes place at the perception level before the production level (Flege, 2009).
- Although the oral production was produced under two speech conditions: controlled and spontaneous, the statistical analyses yielded conflicting results where participants

obtained better results in controlled tasks for the general pronunciation constructs and better results in spontaneous tasks for specific pronunciation constructs. The four tasks were elaborated while taking into consideration the young age of the participants. However, more attention should be given towards the conceptualization and refinement of L2 tasks that could dig into the complex nature of learners' speaking pronunciation abilities.

- We recommend conducting a study in a comparable school setting (either in the public sector or the private sector).
- We suggest administrating empirical research that examines the segmental difficulties for Tunisian learners of French.
- We endorse applying more elaborated tasks as the young age of the participants did not allow us to use more cognitively demanding tasks.

References

Abrahamsson, N., & Hyltenstam, K. (2009). *Age of Onset and Nativelikeness in a Second Language: Listener Perception Versus Linguistic Scrutiny*. *59*(2), 249–306. https://doi.org/10.1111/j.1467-9922.2009.00507.x

Abrahamsson, N. (2012). Age of onset and nativelike L2 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition*, *34*(2), 187–214. https://doi.org/10.1017/s0272263112000022

Abusharifa, Y. (2016). *Les difficultés spécifiques du français pour les élèves palestiniens* [Master's thesis, An-Najah National University]. https://staff-old.najah.edu/sites/default/files/mémoire%20master%20PDF.pdf

Al-Busaidi, S., & Al-Saqqaf, A. (2015). English Spelling Errors Made by Arabic-Speaking Students. *ELT*, *8*(7). https://doi.org/10.5539/elt.v8n7p181

Aldurayheem, A. A. (2020). *The influence of first language aptitude and print exposure (Arabic) on foreign language learning performance (English) and the effectiveness of the general aptitude test in Saudi Arabia* (Publication No. 10746190) [Doctoral dissertation, Liverpool John Moores University].

Ali, E. M. T. (2013). Pronunciation problems- Acoustic analysis of the English vowels produced by Sudanese learners of English. *International Journal of English and Literature*, *4*(10), 495–507.

Almbark, R., Bouchhioua, N., & Hellmuth, H. (2014). Acquiring the phonetics and phonology of English word stress: comparing learners from different L1 backgrounds. In *Proceedings of the International Symposium on the Acquisition of Second Language Speech: Concordia Working Papers in Applied Linguistics* (pp. 19–35).

Alqattan, S. (2015). *Early phonological acquisition by Kuwaiti Arabic children* [Doctoral dissertation, Newcastle University, UK].

Alvarez, E. (2006). Rate and route of acquisition in EFL narrative development at different ages. In C. Muñoz (Ed.), *Age and the Rate of Foreign Language Learning* (pp. 127–155). Clevedon: Multilingual Matters.

Ammar, Z. (2018). *Perception et production des voyelles orales françaises par des enfants tunisiens neo-apprenants du français* [Doctoral dissertation, Université Sorbonne Paris Cité]. https://tel.archives-ouvertes.fr/tel-02149376/document

Aouina, H. (2013). *Globalisation and language policy in Tunisia: Shifts in domains of use and linguistic attitudes* [Doctoral dissertation, University of the West of England]. https://core.ac.uk/download/pdf/16705709.pdf

Archibald, J. (2021). Ease and difficulty in L2 phonology: a mini-review. In J. Archibald, M. G. O'brien, & A. J. Sewell (Eds.), *L2 Phonology Meets L2 Pronunciation. Frontiers in Communication, Language Sciences*. doi: 10.3389/fcomm.2021.626529

Baccouche, T. (2009). Dynamique de la langue arabe. *Synergies Tunisie, 1*, 17–24.

Badwan, K. (2021). Agency in educational language planning: perspectives from higher education in Tunisia. *Current Issues in Language Planning*, *22*(1–2), 99–116. https://doi.org/10.1080/14664208.2019.1700056

Baidoun, T. (2015). *L'influence de la langue arabe sur le français des arabophones migrants* [Doctoral dissertation, Université de Rennes II]. Retrieved from *https://dumas.ccsd.cnrs.fr/dumas-01427030/document*

Baills, F., Alazard-Guiu, C., & Prieto, P. (2022). Embodied Prosodic Training Helps Improve Accentedness and Suprasegmental Accuracy. *Applied Linguistics*. https://doi.org/10.1093/applin/amac010

Ballais, J. L., al Amrawy, M., al Dbiyat, M., Charbel, L., Geyer, B., & Mezedjri, L. (2018). The Place of the French Language in Arabic-Speaking Mediterranean. *Handbook of the Changing World Language Map*, 1–13. https://doi.org/10.1007/978-3-319-73400-2_8-1

Bani Younes, M., & Hellmuth, S. (2020). The role of prosody in disambiguation of disjunctive questions in Jordanian Arabic. In *Arabic Linguistics Forum Conference, University of Leeds*.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL*, *44*(3), 231–250. https://doi.org/10.1515/IRAL.2006.010

Banerjee, J., Lestari, S. B., & Rossi, O. (2020). Choosing Test Formats and Task Types. In *The Routledge Handbook of Second Language Acquisition and Language Testing* (pp. 78-89). Routledge.

Barkat, M., Ohala, J. & Pellegrino, F. (1999). Prosody as a Distinctive Feature for the Discrimination of Arabic Dialects. In *Proceedings of the 6th European Conference on Speech Communication and Technology* (pp. 395–398), Budapest.

Barlow, J. (2001). Optimality Theory and the Assessment and Treatment of Phonological Disorders, *American Speech- Language-Hearing Association*. *32*(4), 242–256. https://doi.org/10.1044/0161-1461(2001/022)

Baumert, J., Fleckenstein, J., Leucht, M., Köller, O., & Möller, J. (2020). The Long-Term Proficiency of Early, Middle, and Late Starters Learning English as a Foreign Language at School: A Narrative Review and Empirical Study. *Language Learning, 70*(4), 1091–1135. https://doi.org/10.1111/lang.12414

Ben Hamida, T. B. A (2009). Erreurs interférentielles arabe-français et enseignement du français. *Synergies Tunisie,1*, 105-117.

Ben Romdhane, A. (2019). *Portrait of a Tunisian Arabic learner second language socialization in study abroad* (Doctoral dissertation). (Publication No. 27668458) [Doctoral dissertation, The University of Iowa]. ProQuest Dissertations and Theses Global.

Bent, T., Bradlow, A., & Smith, B. (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech. In O-S Bohn & M. Munro (Eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 331–347). Amsterdam: John Benjamins.

Bergeron, A., & Trofimovich, P. (2017). Linguistic Dimensions of Accentedness and Comprehensibility: Exploring Task and Listener Effects in Second Language French. *Foreign Language Annals*, *50*(3), 547–566. https://doi.org/10.1111/flan.12285

Bernhardt, B. M., & Stemberger, J. P. (2020). Chapter 11 Phonological development. In F. Li, K. E. Pollock, & R. Gibb (Eds.), *Child Bilingualism and Second Language Learning: Multidisciplinary perspectives* (pp. 223–248). John Benjamins. https://doi.org/10.1075/bpa.10.11ber

Best, C. T. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In H. Goodman & J. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: MIT Press

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109*(2), 775–794. https://doi.org/10.1121/1.1332378.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In O.-S. Bohn & M. J. Munro (eds.). *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). https://doi.org/10.1121/1.1332378.

Betti, M. J., & Ulaiwi, W. A. (2018). Stress in English and Arabic: A Contrastive Study. *English Language and Literature Studies*, *8*(1), 83. https://doi.org/10.5539/ells.v8n1p83

Birdsong, D. (2003). Authenticité de prononciation en français L2 chez des apprenants tardifs anglophones: analyses segmentales et globales. *Acquisition et interaction en langue étrangère, 18*, 17–36. https://doi.org/10.4000/aile.1150

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning, 56*(S1), 9–49. https://doi.org/10.1111/j.1467-9922.2006.00353.x

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O.-S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege. Language learning and language teaching,* Vol. 17 (pp. 99–116). Amsterdam, The Netherlands: John Benjamins.

Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology, 9*. https://doi .org /10 .3389 /fpsyg .2018 .00081.

Birdsong, D., & Molis, M. (2001). On the evidence for Maturational Constraints in second-language acquisition. *Journal of Memory and Language*, *44*(2), 235–249. https://doi.org/10.1006/jmla.2000.2750

Binasfour, H.S.M. (2018). *Investigating the perception and production of the Arabic pharyngealised sounds by L2 learners of Arabic* [Doctoral dissertation, University of Reading]. Retrieved from https://core.ac.uk/download/pdf/222830826.pdf

Boersma, P. & Weenink, D. (2021). *Praat: doing phonetics by computer* [Computer program]. Version 6.2.03, retrieved 3 December 2021 from http://www.praat.org/

Bouchhioua, N. (2009). Stress and accent in Tunisian Arabic. In *First international conference on intonational variation in Arabic* (pp. 28–29).

Bouchhioua, N. (2008). Duration as a cue to stress and accent in Tunisian Arabic, Native English, and L2 English. *Proceeding of 4th International Conference on Speech Prosody, Campinas, Brazil*.

Boukhari, A. (2006). *La réforme de l'enseignement du français en Tunisie : enjeux et difficultés. Le Français Aujourd'hui, 154*(3), 35–41. https://doi.org/10.3917/lfa.154.0035

Boyson, B. A., Semmer, M., Thompson, L. E., & Rosenbusch, M. H. (2013). Does beginning foreign language in kindergarten make a difference? Results of one district's study. *Foreign Language Annals, 46*(2), 246–263. https://doi.org/10.1111/flan.12023

Brkan, A. (2018). *Étude comparative des phénomènes de coarticulation nasale en anglais américain, bosnien, français, norvégien et ourdou*. [Doctoral dissertation, Université Sorbonne Paris Cité]. Retrieved from https://tel.archives-ouvertes.fr/tel-02172372

Brock, S. E. (2017). *Correlational Research*. Retrieved June 26, 2022, from https://www.csus.edu/indiv/b/brocks/courses/eds%20250/eds%20250/powerpoint/pdfs/presentation%208.pdf.

Busà, G. (1992). On the production of English vowels by Italian speakers with different degrees of accent. In J. Leather & A. James (Eds.), *New Sounds 92, Proceedings of the 1992 Amsterdam Symposium on the Acquisition of Second-language Speech* (pp. 47–63). University of Amsterdam.

Burstall, C., Jamieson, M., Cohen, S., & Hargreaves, M. (1974). *Primary French in the balance*. Windsor: NFER Publishing.

Cadette-Blasse, A. (2006). *Testing equivalence classification: The acquisition of French L2 phones by St. Lucian Créole French and St. Lucian Créole English speakers* [Master's thesis, Université Laval]. https://corpus.ulaval.ca/jspui/handle/20.500.11794/18250

Cadierno, T., & Eskildsen, S. W. (2018). The younger, the better? A usage-based approach to learning and teaching of English in Danish primary schools. *European Journal of Applied Linguistics, 6*(1), 171–182. https://doi.org/10.1515/eujal-2017-0018

Calaque, E. (1992). Les erreurs persistantes dans la production de locuteurs arabophones parlant couramment le français. *L'Information Grammaticale*, *54*(1), 48–51. https://doi.org/10.3406/igram.1992.3200

Cardoso, W. (2007). The development of sC onset clusters in interlanguage: Markedness vs. frequency effects. In: Slabakova R, Rothman J, Kempchinsky P, and Gavruseva E (Eds.) *Proceedings of the 9th Generative Approaches to Second Language Acquisition Conference (GASLA 2007)* (pp. 15–29). Somerville, MA: Cascadilla.

Caspers, J. (2010). The influence of erroneous stress position and segmental errors on intelligibility, comprehensibility and foreign accent in Dutch as a second language. *Linguistics in the Netherlands, 27*, 17–29. https://doi.org/10.1075/avt.27.03cas

Cebrian, J., Gorba, C., & Gavaldà, N. (2021). When the Easy Becomes Difficult: Factors Affecting the Acquisition of the English /iː/-/ɪ/ Contrast. *Frontiers in Communication*, *6*. https://doi.org/10.3389/fcomm.2021.660917

Celce-Murcia, M., Brinton, D. and Goodwin, J. (2010) *Teaching Pronunciation: A Course Book and Reference Guide*. Cambridge University Press.

Cenoz, J. (2002). Age differences in foreign language learning. *ITL - International Journal of Applied Linguistics, 135–136,* 125–142. https://doi.org/10.1075/itl.135-136.06cen

Cenoz, J. (2003). Facteurs déterminant l'acquisition d'une L3 : Age, Développement Cognitif et milieu. *Acquisition Et Interaction En Langue Étrangère*, (18), 38–51. https://doi.org/10.4000/aile.1151

Chahal, D. (2007). Intonation. In K . Versteegh (Ed). *Encyclopedia of Arabic Language and Linguistics*. Vol II, (pp. 395–401). Leiden: Brill.

Cheddad, B. (2011). Contribution de la phonétique à l'enseignement/apprentissage du Français dans le cycle primaire. *Synergies Algérie, 14*, p. 145–150.

Collins, B., & Mees, I. M. (2013). *Practical Phonetics and Phonology: A Resource Book for Students*. London: Routledge.

Collins, L., & Muñoz, C. (2016). The foreign language classroom: Current perspectives and future considerations. *Modern Language Journal*, *100*, 133–147. https://doi.org/10.1111/modl.12305

Côté, M.H. (2022). 21 French. In C. Gabriel, R. Gess & T. Meisenburg (Eds.). *Manual of Romance Phonetics and Phonology* (pp. 669–708). Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110550283-022

Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility. *Journal of Second Language Pronunciation*, *2*(2), 160–182. https://doi.org/10.1075/jslp.2.2.02cro

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2018). Linguistic dimensions of L2 accentedness and comprehensibility vary across speaking tasks. *Studies in Second Language Acquisition*, *40*(2), 443–457. https://doi.org/10.1017/s027226311700016x

Dai, W., & Wei, L. (2008). The Study of Chinese Learners' Acquisition of English Article Semantic Features. *Foreign Language Teaching and Research, 40*(2), 136–142.

De Bot, K. (2014). The effectiveness of early foreign language learning in the Netherlands. *Studies in Second Language Learning and Teaching, 3*, 409–418. doi:*10.14746/ssllt.2014.4.3.2*

De Foy, M.P. (2019). *French phonetic and phonological development in simultaneous bilingual acquisition - Individual trajectories and cross-linguistic effects in French-Italian/ Arabic/ Mandarin bilingual preschool children*. [Doctoral dissertation, Université de Mons].

de Jong, K., & Zawaydeh, B. A. (1999). Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics, 27*(1), 3–22. https://doi.org/10.1006/jpho.1998.0088

DeKeyser, R., M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning, 63*(s1), 52–67. https://doi.org/10.1111/j.1467-9922.2012.00737.x

DeKeyser, R., Alfi -Shabta, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics, 31*(3), 413–438. https://doi.org/10.1017/s0142716410000056

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 88–108). New York, NY : Oxford University Press.

Delais-Roussarie, E., Post, B., Avanzi, M., et al. (2015). Intonational phonology of French: Developing a ToBI system for French. In Frota, S., Prieto, P. (Ed), *Intonation in Romance* (pp. 63–100). Oxford: Oxford University Press.

Delattre, P. (1966). Les dix intonations de base du français. *French Review, 40*. 1-14.

Derwing, T. M. (2018). Putting an accent on the positive: New directions for L2 pronunciation research and instruction. *International Symposium on Applied Phonetics (ISAPh 2018)*, 19-21. https://doi.org/10.21437/isaph.2018-3

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in second language acquisition*, *19*(1), 1-16. https://doi.org/10.1017/s0272263197001010

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379-397. https://doi.org/10.2307/3588486

Derwing, T. M., & Munro, M. J. (2009). Comprehensibility as a factor in listener interaction preferences: Implications for the workplace. *Canadian Modern Language Review, 66*(2), 181–202. https://doi.org/10.3138/cmlr.66.2.181

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-Year study. *Language Learning*, *63*(2), 163–185. https://doi.org/10.1111/lang.12000

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

Derwing, T. M., & Munro, M. J. (2022). Pronunciation learning and teaching. In M.J., Munro, & R.I., Thomson (Eds.), *The Routledge Handbook of Second Language Acquisition and Speaking (1st ed.)* (pp. 147–159). Routledge. https://doi.org/10.4324/9781003022497-14

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). *Evidence in favor of a broad framework for instruction. Language Learning, 48*(3), 393–410. https://doi.org/10.1111/0023-8333.00047

Di Cristo, A. (1998). Intonation in French. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 195–218). Cambridge University Press.

Droua-Hamdani, G., Selouani, S. A., Boudraa, M., & Cichocki, W. (2010). Algerian Arabic rhythm classification. In Antonis, B. (Ed.), *Proceedings of the third ISCA tutorial and research workshop on experimental linguistics, ExLing 2010* (pp. 37-41). Athens, Greece.

Durand, J., Prince, T. (2015). Phonological markedness, acquisition and language pathology: what is left of the Jakobsonian legacy? In Astesano, C. Jucla, M. (Eds.), *Neuropsycholinguistic Perspectives on Language Cognition*. Series: Explorations in Cognitive Psychology. London: Routledge, Taylor & Francis group.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27*(2), 315–330. https://doi.org/10.1111/j.1467-1770.1977.tb00124.x

Eckman, F. R. (1985). Some theoretical and pedagogical implications of the markedness differential hypothesis. *Studies in Second Language Acquisition, 7*(3), 289-307. https://doi.org/10.1017/s0272263100005544

Eckman, F. R. (1996). A functional-typological approach to second language acquisition theory. In W. C. Ritchie, & T. K. Bhatia (Eds). *Handbook of second language acquisition*. *New* York: Academic Press.

Eckman, F. (2013). Markedness Differential Hypothesis (MDH). In Robinson, P. (Ed.) *Routledge Encyclopedia of Second Language Acquisition* (pp. 409-411). New York: Routledge.

Edwards, J., Zampini, M., & Cunningham, C. (2019). Listener proficiency and shared background effects on the accentedness, comprehensibility and intelligibility of four varieties of English. *Journal of Monolingual and Bilingual Speech, 1*(2), 333–356. https://doi.org/10.1558/jmbs.v1i2.11867

Ellis, R. (1985). A variable competence model of second language acquisition. *IRAL-International Review of Applied Linguistics in Language Teaching, 23*, 47–59.

El Zarka, D. (2017). Arabic Intonation. *Oxford Handbooks Online.* Retrieved from https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935345.001.0001/oxfordhb-9780199935345-e-77

Embarki, M. (2008). Les dialectes arabes modernes : état et nouvelles perspectives pour la classification géo-sociologique. *Arabica*, *55*(5), 583–604. https://doi.org/10.1163/157005808x364616

Embarki, M., Abou Haidar, L., Zeroual, C. & Naboulsi, R. (2016). Les apprenants et leur prononciation du français : description pédagogique. Les arabophones. In S. Detey, I.

Racine, Y. Kawaguchi, & J. Eychenne (Eds.), *La prononciation du français dans le monde. Du natif à l'apprenant* (pp.103-110). Paris : Clé International.

Esteve-Gibert, N., & Muñoz, C. (2021). Preschoolers benefit from a clear sound-referent mapping to acquire nonnative phonology. *Applied Psycholinguistics, 42*(1), 77-100. doi:10.1017/S0142716420000600

Fagyal, Z., Douglas, K., & Fred Jenkins. (2006). *French: A linguistic introduction*. Cambridge: Cambridge University Press.

Field, J. (2005). Intelligibility and the Listener: The Role of Lexical Stress. *TESOL Quarterly, 39*(3). 399-423. https://doi.org/10.2307/3588487

Fisch, S. M. (2017). Bridging theory and practice: Applying cognitive and educational theory to the design of Educational Media. *Cognitive Development in Digital Contexts*, 217–234. https://doi.org/10.1016/b978-0-12-809481-5.00011-0

Flege, J. E. (1992). Speech Learning in a Second Language. In C. A. Ferguson, L. Menn, and C. Stoel-Gammon (Eds.), *Phonological Development: Models, Research, Implications* (pp. 565–604). Timonium, MD: York Press.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America, 93*(3), 1589–1608. https://doi.org/10.1121/1.406818

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). Timonium, MD: York Press.

Flege, J. E. (1999). The relation between L2 production and perception. In *Proceedings of the XIVth International Congress of Phonetics* Sciences (pp. 1273–1276), Berkeley.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception . In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319 – 355). Berlin, Germany : Mouton de Gruyter .

Flege, J. E. (2009). Give input a chance! In T. Piske & M. Young- Scholten (Eds.), *Input Matters in SLA* (pp. 175–190). Bristol: Multilingual Matters.

Flege, J. E., Aoyama, K., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r) applied. In R. Wayland (Ed.) *Second language speech learning: Theoretical and empirical progress* (pp. 84–118). Cambridge, UK: Cambridge University Press.

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In Wayland, R. (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge: Cambridge University Press.

Flege, J. E., Bohn, O.-S., & Sunyoung, J. (1997). Effects of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470. https://doi.org/10.1006/jpho.1997.0052

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics, 15*, 67-83. https://doi.org/10.1016/S0095-4470(19)30538-8

Flege, J., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*(4), 527–552. https://doi.org/10.1017/s0272263101004041

Flege, J. E., & MacKay, I. R. A. (2010). ‘‘Age’’ effects on second language acquisition. In *Proceedings of the sixth international symposia on the acquisition of second-language speech*. May 1–3, 2010, Poznan, Poland.

Flege, J. E., & MacKay, I. R. A. (2011). What accounts for age effects on overall degree of foreign accent? In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Polish studies in English language and literature: Achievements and perspectives in SLA of speech* (pp. 65–82). Frankfurt am Main, Germany: Peter Lang.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Effects of age of second- language learning on the production of English consonants. *Speech Communication, 16*(1), 1–26. https://doi.org/10.1016/0167-6393(94)00044-b

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78–104. https://doi.org/10.1006/jmla.1999.2638

Fogerty, D., & Humes, L. E. (2012). The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. *The Journal of the Acoustical Society of America*, *131*(2), 1490–1501. https://doi.org/10.1121/1.3676696

Foote, J.A. (2015). *Pronunciation Pedagogy and Speech Perception: Three Studies.* [Doctoral dissertation, Concordia University]. https://spectrum.library.concordia.ca/id/eprint/980242/

Foote, J. A., & Trofimovich, P. (2018). Second language pronunciation learning: An overview of theoretical perspectives. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The*

Routledge handbook of contemporary English pronunciation (pp. 75-90). London: Routledge.

Fougeron, C., & Smith, C. (1993). French: Illustrations of the IPA. *Journal of the International Phonetic Association, 23*(2), 73-76.

Fox, A. (2000). *Prosodic features and prosodic structure: The phonology of suprasegmentals*. Oxford: Oxford University Press.

Fraser, H. (2001). *Teaching Pronunciation: A Handbook for Teachers and Trainers. Three Frameworks for an Integrated Approach*. Sydney, NSW: Department of Education Training and Youth Affairs.

Fraser, H. (2010). Cognitive theory as a tool for teaching pronunciation. In *Cognitive theory as a tool for teaching pronunciation* (pp. 357–380). De Gruyter Mouton. https://doi.org/10.1515/9783110245837.357

Ftita A. (2019). Améliorer la lutte contre l'échec en FLE/FLS pris pour facteur aggravant le décrochage scolaire en Tunisie. L'exploitation des erreurs lexicales intra-linguales et interlinguales. In P. Y. Bernard (dir). *Actes du colloque International* (TEDS- 2018) (pp. 1-15), Nantes, France : CREN.

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, *42*(1), 120–133. https://doi.org/10.1002/ejsp.862

Fullana, N. (2006). The Development of English (FL) Perception and Production Skills: Starting Age and Exposure Effects. In *Chapter 2. The Development of English (FL) Perception and Production Skills: Starting Age and Exposure Effects* (pp. 41–64). Multilingual Matters. https://doi.org/10.21832/9781853598937-004

Fullana, N., & MacKay, I. R. A. (2003). Production of English sounds by EFL learners: The case of /i:/ and /ɪ/. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1525–1528.

Fullana, N., & MacKay, I. R. A. (2010). L2 fluency in the production of English extemporaneous speech by Catalan/Spanish bilinguals. In *Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 125–130).

Fullana, N. & Mora, J. C. (2007). Production and perception of voicing contrasts in English word-final obstruents: Assessing the effects of experience and starting age. In M. Watkins, A. Rauber & B. Baptista (Eds.), *Recent research in second language phonetics/ phonology: perception and production* (pp. 97-117). Newcastle: Cambridge Scholar Publishing.

Garcia Lecumberri, M. L. G. & Gallardo, F. (2003). English FL sounds in school learners of different ages. In M. P. G. Mayo & M. L. G. Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 115–135). Clevedon: Multilingual Matters.

García Mayo, M. del P., & García Lecumberri, M. L. (Eds.). (2003). *Age and the acquisition of English as a Foreign Language.* Clevedon: Multilingual Matters.

Gass, S. M., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An introductory course* (4th ed.). New York: Routledge.

Gass, S. M., & Selinker, L. (2008). *Second Language Acquisition: An introductory Course*, *3rd edition*. Taylor and Francis.

Georgiou, G. P., Perfilieva, N., Denisenko, V., & Novoskaspaya, N. (2020). Perceptual realization of Greek consonants by Russian monolingual speakers. *Speech Communication*, *125*, 7–14. https://doi.org/10.1016/j.specom.2020.09.008

Ghanem, R., & Kang, O. (2018). Pronunciation features in rating criteria. In O. Kang & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 115–136). London: Routledge.

Ghazali, S. (1973). *Tunisian Arabic and French Phonological Interference with English as a Second Foreign Language.* [Master's thesis, University of California].

Ghazali S., Hamdi R. and Barkat M. (2002). Speech rhythm variation in Arabic dialects. In *Speech Prosody 2002, International Conference*, Aix-en-Provence, France. International Speech Communication Association (ISCA).

Gooch, R., Saito, K., & Lyster, R. (2016). Effects of recasts and prompts on L2 pronunciation development: Teaching English /ɹ/ to Korean adult EFL learners. *System, 60*, 117–127. https://doi.org/10.1016/j.system.2016.06.007

Goorhuis-Brouwer, S., & de Bot, K. (2010). Impact of early English language teaching on L1 and L2 development in children in Dutch schools. *International Journal of Bilingualism, 14*(3), 289–302. https://doi.org/10.1177/1367006910367846

Graham, S., Courtney, L., Marinis, T., & Tonkyn, A. (2017). Early language learning: The impact of teaching and teacher factors. *Language Learning, 67*(4), 922–958. https://doi.org/10.1111/lang.12251

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research 29*(3), 311–343. https://doi.org/10.1177/0267658312461497

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge, UK: Cambridge University Press.

Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*(2), 201. https://doi.org/10.2307/3588378

Hamrouni, N. (2010). *Structure and processing in Tunisian Arabic: speech error data* (Doctoral dissertation). [Doctoral dissertation, The University of Arizona]. ProQuest Dissertations and Theses Global. http://hdl.handle.net/10150/195969

Hannah, L., Kim, H., & Jang, E. E. (2022). Investigating the Effects of Task Type and Linguistic Background on Accuracy in Automated Speech Recognition Systems: Implications for Use in Language Assessment of Young Learners. *Language Assessment Quarterly*, *19*(3), 289–313. https://doi.org/10.1080/15434303.2022.2038172

Harding, L. (2018). Validity in pronunciation assessment. *Assessment in Second Language Pronunciation*, 30–48. https://doi.org/10.4324/9781315170756-3

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hellmuth, S. (2013). Phonology. In: Owens, J. (Ed.), *The Oxford handbook of Arabic linguistics* (pp. 45–70). Oxford University Press. DOI: 10.1093/oxfordhb/9780199764136.013.0003

Hellmuth, S. (2018). Variation in polar interrogative contours within and between Arabic dialects. In *Proceedings of Speech Prosody. Poznań, Poland, 13-16 June.*

Hopp, H.,& Schmid, M. (2013). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics, 34*(2), 361–394. https://doi.org/10.1017/s0142716413000015

Hopp, H., Vogelbacher, M., Kieseier, T., & Thoma, D. (2019). Bilingual advantages in early foreign language learning: Effects of the minority and the majority language. *Learning and Instruction, 61*, 99–110. https://doi.org/10.1016/j.learninstruc.2019.02.001

Huang, B. H. (2014). The effects of age on second language grammar and speech production. *Journal of Psycholinguistic Research, 43*(4), 397–420. https://doi.org/10.1007/s10936-013-9261-7

Huang, B. H. (2016). A synthesis of empirical research on the linguistic outcomes of early foreign language instruction. *International Journal of Multilingualism, 13*(3), 257–273. https://doi.org/10.1080/14790718.2015.1066792

Huang, B. H., & Jun, S-A. (2011). The effect of age on the acquisition of second language prosody. *Language and Speech, 54*(3), 387–414. https://doi.org/10.1177/0023830911402599

Hume, E. (2011). Markedness. In M. V. Oostendorp, C. Ewen, E. Hume & K. Rice (Eds.), *Companion to Phonology*. Blackwell.

Ibrahim, O., Asadi, H., Kassem, E., & Dellwo, V. (2020). Arabic speech rhythm corpus: Read and spontaneous speaking styles. In *Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 5337-5342).

Inceoglu, S. (2018). Exploring the effects of instruction on L2 French learner pronunciation, accentedness, comprehensibility, and fluency: An online classroom study. *Journal of Second Language Pronunciation*, *5*(2), 223–247. https://doi.org/10.1075/jslp.18004.inc

Inceoglu, S. (2019). Individual Differences in L2 Speech Perception: The Role of Phonological Memory and Lipreading Ability. *The Modern Language Journal*, *103*(4), 782–799. https://doi.org/10.1111/modl.12591

Isaacs, T., & Trofimovich, P. (2012). DECONSTRUCTING COMPREHENSIBILITY: Identifying the Linguistic Influences on Listeners' L2 Comprehensibility Ratings. *Studies in Second Language Acquisition, 34*(3), 475-505. doi:10.1017/S0272263112000150

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*(2), 135–159. https://doi.org/10.1080/15434303.2013.769545

Isaacs, T., & Trofimovich, P. (2016). *Second Language Pronunciation Assessment: Interdisciplinary Perspectives (Second Language Acquisition)* (Illustrated ed.). Multilingual Matters.

Isaacs, T., Trofimovich, P., & Foote, J. A. (2018). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing, 35*(2), 193–216. https://doi.org/10.1177/0265532217703433

Jaekel, N., Schurig, M., Florian, M., & Ritter, M. (2017). From Early Starters to Late Finishers? A Longitudinal Study of Early Foreign Language Learning in School. *Language Learning, 67*(3), 631–664. https://doi.org/10.1111/lang.12242

Jaekel, N., Schurig, M., van Ackern, I., and Ritter, M. (2022). The impact of early foreign language learning on language proficiency development from middle to high school. *System, 106*. doi: 10.1016/j.system.2022.102763

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford: Oxford University Press.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in Second language learning: The influence of maturational state on the acquisition of English as a Second language. *Cognitive Psychology*, *21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Jun, S.A., and Cowie, I. (1994). Interference for 'New' and 'Similar' Vowels in Korean Speakers of English. Ohio State Univ. *Working Papers. 43*, 117–130.

Jun, S. A. and Fougeron, C. (2000). A phonological model of French intonation. In: Botinis, A. (ed.) *Intonation: Analysis, modelling and technology* (pp. 209–242). Dordrecht: Kluwer.

Kallio, H. (2022). *The prosody underlying spoken language proficiency. Cross-lingual investigation of non-native fluency and syllable prominence* [Doctoral dissertation, University of Helsinki]. http://urn.fi/URN:ISBN:978-951-51-7845-9

Kamerhuber, J., Horvath, J., & Pustka, E. (2020). Lecture, répétition, Parole Spontanée : L'impact de la tâche sur le comportement du schwa en fle. *Journal of French Language Studies*, *30*(2), 161–188. https://doi.org/10.1017/s095926952000006x

Kamiyama, T. (2009). *Apprentissage phonétique des voyelles du français langue étrangère chez des apprenants japonophones.* [Doctoral dissertation, Université Sorbonne Nouvelle – Paris 3]. Retrieved from https://tel.archives-ouvertes.fr/tel-00473029v2

Kandeel, H. R., & Marcy, V. (2019). La correction phonétique de la prononciation du français dans un contexte universitaire Jordanien. *Onomázein Revista de Lingüística Filología Y Traducción*, *45*, 79–97. https://doi.org/10.7764/onomazein.45.03

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, *38*(2), 301–315. https://doi.org/10.1016/j.system.2010.01.005

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental Measures of Accentedness and Judgments of Language Learner Proficiency in Oral English. *The Modern Language Journal*, *94*(4), 554–566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Kang, O., Thomson, R. I., & Moran, M. (2018). Empirical approaches to measuring the intelligibility of different varieties of English in predicting listener comprehension. *Language Learning, 68*(1), 115–146. https://doi.org/10.1111/lang.12270

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review, 64*(3),459–490. https://doi.org/10.3138/cmlr.64.3.459

Khalifa, M. F. (2020). *Errors in English Pronunciation Among Arabic Speakers: Analysis and Remedies*. Cambridge Scholars Publishing.

Kossentini, R. (2021). Éléments de réflexion pour le discours formatif en matière de politiques linguistiques en Tunisie. *Synergies France, (14-15)*,47-59.

Kuiken, V., & Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA* (pp. 143-169). Amsterdam: John Benjamins.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.

Lado, R. (1957). *Linguistics Across Cultures: Applied linguistics for language teachers.* Ann Arbor: University of Michigan Press.

Lambelet, A., & Berthele, R. (2015). *Age and foreign language learning in school*. New York, NY: Palgrave Macmillan.

Lan, Y. (2014). Detecting L2 speech deviations by a communicative experiment procedure: Cantonese speakers' realizations of English /r/. In J. Levis & S. McCrocklin (Eds). *Proceedings of the 5th Pronunciation in Second Language Learning and Teaching Conference* (pp. 75-86). Ames, IA: Iowa State University.

Langerová, M. (2012). *Diglossie au Maghreb - Histoire et situation actuelle* [Bachelor degree, Universitas Masarkiana Brunensis]. https://is.muni.cz/th/262295/ff_m/magisterska_diplomova_prace.pdf

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24*(1), 35-63. doi: 10.1177/0267658307082981

Lauret, B. (2007). *Enseigner la prononciation du français: questions et outils*. Paris, Hachette.

Lenneberg, E. (1967). *The biological foundations of language*. New York: John Wiley & Sons.

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

Levis, J. M. (2006). Pronunciation and the assessment of spoken language. In R. Hughes (Ed.), *Spoken English, TESOL and applied linguistics: Challenges for theory and practice* (pp. 245–270). New York: Palgrave Macmillan.

Levis, J. M. (2018a). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge: Cambridge University Press.

Levis, J. M. (2018b). Segmentals and Intelligibility. In *Intelligibility, Oral Communication, and the Teaching of Pronunciation* (Cambridge Applied Linguistics, pp. 61-85). Cambridge: Cambridge University Press. doi:10.1017/9781108241564.006

Levis, J. (2020). Revisiting the nativeness and intelligibility principles. *Journal of Second Language Pronunciation, 6*(3), 310–328. https://doi.org/10.1075/jslp.20050.lev

Levis, J. M., & Silpachai, A. O. (2022). Speech intelligibility. In M.J., Munro, & R.I., Thomson (Eds.), *The Routledge Handbook of Second Language Acquisition and Speaking (1st ed.)* (147–159). Routledge.

Levis, J., & Wichmann, A. (2015). English intonation-form and meaning. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 139–155). West Sussex. UK: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118346952.ch8

Lightbown, P. M., & Spada, N. (2020). Teaching and learning L2 in the classroom: It's about time. *Language Teaching, 53*(4), 422–432. https://doi.org/10.1017/S0261444819000454

Liu, D., & Reed, M. (2019). Asymmetrical cognitive load imposed by processing native and non-native speech. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 179-190). Ames, IA: Iowa State University.

Liu, D., & Reed, M. (2021). Exploring the Complexity of the L2 Intonation System: An Acoustic and Eye-Tracking Study. *Front. Commun.* 6, 627316. doi:10.3389/fcomm.2021.627316

Long, M. H. (1990). Maturational Constraints on language development. *Studies in Second Language Acquisition*, *12*(3), 251–285. https://doi.org/10.1017/s0272263100009165

Long, M. (2005). Problems with supposed counter-evidence to the critical period hypothesis. *IRAL - International Review of Applied Linguistics in Language Teaching*, *43*(4), 287–317. https://doi.org/10.1515/iral.2005.43.4.287

Ludwig, A., & Mora, J. C. (2017). Processing time and comprehensibility judgments in non-native listeners' perception of L2 speech. *Journal of Second Language Pronunciation*, *3*(2), 167–198. https://doi.org/10.1075/jslp.3.2.01lud

MacKay, I. R. A., & Fullana, N. (2007). Starting age and exposure effects on EFL learners' sound production in a formal learning context. In M. Watkins, A. Rauber & B. Baptista (Eds.), *Recent research in second language phonetics/ phonology: perception and production* (pp. 43-63). Newcastle: Cambridge Scholar Publishing.

Major, R. C. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah, NJ: Lawrence Erlbaum Associates

Major, R. C. (2008). Transfer in second language phonology: A review. *Phonology and second language acquisition, 36*, 63-94. https://doi.org/10.1075/sibil.36.05maj

Major, R. C., & Kim, E. (1996). The similarity differential rate hypothesis. *Language Learning*, *46*(3), 465–496. https://doi.org/10.1111/j.1467-1770.1996.tb01244.x

Marçais, P. (1977). *Esquisse grammaticale de l'arabe maghrébin*. Paris: Adrien Maisonneuve

Marinova-Todd, S.H., Marshall, B., & Snow, C.E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly, 34*(1), 9–34. https://doi.org/10.2307/3588095

Masmoudi, A., Khmekhem, M., Estéve, Y., F. Bougares, F., & Belguith, L. (2014). Phonetic tool for the Tunisian Arabic. In *Proceedings of the 4th International Workshop on Spoken Language Technologies for Under-resourced Languages*. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1083.4485&rep=rep1&type=pdf

Maume, J.L. (1973). L'apprentissage du Français chez les Arabophones Maghrébins (diglossie et plurilinguisme en Tunisie). *Langue Française*, *19*(1), 90–107. https://doi.org/10.3406/lfr.1973.5642

Mayo, M. del P. (2003). Age, Length of Exposure and Grammaticality Judgements in the Acquisition of English as a Foreign Language. In M. Garcia Mayo & M. Lecumberri (Eds.), *Chapter 5. Age and the acquisition of English as a foreign language* (pp. 94-114). Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853596407-006

McCarthy, J. J. (2007). What is optimality theory? *Language and Linguistics Compass*, *1*(4), 260–291. https://doi.org/10.1111/j.1749-818x.2007.00018.x

Mehrdad, A. G., & Ahghar, M. R. (2015). Markedness and Syllabus Design in SLA. *Procedia - Social and Behavioral Sciences*, *177*, 104–108. https://doi.org/10.1016/j.sbspro.2015.02.348

Mejri, S. (2003). *Syntaxe et sémantique, revue de l'Université de Caen, numéro sur Polysémie et polylexicalité.* Presses Universitaires de Caen.

Meisel, J.M. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft, 28*(1), 5–34.

Mejri, S., Said, M., & Sfar, I. (2009). Pluringuisme et diglossie en Tunisie. *Synergies Tunisie. 1*, 53-74.

Mestre, L. S. (2012). Pedagogical Considerations for Tutorials. *Designing Effective Library Tutorials*, 141–169. https://doi.org/10.1016/b978-1-84334-688-3.50007-x

Michel, M. (2017). Complexity, accuracy, and fluency in L2 production. In S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 50-68). London-New York: Routledge.

Mihaljevic Djigunovic, J., Nikolov, M., & Otto, I. (2008). A comparative study of Croatian and Hungarian EFL students. *Language Teaching Research, 12*, 433–452. https://doi.org/10.1177/1362168808089926

Miled, M. (2007). Le français langue seconde en Tunisie : une évolution sociolinguistique et didactique spécifique. *Le français aujourd'hui*, 156, 79-86. https://doi.org/10.3917/lfa.156.0079

Miller, C., & Caubet D. (2010). Arabic sociolinguistics in the Middle East and North Africa (MENA). In M. J. Bal (ed.), *The Routledge Handbook of Sociolinguistics Around the World* (pp. 238-256). London-New York: Routledge.

Ministère de l'Éducation. (2004). *Les programmes officiels : 2e degré de l'enseignement de base*. Tunis. CNP.

Miralpeix, I. (2006). Age and vocabulary acquisition in English as a foreign language. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 89–106). Clevedon: Multilingual Matters.

Mora, J. C. (2006). Age effects on oral fluency development. In C. Muñoz (Ed.), *Chapter 3. Age and the rate of foreign language learning* (Vol. 19, pp. 65-88). Clevedon: Multilingual Matters.

Mora, J. C. & Fullana, N. (2007). Production and perception of English /i:/-/ɪ/ and /æ/-/ʌ/ in a formal setting: Investigating the effects of experience and starting age. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th international congress of phonetic sciences* (Vol. 3, pp. 1613-1616). Saarbrücken: Universität des Saarlandes.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, *21*(1), 81–108. https://doi.org/10.1017/s0272263199001035

Moyer, A. (2014). What's age got to do with it? Accounting for individual factors in second language accent. *Studies in Second Language Learning and Teaching, 4*(3), 443-464. doi: 10.14746/ssllt.2014.4.3.4

Muñoz, C. (2003a). Variation in Oral Skills Development and Age of Onset. In Muñoz, C. (ed.), *Age and the Acquisition of English as a Foreign Language*. Clevedon: Multilingual Matters, 161–181.

Muñoz, C. (2003b). Le rythme d'acquisition des savoirs communicationnels chez des apprenants Guidés. *Acquisition Et Interaction En Langue Étrangère*, (18), 53–77. https://doi.org/10.4000/aile.1152

Muñoz, C. (2006a). *Age and the Rate of Foreign Language Learning*. Clevedon: Multilingual Matters.

Muñoz, C. (2006b). The BAF project: Research on the effects of age on foreign language acquisition. In C. Abelló-Contesse, R. Chacón- Beltrán & M. Dolores López-Jiménez (Eds.), *Age in L2 Acquisition and Teaching* (pp. 81-92). Bern: Peter Lang.

Muñoz, C. (2006c). The effects of age on foreign language learning: The BAF project. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 1-40). Clevedon: Multilingual Matters.

Muñoz, C. (2008). Age-related differences in foreign language learning. Revisiting the empirical evidence. I*RAL - International Review of Applied Linguistics in Language Teaching*, 46(3), 197–220. https://doi.org/10.1515/IRAL.2008.009

Muñoz, C. (2010). On how age affects foreign language learning. *Advances in Research on Language Acquisition and Teaching: Selected Papers*, 39–49.

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics, 35*(4), 463-482. doi:10.1093/applin/amu024

Muño , C. (2019). A new look at "age": Young and old l2 learners. In *The Cambridge Handbook of Language Learning* (pp. 430–450). https://doi.org/10.1017/9781108333603.019

Muñoz, C., & Singleton, D. (2007). Foreign accent in advanced learners. Two successful profiles. *Eurosla Yearbook, 7*, 171–190. https://doi.org/10.1075/eurosla.7.10mun

Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching, 44* (1), 1-35. https://doi.org/10.1017/S0261444810000327

Muñoz, C., & Spada, N. (2018). Foreign Language Learning from Early Childhood to Young Adulthood. In A. De Houwer & L. Ortega (Eds.), *The Cambridge Handbook of Bilingualism* (pp. 233-249). Cambridge: Cambridge University Press. doi:10.1017/9781316831922.013

Munro, M. J. (2017). Dimensions of pronunciation. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 413–431). New York, NY: Routledge.

Munro, M. (2011). Intelligibility: Buzzword or buzzworthy? In. J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd pronunciation in second language learning and teaching conference*, Sept. 2010. (pp. 7–16). Ames, IA: Iowa State University.

Munro, M.J., & Derwing, T.M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73–97. Doi: https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289–306. https://doi.org/10.1177/002383099503800305

Munro, M. J. & Derwing, T. M. (1999). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning, 49* (Supp. 1), 285-310. doi: 10.1111/0023-8333.49.s1.8

Munro, M. J., & Derwing, T. M. (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. https://doi.org/10.1017/s0272263101004016

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, *34*(4), 520–531. https://doi.org/10.1016/j.system.2006.09.004

Munro, M. J. & Derwing, T. M. (2015). Intelligibility in research and practice: Teaching priorities. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 377-396). Chicester: Wiley Blackwell.

Munro, M. J., & Derwing, T. M. (2019). Phonetics and second language teaching research. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge handbook of phonetics* (pp. 473–495). Routledge.

Munro, M. J., & Derwing, T. M. (2020). Foreign accent, comprehensibility and intelligibility, redux. *25 Years of Intelligibility, Comprehensibility and Accentedness*, *6*(3), 283–309. https://doi.org/10.1075/jslp.20038.mun

Munro, M. J., Derwing, T. M., & Burgess, C. S. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, *52*(7-8), 626–637. https://doi.org/10.1016/j.specom.2010.02.013

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility Of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 111–131. *https://doi.org/10.1017/S0272263106060049*

Munro, M. J., Flege, J. E., & Mackay, I. R. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, *17*(3), 313–334. https://doi.org/10.1017/s0142716400007967

Murphy, V. A. (2014). S*econd Language Learning in the Early School Years: Trends and Contexts - Oxford Applied Linguistics*. Oxford University Press.

Mustafawi, E. (2018). Arabic phonology. In E. Benmamoun, & R. Bassiouney (Eds.), *The Routledge handbook of Arabic linguistics* (pp. 11-31). New York, NY: Routledge.

Mzoughi, I. (2015). *Intégration des emprunts lexicaux au français en arabe dialectal tunisien* [Doctoral dissertation, Université de Cergy Pantoise]. Retrieved from *https://tel.archives-ouvertes.fr/tel-01220289/document*

Nagle, C. (2022). Individual Difference Factors for Second Language Pronunciation. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge Handbook of Second Language Acquisition and Individual Differences* (pp. 269-281). New York and London: Routledge. ISBN 9781032219141. DOI: 10.4324/9781003270546-23

Nagle, C. L., & Baese-Berk, M. M. (2021). Advancing the state of the art in L2 speech perception-production research: Revisiting theoretical assumptions and Methodological Practices. *Studies in Second Language Acquisition*, *44*(2), 580–605. https://doi.org/10.1017/s0272263121000371

Nagle, C. L., & Huensch, A. (2020). Expanding the scope of L2 intelligibility research. *25 Years of Intelligibility, Comprehensibility and Accentedness*, *6*(3), 329–351. https://doi.org/10.1075/jslp.20009.nag

Navés, T., Torras, M. R., & Celaya, M. L. (2003). Long-term effects of an earlier start. *EUROSLA Yearbook*, *3*, 103–129. https://doi.org/10.1075/eurosla.3.08nav

Nawafleh, A. (2016). Le rôle de la phonétique dans l'enseignement/apprentissage du français langue étrangère, étude acoustico-perceptive. *Jordan Journal of Modern Languages and Literature. 9* (1). 73-94. Retrieved from http://journals.yu.edu.jo/jjmll/Issues/vol9no12017/Nom5.pdf

Nawafleh, A., & Alrabadi, E. (2017). Étude acoustique et perceptive des voyelles du français réalisées par des apprenants jordaniens et des locuteurs français. *Çédille, revista de estudios franceses*, (13), 351-378.

Newton, J. M. (2018). Introduction to Speaking. In Newton, J., Ferris, D., Goh, C. Grabe, W., Stoller, F., & Vandergrift, L. (Eds.) Griffith, C., Tajeddin, Z. & Brown, A. (Eds.). *Teaching English to Second Language Learners in Academic Contexts.* New York: Routledge.

Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In E. Dupoux (Ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler*. Cambridge, MA: MIT Press.

Nikolov, M., & Mihaljević Djigunović, J. (2006). Recent research on age, second language acquisition, and early foreign language learning. *Annual Review of Applied Linguistics, 26*, 234–260. doi:10.1017/S0267190506000122

Nikolov, M., & Mihaljević Djigunović, J. (2011). All shades of every color: An overview of early teaching and learning of foreign languages. *Annual Review of Applied Linguistics, 31*, 95–119. https://doi.org/10.1017/S0267190511000183

Norris, J. M., & Ortega, L. (2012). Assessing learner knowledge. In T*he Routledge Handbook of Second Language Acquisition*. Routledge.

O'Connor, J. D. (1973). *Phonetics*. Harmondsworth, Middlesex: Penguin Books.

Odisho, E. Y. (2005). *Techniques of Teaching Comparative Pronunciation in Arabic and English*. New Jersey: Gorgias Press.

Official Journal of the Tunisian Republic. (2002). *Loi d'orientation n° 2002-80 du 23 juillet 2002, relative à l'éducation et à l'enseignement scolaire.* Imprimerie Officielle de la République Tunisienne. *https://www.ilo.org/dyn/natlex/natlex4.detail?p_isn=61806&p_lang=fr*

Oller, J., Jr., & Nagato, N. (1974). The long-term effect of FLES: An experiment. *Modern Language Journal, 58*(1/2), 15–19. https://doi.org/10.2307/323984.

Osifeso, E. (2022). An Optimality Approach to Word Stress Analysis in Yoruba-English. *Journal of the Phonetician & Phonologists in Nigeria, 1*, 27-54.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*(3), 261–283. https://doi.org/10.1007/bf01067377

Papi, M., Vasylets, O., & Ahmadian, M. J. (2022). Individual difference factors for second language writing. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge Handbook of Second Language Acquisition and Individual Differences* (pp. 381–396). New York and London: Routledge. ISBN 9781032219141. https://doi.org/10.4324/9781003270546-30

Parlak, Ö. (2018). *Interaction, corrective feedback, and the development of lexical stress* [Doctoral dissertation, Lancaster University]. https://doi.org/10.17635/lancaster/thesis/520

Penfield, W. & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, NJ: Princeton University Press.

Pfenninger, S. E. (2011). Age effects on the acquisition of nominal and verbal inflections in an instructed setting. *Studies in Second Language Learning and Teaching, 1*(3), 401–420. https://doi.org/10.14746/ssllt.2011.1.3.5

Pfenninger, S. E. (2014). The misunderstood variable: Age effects as a function of type of instruction. *Studies in Second Language Learning and Teaching, 3*, 529–556. https://doi.org/10.14746/ssllt.2014.4.3.8

Pfenninger, S. E., & Singleton, D. (2017a). *Beyond age effects in instructional L2 learning: Revisiting the age factor*. Bristol: Multilingual Matters.

Pfenninger, S. E., & Singleton, D. (2017b). Beyond Age Effects in Instructional L2 Learning: Revisiting the Age Factor. In S. E. Pfenninger & D. Singleton (Eds.), *Beyond Age Effects in Instructional L2 Learning. Multilingual Matters*. https://doi.org/10.21832/9781783097630

Pfenninger, S. E., & Singleton, D. (2019). Starting Age Overshadowed: The Primacy of Differential Environmental and Family Support Effects on Second Language Attainment in an Instructional Context. *Language Learning, 69*, 207–234. https://doi.org/10.1111/lang.12318

Phakiti, A (2014) *Experimental research methods in languages learning*. Bloomsbury.

Pinter, A. (2011). *Children learning second languages*. Palgrave Macmillan.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A Review. *Journal of Phonetics*, *29*(2), 191–215. https://doi.org/10.1006/jpho.2001.0134

Price, G. (2005). *An introduction to French pronunciation*. Malden, MA: Blackwell Pub.

Prieto, P., & Roseano, P. (2018). Prosody: Stress, rhythm, and intonation. In K. L. Geeslin (Ed.), *The cambridge handbook of Spanish linguistics* (1st ed., pp. 211–236). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316779194.011

Prince, A., & Smolensky, P. (2004). *Optimality Theory: constraint interaction in generative grammar*. Malden, Mass. & Oxford: Blackwell.

Rau, D. V., Chang, H.-H. A., & Tarone, E. E. (2009). Think or sink: Chinese learners' acquisition of the English voiceless interdental fricative. *Language Learning*, *59*(3), 581–621. https://doi.org/10.1111/j.1467-9922.2009.00518.x

Reetz, H., & Jongman, A. (2011). *Phonetics, transcription, production, acoustics, and perception*. Wiley.

Richards, J. C., & Schmidt, R. W. (2010). *Dictionary of Language Teaching and Applied Linguistics (Fourth ed)*. Pearson Education, Harlow: Longman.

Roach, P.J. (1991). *Phonetic transcription.* Oxford International Encyclopedia of Linguistics.

Robinson, P. (2001a). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*, 27–57.

Robinson, P. (2001b). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In Robinson, P. (Ed.), *Cognition and second language instruction* (pp. 287–318). Cambridge: Cambridge University Press.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, *43*, 1–32.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching, 45*. 193–213. doi:10.1515/IRAL.2007.009

Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In Robinson, P. (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–38). Amsterdam: Benjamins.

Robinson, P. (2022). The cognition hypothesis and individual difference factors in SLA. In S. Li, P. Hiver, and M. Papi (Eds.), *The Routledge handbook of SLA and individual differences* (pp. 3–33). Chapter 1: Individual Differences in Second Language Acquisition: Theories, Research, and Pedagogy. New York: Routledge.

Rokita-Jaksow, J., & Ellis, M. (2019). *Early instructed second language acquisition pathways to competence*. Bristol, UK. Multilingual Matters.

Saito, K. (2015). The role of age of acquisition in late Second Language Oral proficiency attainment. *Studies in Second Language Acquisition*, *37*(4), 713–743. https://doi.org/10.1017/s0272263115000248

Saito, K. (2019). Individual differences in second language speech learning in classroom settings: Roles of awareness in the longitudinal development of Japanese learners' English /ɹ/ pronunciation. *Second Language Research, 35*, 149–172. https://doi.org/10.1177%2F0267658318768342

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, Rater, and instructional factors. *TESOL Quarterly*, *55*(3), 866–900. https://doi.org/10.1002/tesq.3027

Saito, K., & Brajot, F. (2013). Scrutinizing the role of length of residence and age of acquisition in the interlanguage pronunciation development of English /r/ by late Japanese bilinguals.

Bilingualism: *Language and Cognition, 16*, 847-863. https://doi.org/10.1017/S1366728912000703

Saito, K., & Lyster, R. (2011). Effects of Form-Focused Instruction and Corrective Feedback on L2 Pronunciation Development of /ɹ/ by Japanese Learners of English. *Language Learning*, *62*(2), 595–633. https://doi.org/10.1111/j.1467-9922.2011.00639.x

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652–708. https://doi.org/10.1111/lang.12345

Saito, Y., & Saito, K. (2017). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research, 21*, 589-608. https://doi.org/10.1177/1362168816643111

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics, 37*(2), 217–240. https://doi.org/10.1017/s0142716414000502

Saito, K., Trofimovich, P., Isaacs, T., & Webb, S. (2017b). Re-examining phonological and lexical correlates of second language comprehensibility: The role of rater experience. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 141–156). Multilingual Matters.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, *20*, 213–223. https://doi.org/10.1017/s0267190500200135

Schwartz, B. D. (2004). Why child L2 acquisition? In J. Van Kampen & S. Baauw (Eds.), *Proceedings of Generative Approaches to Language Acquisition 2003* (LOT Occasional Series). Utrecht: Utrecht University.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10*, 219–231. https://doi.org/ 10.1515/iral.1972.10.1-4.209

Sheldon, A., & Strange, W. (1982). The acquisition of / r / and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261. https://doi.org/10.1017/s0142716400001417

Shifflett, E. (2011). *Pitch, Frequency and Duration: Using Phonetic Correlates to Determine the Word-Level Stress Pattern of Tunisian Arabic* [Bachelor's thesis, College of William and Mary, Williamsburg].

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Singleton, D., & Muñoz, C. (2011). Around and beyond the critical period hypothesis. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 407–425). London, UK: Routledge.

Singleton, D., & Pfenninger, S. E. (2017). Reporting on politically sensitive issues: The case of telling the truth about early L2 instruction. In H. Rose & J. McKinley (Eds.), *Doing real research in applied linguistics* (pp. 214–224). London: Routledge. https://doi.org/10.4324/9781315389608-20

Singleton, D., & Pfenninger, S. E. (2018). L2 acquisition in childhood, adulthood and old age. *Journal of Second Language Studies*, *1*(2), 254–276. https://doi.org/10.1075/jsls.00003.sin

Singleton, D., & Pfenninger, S. E. (2019). Exploring a poorly understood variable: An agenda for classroom research on the age factor. *Language Teaching, 52*(1), 111–127. https://doi.org/10.1017/S0261444818000319

Singleton, D., & Pfenninger, S.E. (2022). Age. In S. Li, P. Hiver, M. Papi (Eds.), *The Routledge Handbook of Second Language Acquisition and Individual Differences* (pp. 251-266). New York and London: Routledge.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge, UK: Cambridge University Press.

Skoruppa, K., Pons, F., Bosch, L., Christophe, A., Cabrol, D., & Peperkamp, S. (2013). The development of word stress processing in French and Spanish infants. *Language Learning and Development*, *9*(1), 88–104. https://doi.org/10.1080/15475441.2012.693881

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*(2), 263–308. https://doi.org/10.1111/j.1467-9922.2010.00562.x

Stefanich, S., & Cabrelli, J. (2021). The effects of L1 English constraints on the acquisition of the L2 Spanish alveopalatal nasal. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.640354

Stölten, K. (2005). Effects of age of learning on VOT in voiceless stops produced by near-native L2 speakers. In A. Eriksson, & J. Lindh (Eds.), *Proceedings FONETIK 2005. The XVIIIth*

Swedish phonetics conference, May 25–27 2005 (pp. 91–94). Department of Linguistics, University of Gothenburg.

Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2014). Effects of age and speaking rate on voice onset time. *Studies in Second Language Acquisition, 37*(1), 71–100. https://doi.org/10.1017/s0272263114000151

Sweller, J. (1988) Cognitive load during problem solving: Effects on learning, *Cognitive Science*, 12, 257-285. Retrieved from https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8.

Suzukida, Y., & Saito, K. (2019). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Language Teaching Research, 25*(3), 431- 450. https://doi.org/10.1177/1362168819858246

Takagi, N. (1993). *Perception of American English /r/ and /l/ by adult Japanese learners of English: A unified view* [Doctoral dissertation, University of California, Irvine].

Tanner, M., & Landon, M. (2009). The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning and Technology, 13*, 51-65.

Thomson, R. (2018). Measurement of accentedness, intelligibility and comprehensibility. In O. Kang & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 11–29). London: Routledge.

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326–344. https://doi.org/10.1093/applin/amu076

Torras, M. R. & Celaya, M. L. (2001). Age-Related Differences in the Development of Written Production. An empirical study of EFL school learners. *International Journal of English Studies, 1*(2), 103–126.

Torras, M. R., Navés, T., Celaya, M. L., & Pérez-Vidal, C. (2006). Age and IL development in writing. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 156-182). Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853598937-009

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in L2 Spanish word processing: Implications for pronunciation instruction. *Modern Language Journal, 90*(4), 519–535.https://doi.org/10.1111/j.1540-4781.2006.00464.x

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition, 15*(4), 905-916. doi:10.1017/S1366728912000168

Trofimovich, P., Isaacs, T., Kennedy, S., Tsunemoto, A. (2022). Speech comprehensibility. In T. M. Derwing, M. J. Munro, R. Thomson (Eds.), *The Routledge handbook of second language acquisition and speaking* (pp. 174–187). Routledge. ISBN: 9780367900847.

Trofimovich, P., Kennedy, S., & Blanchet, J. (2015). Phonetics instruction in l2 French: contributions of segments, prosody and fluency to speech ratings. In J. Przedlacka, J. Maidment, and M. Ashby (Eds.), *Proceedings of the Phonetics Teaching and Learning Conference* (pp. 101-105). London: University College London.

Trofimovich, P., Nagle, C. L., O'Brien, M. G., Kennedy, S., Taylor Reid, K., & Strachan, L. (2020). Second language comprehensibility as a dynamic construct. *25 Years of Intelligibility, Comprehensibility and Accentedness*, *6*(3), 430–457. https://doi.org/10.1075/jslp.20003.tro

Tuniscop. (2018, March 6). *Hassen Zargouni : 50000 élèves sont inscrits dans des écoles primaires privées en Tunisie*. Tuniscope.com. https://www.tuniscope.com/ar/article/80430/actualites/tunisie/hassen-zargouni-340812

Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A sound approach to language matters - In honor of Ocke-Schwen Bohn* (pp. 607–630). Aarhus, Denmark: Dept. of English, School of Communication & Culture, Aarhus University. https://doi.org/10.7146/aul.322.218

Unsworth, S., Persson, L., Prins, T., & de Bot, K. (2015). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics, 36*, 527–548. https://doi.org/10.1093/applin/amt052

Vaissière, J. (1983). Language-independent prosodic features. In: Cutler, A. & Ladd, D. R. (Eds.) *Prosody: Models and measurements* (pp. 53–66). Berlin: Springer Verlag,

Vaissière, J. (2011). *La phonétique*. Paris, France : Presses universitaires de France.

van Boxtel, S, Bongaerts, T., & Coppen, P.A. (2005). Native-like attainment of dummy subjects in Dutch and the role of the L1. *IRAL - International Review of Applied Linguistics in Language Teaching*, *43*(4), 355–380. https://doi.org/10.1515/iral.2005.43.4.355

Verissimo, J., Heyer, V., Jacob, G., & Clahsen, H. (2018). Selective effects of age of acquisition on morphological priming: Evidence for a sensitive period. *Language Acquisition, 25*(3). 315–326. doi: 10.1080/10489223.2017.1346104

Walker, D. (2001). *French Sound Structure.* Calgary: University of Calgary Press.

Walter, H. (1976). *La dynamique des phonèmes dans le lexique français contemporain*. Gnenève: Librairie Droz.

White, J. L., & Turner, C. E. (2005). Comparing children's oral ability in two ESL programs. *The Canadian Modern Language Review, 61*, 491–517. doi:10.1353/cml.2005.0037

Whitman, R. L., & Jackson, K. L. (1972). The unpredictability of Contrastive Analysis. *Language Learning*, *22*(1), 29–41. https://doi.org/10.1111/j.1467-1770.1972.tb00071.x

Youssef, I (2013). *Place assimilation in Arabic: Contrasts, features, and constraints*. [Doctoral dissertation, University of Tromsø]. Retrieved from https://munin.uit.no/bitstream/handle/10037/5347/thesis.pdf?isAllowed=y&sequence=2

Zarobe, Y. R. D. (2005). Age and Third Language Production: A longitudinal study. International *Journal of Multilingualism, 2*(2), 105–112. https://doi.org/10.1080/14790710508668379

Zheng, Y., Cheng, L. (2018). How does anxiety influence language performance? From the perspectives of foreign language classroom anxiety and cognitive test anxiety. *Lang Test Asia, 8*(13). https://doi.org/10.1186/s40468-018-0065-4

Zielinski, B. (2015). The segmental/suprasegmental debate. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 397–410). Hoboken, NJ: John Wiley and Sons.

Zribi, I., Graja, M., Khmekhem, M. E., Jaoua, M., & Belguith, L. H. (2013). Orthographic transcription for spoken Tunisian Arabic. In *Computational linguistics and intelligent text processing* (pp. 153–163). Springer.

Appendices

Appendix A

Examples of transcriptions by the listeners

Example 2 of an original transcript by participant 14 from test 2

Le dimanche alain vient en velé pour voi le tortue et dit : lu la po torti et dire veut ti des raisons, de conture et une laitu. Le garçon intelligent pé pa pala des linettes et écoute les musuque va à école.

Note: the words in red are incorrectly pronounced words by the participant and were considered incorrect for listeners who transcribed them as heard.

Transcription of listeners:

→Listener 1

Le dimanche ala vient en vélo pour voi la torti et dire : lila torti, x x veut tu des raisins, de la confiture et une laitu. Le garçon intelligent qui pépa x des linettes et écoute le musique va à incole.

Number of total transcribed words: 42

Number of errors: 6

% of correctly transcribed words = $\frac{42-6}{42} \times 100 = 85{,}71\%$

→ Listener 2

Le dimanche alin vient on vélo pour voir le torture et dit : lila la torti, est dire veut tu des raisins, des conture et une laitu. Le garçon intelligent pai x x par des linettes et écoute la mesique et va à l'école.

Number of total transcribed words: 44

Number of errors: 5

% of correctly transcribed words = $\frac{44-5}{44} \times 100 = 88{,}63\%$

→ Listener 3

Le dimanche alain vient en velé po voi le tortue et dit lu la tortue et dire veux tu des raisons de contur et une laiti. Le garçon intelligent pé palle des lunettes et écoute lé musique va à acole.

Number of total transcribed words: 40

Number of errors: 5

% of correctly transcribed words = $\frac{40-5}{40}$ x 100 = 87,5%

- **Example 3 of an original transcript by participant 50 from test 2**

Le garçon interrigent qui porte des limalettes et acote la micique voi à la à l'école. et je dimange Alien von au velo pour vo la tor torti. Il dit : juli la la trou, juli la torti, a dinnor vé té vi té des rinciens, des cramelles et a o un laite.

Note: the words in red are incorrectly pronounced words by the participant and were considered incorrect for listeners who transcribed them as heard.

Transcription of listeners:

→ Listener 1

Le garçon intelligent qui porte des lunettes et écoute la musique va à l'école. X dimanche Alain vin au vélo pour avar la torti. Il dit : lili laX X X torti, au dinner vé tu X des rinciens, des caramelles et o une laitue.

Number of total transcribed words: 45

Number of errors: 11

% of correctly transcribed words = $\frac{45-11}{45}$ x 100 = 75,55%

→ Listener 2

Le garçon interregent qui porte des lilamettes X acote la misique va à l'école et jai dimanche unlien von an vélo pour vo la torti. Il dit : jili X la datrou, jili la tarti au diton vé tu X un ciens, des camelles et o un laiti.

Number of total transcribed words: 48

Number of errors: 5

% of correctly transcribed words = $\frac{48-15}{48}$ x 100 = 68,75%

→ Listener 3

Le garçon intelligent qui porte des limettes et acote la musique va à X l'école. X je dimange alien vin un velo poure vour la tortue il dit juli la la trou juli la torti X dinore vé ti X des raciens des cramelle é a o X laite.

Number of total transcribed words: 49

Number of errors: 16

% of correctly transcribed words = $\frac{49-16}{49}$ x 100 = 67,34%

- **Example 2 of an original transcript by participant 46 from test 4**

Dans dans le **vé** il y a des animaux. De lepin di lapon dit à la tortie, moi je suis puis **vie** que **po**. Ils font une course. Le papan pour très vitre et les ten et les tente marche lentement.

Transcription of listeners:

→ Listener 1

Dans X la foret il y a des animaux. Le lapin X X dit à la tortie, moi je suis plus vite que toi. Ils font une corse. Le lapin court très vite et la X X X torte marche lentement.75

Number of total transcribed words: 41

Number of errors: 23

% of correctly transcribed words = $\frac{41 - 23}{41}$ x 100 = 37,83%

→ Listener 2

Dans dans le vé il ya des animaux des animaux de lapon de lapon dire à la torture, moi je suis puis vet que te. Ils font une course. X Lepin pourt très vitre elle X X X X est tente marche lentement.53,48

Number of total transcribed words: 43

Number of errors: 23

% of correctly transcribed words = $\frac{43 - 23}{43}$ x 100 = 37,83%

→ Listener 3

dans X la veille il ya des animaux du lepins du lapons dit à la torti moi je suis pui vi di kepo ils font une course le papin poure très vite et lé X X X tante marche lentement.62,5

Number of total transcribed words: 40

Number of errors: 23

% of correctly transcribed words = $\frac{40 - 23}{40}$ x 100 = 37,83%

- **Example 3 of an original transcript by participant 35 from test 4**

Aujourd'hui le lapin se promène dans le jardin et il s'amuse beaucoup. Le torti dit tortue dit au lapin, moi je pour plus pi le le et tu es lent. Ils font un course, le lapin dit le lapin coure très rig.

Transcription of listeners:

→Listener 1

Aujourd'hui le lapin se promène dans le jardin et il s'amuse beaucoup. Le torti dit au X X X lapin, moi je pos X le pier X et tu es lent. Ils font un course, le lapin X X X coure très vite.

Number of total transcribed words: 43

Number of errors: 11

% of correctly transcribed words = $\frac{43 - 11}{43} \times 100 = 74{,}41\%$

→ Listener 2

Aujourd'hui le lapin se promène dans le jardin et il se mise beaucoup. Le tortu dit tortie dit au lapin, moi je pour X le pier le le et le tu es lent. Ils font un course, le lapin dit le lapin coure très vit.

Number of total transcribed words: 45

Number of errors: 7

% of correctly transcribed words = $\frac{45 - 7}{45} \times 100 = 84{,}44\%$

→ Listener 3

aujourd'hui le lapin se promène dans le jardin et il s'amuse beaucoup le tortu dit tortue dit au lapin moi je por pli pi le X et tu est lent ils font un course le lapin dit le lapin court très ri.

Number of total transcribed words: 42

Number of errors: 5

% of correctly transcribed words = $\frac{42 - 5}{40} \times 100 = 88{,}09\%$

Appendix B

Questionnaire

Monsieur/ Madame,

Comme convenu, nous vous adressons le questionnaire concernant notre étude intitulée 'Effects of age and exposure on the acquisition of pronunciation: Focus on Arabic learners of French' portant sur l'effet de l'âge et l'exposition sur l'acquisition de la prononciation du français langue seconde.

Le questionnaire sert à obtenir des informations générales sur l'exposition de votre enfant au français à l'extérieur de la classe. Les réponses données resteront confidentielles et ne serviront qu'à des fins de recherche. Le nom de votre enfant ne sera pas utilisé, car nous utiliserons un code pour permettre de classifier les résultats.

Nous vous prions de bien vouloir répondre à notre questionnaire et nous vous remercions d'avance pour votre temps alloué. Nous tenons à vous assurer que les réponses fournies par vous seront utilisées à des fins scientifiques, et que la confidentialité des réponses fournies sera assurée.

Pour plus de détails et d'informations, veuillez me contacter par téléphone ou par email.

INFORMATION GENERALE

1. Votre nom : ______________________________
   Le prénom de votre enfant : ______________________________
2. Votre lien de parenté avec l'enfant : __________________
3. Quelle est votre langue maternelle ? (La langue maternelle est la langue que vous avez apprise dès la naissance)
   □ Arabe
   □ Français
   □ Arabe et français
   □ Autre : ______________________________

4. Quelle est la langue maternelle de votre conjoint(e) ?
   □ Arabe
   □ Français
   □ Arabe et français
   □ Autre : ______________________________

5. Quelle(s) langue(s) parlez- vous avec votre enfant à la maison ?

□ Arabe
□ Français
□ Arabe et français
□ Autre : ________________________________

6. Quelle(s) langue(s) votre conjoint parle-t-il avec votre enfant à la maison ?
   □ Arabe
   □ Français
   □ Arabe et français
   □ Autre : ________________________________

8. Quel est votre pays de naissance?

   Ville : _______________________________

   Pays : _______________________________

9. Quel est le pays de naissance de votre conjoint(e) ?

   Ville : _______________________________

   Pays : _______________________________

10. À quel âge votre enfant a-t-il commencé l'apprentissage du français ?
    □ Dès la maternelle (4 – 5 ans)
    □ Dès la 1re année du primaire (6-7 ans)
    □ Dès la 3e année du primaire (8-9)

11. Votre enfant a-t-il des frères ou des sœurs?
    □ Oui
    □ Non

12. Est-ce qu'il y a une langue parlée autre que l'arabe entre les enfants dans la famille?

    □ Oui
    □ Non

    Si oui, quel est cette langue? _______________________________

13. Visitez-vous souvent avec vos enfants des pays francophones ?
    □ Oui
    □ Non

    Si oui, combien de jours, semaines par an ? Quels pays ?
    ____

14. Sur une échelle de 0 à 10, quelle importance accordez-vous à l'acquisition d'une bonne prononciation du français : (0 = aucune importance ; 10 = essentielle)

    0 1 2 3 4 5 6 7 8 9 10

15. Sur une échelle de 0 à 10, quelle importance accordez-vous à l'acquisition d'une prononciation la plus proche possible de celle des francophones (0 = aucune importance ; 10 = essentiel)

    0 1 2 3 4 5 6 7 8 9 10

EXPOSITION DE L'ENFANTAU FRANÇAIS À l'ÉXTERIEUR DE LA CLASSE

1- Écoute de la musique francophone
Jamais □ Quelques fois par mois □ Quelques fois par semaine □ Tous les jours □

2- Regarde des films et des émissions télévisées en français.

Jamais □ Quelques fois par mois □ Quelques fois par semaine □ Tous les jours □

3- Lit des comptines en français

Jamais □ Quelques fois par mois □ Quelques fois par semaine □ Tous les jours □

4- Reçoit des leçons particulières en français à l'extérieur de la classe.
- □ Oui
- □ Non

Si oui, combien d'heures par semaine : ___________________

6- Est-ce qu'il y a des membres de votre famille qui utilisent la langue fréquemment avec votre enfant (grands parents par exemple)?
- □ Oui
- □ Non

MERCI POUR VOTRE COLLABORATION

Appendix C

Certification of ethical acceptability for research involving human subjects

CERTIFICATION OF ETHICAL ACCEPTABILITY
FOR RESEARCH INVOLVING HUMAN SUBJECTS

Name of Applicant:	Amina Affes
Department:	N/A\Individualized program
Agency:	N/A
Title of Project:	Effects of age and language exposure on the acquisition of pronunciation: Focus on primary school Arabic learners of French
Certification Number:	30015530

Valid From: October 19, 2021 To: October 18, 2022

The members of the University Human Research Ethics Committee have examined the application for a grant to support the above-named project, and consider the experimental procedures, as outlined by the applicant, to be acceptable on ethical grounds for research involving human subjects.

Dr. Richard DeMont, Chair, University Human Research Ethics Committee

Appendix D

Consent Form (School administration)

DEMANDE D'AUTORISATION DE L'ADMINISTRATION DE L'ÉCOLE

Bonjour

Je suis étudiante en cotutelle à Concordia University, Montreal. En ce moment, je prépare une étude dans le cadre de mon doctorat en linguistique en cotutelle, sous la direction de Dr. Adel Jebali.

Avant d'accepter à ce projet de recherche d'être mené dans votre école, veuillez prendre le temps de lire les renseignements qui suivent. Ce document vous explique le but de ce projet de recherche, ses procédures et ses avantages. Nous vous invitons à nous poser toutes les questions que vous jugerez utiles.

Mon projet de recherche intitulé « Effects of age and exposure on the acquisition of pronunciation: Focus on Arabic learners of French » vise à comparer la maîtrise de la prononciation acquise au cours du 1er cycle du primaire de trois groupes d'élèves qui ont commencé l'apprentissage du français à des âges différents et ont été exposés à la langue cible pour des durées variables.

J'ai besoin de faire mon étude dans votre école car vos élèves possèdent toutes les caractéristiques requises pour mon étude. Nous vous demanderions de bien vouloir accepter que vos élèves passent une seule entrevue individuelle (avec enregistrement audio uniquement) d'environ 15 minutes. Pendant cette entrevue, nous allons présenter à l'élève quelques images individuelles qu'il va identifier, une série d'images qui constitue une histoire qu'il va raconter et une série de mots et phrases qu'il va lire.

Le fait que les élèves de votre école participent à cette recherche offre une occasion de participer dans l'avancement des recherches dans le domaine de la linguistique appliquée.

Je tiens à vous assurer que les réponses fournies par les élèves seront utilisées à des fins scientifiques, et que la confidentialité des réponses et données fournies sera assurée.

Les mesures suivantes seront appliquées pour assurer la confidentialité des renseignements fournis par les participants :

- Les noms des participants ne paraîtront dans aucun rapport ;
- Les divers documents de la recherche seront codifiés et seul le chercheur aura accès à la liste des noms et des codes ;
- Les résultats individuels des participants ne seront jamais communiqués ;
- Les matériaux de la recherche, incluant les données et les enregistrements, seront conservés (sur ordinateur protégés par un mot de passe). Ils seront détruits trois ans après la fin de la recherche, soit en décembre 2024.
- La recherche fera l'objet de publications dans des revues scientifiques, et aucun participant ne pourra y être identifié.

Les résultats de l'étude vous seront communiqués après la soutenance de thèse.

Nous vous remercions à l'avance de votre précieuse collaboration et nous vous prions d'agréer l'assurance de nos sentiments les meilleurs.

Toute question concernant le projet pourra être adressée à :

- Amina Affes : amina.affes@gmail.com / amina.affes@mail.concordia.ca
  Téléphone: (+216) 27 403 717 / (+1) (438) 969-8747

- Dr. Adel Jebali: adel.jebali@concordia.ca;
  Téléphone: (514) 848-2424 ext. 7873;
  Bureau: S-LB 620-03 J.W. McConnell Building,1400 De Maisonneuve Blvd. W, Montréal, Canada.

CONSENTEMENT DE L'ETABLISSEMENT

Je reconnais par la présente avoir lu et compris le présent document. J'ai eu l'occasion de poser des questions et d'obtenir des réponses. Je consens que mon établissement ____ participe à l'étude dans les conditions décrites ci-dessus.

NOM (en majuscules) ____

SIGNATURE ____

DATE ____

- Mesures préventives exigées par l'administration de l'établissement liées à la COVID-19

____
____
____
____
____
____
____
____

Si vous avez des questions sur l'aspect scientifique ou savant de cette étude, communiquez avec le chercheur. Vous trouverez ses cordonnées sur la première page. Vous pouvez aussi communiquer avec son professeur-superviseur.

Pour toute préoccupation d'ordre éthique relative à ce projet de recherche, veuillez communiquer avec le responsable de l'éthique de la recherche de l'Université Concordia au 514-848-2424, poste 7481, ou à oor.ethics@concordia.ca.

Appendix E

Invitation letters

Invitation de participation à une étude

Madame, Monsieur,

Je m'appelle Amina Affes, je suis une étudiante au doctorat à l'Université Concordia, Montréal Canada et à la Faculté des Lettres et sciences humaines Sfax. Je mène actuellement une étude portant sur l'effet de l'âge sur l'acquisition de la prononciation du Français langue seconde.

Afin de réaliser cette étude, nous sommes en train d'identifier des enfants qui présentent les critères suivants :

- L'âge de début de l'apprentissage de la langue française soit aux alentours de 4 ans.
- La langue maternelle de tous les participants ainsi que celle de leurs parents est l'arabe.

Si votre enfant possède les critères mentionnés ci-dessus, nous vous demanderions de bien vouloir accepter qu'il participe à notre étude qui consiste à une seule entrevue individuelle (avec enregistrement audio uniquement) d'environ 15 minutes. Pendant cette entrevue, nous allons présenter à votre enfant quelques images individuels qu'il va identifier et une série d'images qui constitue une histoire qu'il va raconter. L'entrevue sera menée dans l'école de votre enfant pendant l'heure du déjeuner.

Nous vous informons que nous avons obtenu l'autorisation de la direction de l'école pour mener notre étude. Si vous donneriez votre consentement préliminaire, nous vous contacterons par téléphone pour vous expliquer le formulaire de consentement, répondre à vos questions et pour vous inviter à remplir un questionnaire concernant l'exposition de votre enfant à la langue française en dehors de la salle de classe. Ces documents vous seront communiqués par email ultérieurement.

Le fait que votre enfant participe à cette recherche lui offre une occasion qui lui permet de participer à l'avancement des recherches dans le domaine de la linguistique. Si vous le souhaiter, les résultats de l'étude vous seront communiqués après la soutenance de thèse.

Il va sans dire que les renseignements ainsi recueillis ultérieurement seront utilisés uniquement pour les besoins de la recherche et demeureront tout à fait confidentiels.

Nous vous remercions à l'avance de votre précieuse collaboration et nous vous prions d'agréer l'assurance de nos sentiments les meilleurs.

J'accepte ☐

Je n'accepte pas ☐

دعوة لمشاركة طفلكم في بحث علمي

سيدتي، سيدي

اسمي أمينة الفعاس ، طالبة دكتوراه في جامعة كونكورديا ، مونتريال كندا في كلية الآداب والعلوم الإنسانية قسم علم النفس. حاليًا أقوم بإجراء دراسة حول تأثير العمر على اكتساب مهارة النطق في اللغة الفرنسية عند الأطفال الذين لهم المعايير التالية

سن بداية تعلم اللغة الفرنسية حوالي 4 سنوات

العربية هي اللغة الام للأطفال وللأولياء

إذا كان طفلكم يستوفي المعايير المذكورة أعلاه، فإننا نرجو من جنابكم الموافقة على مشاركة ابنكم في دراستنا التي تتكون من مقابلة فردية واحدة (مع تسجيل صوتي فقط) لمدة 15 دقيقة تقريبًا. خلال هذه المقابلة سيقدم لطفلكم بعض الصور الفردية, سيروي قصة عن طريق سلسلة من الصور وسيقرأ بعض الكلمات و جمل. ستجرى المقابلة في المدرسة خلال ساعة الغداء

كما نعلمكم أننا حصلنا على إذن من إدارة المدرسة لإجراء دراستنا. إذا منحتمونا موافقتكم المبدئية فإننا سنرسل إليكم وثيقة استبيان ووثيقة موافقة للولي. سيتم إرسال هذه المستندات إليكم من قبل إدارة المدرسة لاحقًا

مشاركة طفلكم في هذا البحث يوفر له فرصة المساهمة في تقدم الأبحاث في مجال اللغويات

إذا كنتم ترغبون، فسيتم إبلاغكم بنتائج الدراسة بعد مناقشة الأطروحة .

نلفت انتباهكم أن المعلومات التي سيتم جمعها ستستخدم فقط لأغراض البحث وستبقى سرية تمامًا

نشكركم مسبقاً على تعاونكم الثمين ونرجوا منكم قبول أحر تحياتنا

□ أوافق

□ لا أوافق

Appendix F

Consent Form (parents of participants)

CONSENTEMENT ÉCLAIRÉ À LA PARTICIPATION À UNE ÉTUDE

Remarque : Le masculin est utilisé pour faciliter la lecture.

Titre de l'étude: Effects of age and exposure on the acquisition of pronunciation: Focus on Arabic learners of French

Chercheur: Amina Affes (PhD student in Individualized program)

Coordonnées du chercheur : amina.affes@mail.concordia.ca

Professeur-superviseur : Dr. Adel Jebali Départment d'études françaises

Coordonnées du professeur-superviseur :
Dr. Adel Jebali adel.jebali@concordia.ca; (514) 848-2424 ext. 7873; office S-LB639-05, J.W. McConnell Building,1400 De Maisonneuve Blvd. W.

Sources de financement de l'étude : -

Nous invitons votre enfant à prendre part au projet de recherche susmentionné. Le présent document vous renseigne sur les conditions de participation à l'étude; veuillez le lire attentivement. Au besoin, n'hésitez pas à communiquer avec le chercheur pour obtenir des précisions.

A. BUT DE LA RECHERCHE

Cette étude a pour but d'examiner les effets de l'âge et de l'exposition à la langue sur l'acquisition de la prononciation du français.

B. PROCÉDURES DE RECHERCHE

Si votre enfant participait à l'étude, il devrait lire une liste de mots, lire des phrases nommer des objets à partir de quelques photos (l'image d'une pomme par exemple) et raconter une histoire à partir d'une série d'images en français.

Somme toute, sa participation s'étendra sur une durée de 15 minutes.

L'entretien sera effectué individuellement au sein de l'école de votre enfant.

C. RISQUES ET AVANTAGES

Il n'y a aucun risque associé à cette étude.
Le fait que votre enfant participe à cette recherche offre une occasion de participer dans l'avancement des recherches dans le domaine de la linguistique appliquée.

D. CONFIDENTIALITÉ

Dans le cadre de cette étude, nous recueillerons les renseignements suivants :

Le nom et prénom de votre enfant et le lieu de sa naissance.

En tant que parent du participant, vous permettez aux chercheurs d'avoir accès à des renseignements sur l'exposition extracurriculaire à la langue française. Ceux-ci seront obtenus à partir d'un questionnaire que vous remplissez.

Excepté les situations précisées aux présentes, seules les personnes qui mènent cette recherche auront accès aux renseignements fournis. Nous n'utiliserons l'information qu'aux fins de l'étude décrite dans ce document.

Les renseignements recueillis resteront confidentiels. On ne pourra donc établir aucun lien entre votre identité ou l'identité de votre enfant et l'information que vous fournissez.

Les renseignements recueillis seront codés; ils seront donc associés à un code. Le chercheur aura une liste établissant le lien entre le code et votre nom sur un ordinateur protégé par une mot de passe.

Nous avons l'intention de publier les résultats de cette étude. Cependant, on ne pourra pas identifier votre enfant dans la publication.

Nous détruirons les données trois ans après la fin de l'étude.

F. CONDITIONS DE PARTICIPATION

Vous pouvez refuser de ne pas inclure votre enfant dans la recherche ou vous en retirer à n'importe quel moment. Vous pouvez aussi demander que les données que votre enfant a fournies ne soit pas utilisée; le cas échéant, votre choix sera respecté. Si vous prenez une décision en ce sens, vous devrez en avertir le chercheur au plus tard le 20 décembre 2021.

L'indemnité compensatoire pour la participation de votre enfant à la recherche sera d'une broche smiley. Si votre enfant se retire avant la fin du projet, il recevra la broche.

Vous ne subirez aucune conséquence négative si vous décidez de ne pas inclure votre enfant dans l'étude, d'interrompre sa participation à celle-ci ou de nous demander de ne pas utiliser ses données.

G. CONSENTEMENT DU PARENT DU PARTICIPANT

Je reconnais par la présente avoir lu et compris le présent document. J'ai eu l'occasion de poser des questions et d'obtenir des réponses. Je consens à participer à l'étude dans les conditions décrites ci-dessus.

NOM DE VOTRE ENFANT (en majuscules)

____

VOTRE NOM (en majuscules)

____

SIGNATURE ____

DATE ____

Si vous avez des questions sur l'aspect scientifique ou savant de cette étude, communiquez avec le chercheur. Vous trouverez ses cordonnées sur la première page. Vous pouvez aussi communiquer avec son professeur-superviseur.

Pour toute préoccupation d'ordre éthique relative à ce projet de recherche, veuillez communiquer avec le responsable de l'éthique de la recherche de l'Université Concordia au 514-848-2424, poste 7481, ou à oor.ethics@concordia.ca.

Appendix G

Consent Form (Participants)

CONSENTEMENT ÉCLAIRÉ À LA PARTICIPATION À UNE ÉTUDE

Titre de l'étude: Effects of age and exposure on the acquisition of pronunciation: Focus on Arabic learners of French

Chercheure: Amina Affes (étudiante PhD)

Coordonnées du chercheur : amina.affes@mail.concordia.ca

Le Professeur-superviseur : Dr. Adel Jebali Départment d'études françaises

Coordonnées du professeur-superviseur :
Dr. Adel Jebali adel.jebali@concordia.ca; (514) 848-2424 ext. 7873; office S-LB639-05, J.W. McConnell Building,1400 De Maisonneuve Blvd. W.

A. BUT DE LA RECHERCHE

Cette étude a pour but d'examiner les effets de l'âge et de l'exposition à la langue sur l'acquisition de la prononciation du français.

B. PROCÉDURES DE RECHERCHE

Vous avez 4 taches à faire :

- Lire une liste de mots
- Lire 4 phrases
- Identifier des images
- Raconter une histoire

Somme toute, sa participation s'étendra sur une durée de 15 minutes. L'entretien sera effectué individuellement au sein de votre l'école. Seulement votre voix sera enregistrée.

D. CONFIDENTIALITÉ

Dans le cadre de cette étude, nous recueillerons les renseignements suivants :
Votre nom et prénom.
Les renseignements recueillis resteront confidentiels. Seulement moi et mon directeur de recherche auront accès à vos renseignements. Personne ne peut vous identifier.

F. CONDITIONS DE PARTICIPATION

Vous pouvez refuser de ne pas participer à la recherche ou vous en retirer à n'importe quel moment.
L'indemnité compensatoire pour votre participation sera d'une broche smiley. Si vous vous retirez avant la fin du projet, vous recevrez la broche.
Vous ne subirez aucune conséquence négative si vous décidez de ne pas participer dans l'étude, d'interrompre votre participation à celle-ci.

G. CONSENTEMENT DU PARTICIPANT

Je reconnais par la présente avoir lu et compris le présent document. J'ai eu l'occasion de poser des questions et d'obtenir des réponses. Je consens à participer à l'étude dans les conditions décrites ci-dessus.

NOM (en majuscules)
____

SIGNATURE ____

DATE ____

Si vous avez des questions sur l'aspect scientifique ou savant de cette étude, communiquez avec le chercheur. Vous trouverez ses cordonnées sur la première page. Vous pouvez aussi communiquer avec son professeur-superviseur.

Pour toute préoccupation d'ordre éthique relative à ce projet de recherche, veuillez communiquer avec le responsable de l'éthique de la recherche de l'Université Concordia au 514-848-2424, poste 7481, ou à oor.ethics@concordia.ca.

<u>دراسة في مشاركة على موافقة استمارة</u>

عنوان الدراسة:
Effects of age and exposure on the acquisition of pronunciation: Focus on Arabic learners of French

الباحث: أمينة العفاس (طالبة دكتوراه)
تفاصيل الاتصال: amina.affes@mail.concordia.ca

الأستاذ المشرف: الدكتور عادل الجبالي
تفاصيل الاتصال بالأستاذ المشرف:

Dr. Adel Jebali adel.jebali@concordia.ca; (514) 848-2424 ext. 7873; office S-LB639-05, J.W. McConnell Building,1400 De Maisonneuve Blvd. W.

1. الغرض من البحث
الغرض من هذه الدراسة هو دراسة تأثير العمر والتعرض للغة على اكتساب النطق الفرنسي.

2. إجراءات البحث
لديك 4 مهام للقيام بها:
- قراءة قائمة كلمات
- قراءة 4 جمل
- التعرف على الصور
- رواية قصة بالفرنسية

بشكل عام، مدة المشاركة هي 15 دقيقة. سيتم إجراء المقابلة بشكل فردي داخل مدرستك وسيتم تسجيل صوتك فقط.

3. السرية
كجزء من هذه الدراسة، سنقوم بجمع المعلومات التالية:
– اسمك ولقبك
ستبقى المعلومات التي يتم جمعها سرية. إن الباحث والمشرف فقط سيتاح لنا الوصول لمعلوماتك. لا أحد يستطيع التعرف عليك

4. شروط المشاركة
1. يجوز لك رفض المشاركة في البحث أو الانسحاب منه في أي وقت.
2. مقابل مشاركتك سوف تتلقى بونس. إذا قررت الانسحاب قبل نهاية الدراسة، سوف تتلقى البونس.
3. لن تكون هناك أي عواقب سلبية إذا قررت عدم المشاركة أو الانسحاب في الدراسة.

5. الموافقة

أعترف بموجب هذا بأنني قرأت هذه الوثيقة وفهمتها. أتيحت لي الفرصة لطرح الأسئلة والحصول على إجابات.
أوافق على المشاركة في الدراسة وفقا للشروط المذكورة أعلاه.

الاسم واللقب ________________________________

التوقيع ________________________________

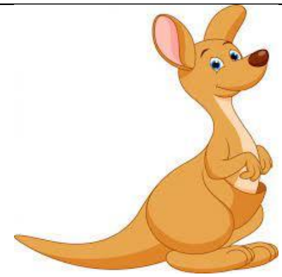
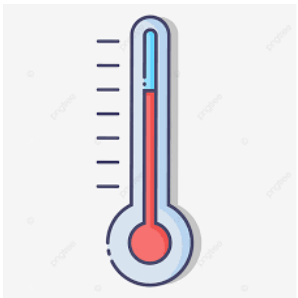
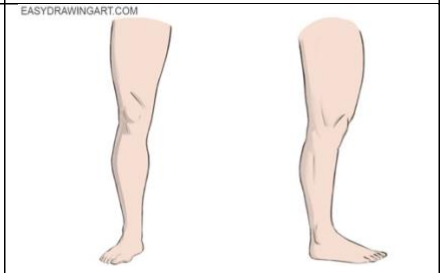
التاريخ ________________________________

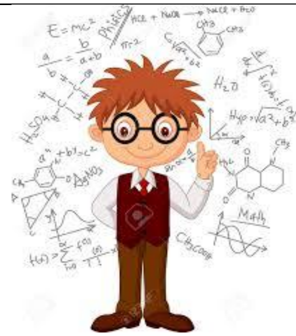
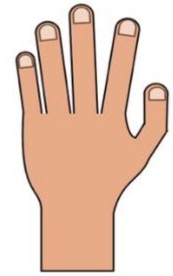
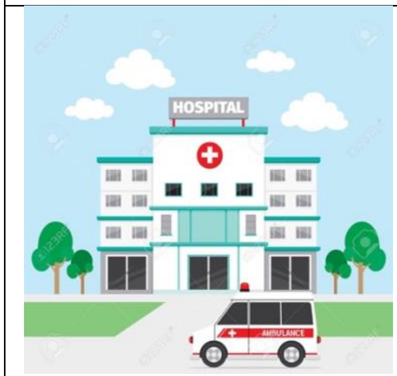
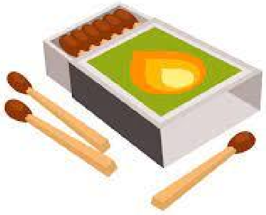
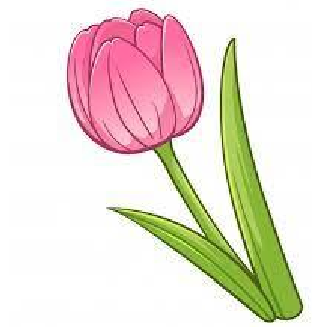
إذا كان لديك أي أسئلة حول الجانب العلمي أو العملي لهذه الدراسة، اتصل بالباحث. ستجد تفاصيل الاتصال به في الصفحة الأولى. يمكنك أيضا الاتصال بالأستاذ المشرف.

لأي مخاوف أخلاقية تتعلق بهذا المشروع البحثي، يرجى الاتصال بمسؤول أخلاقيات البحوث بجامعة كونكورديا على 514-848-2424 ext. 7481 أو oor.ethics@concordia.ca.

Appendix H
Task 1 : Word list reading

Instruction: Veuillez lire les mots suivants

 **Ampoule**	 **Pamplemousse**	 **Kangourou**	 **Tambour**
 **Temperature**	 **Ange**	 **BLANC**	 **Jambe**
 **Serpent**	 **Argent**	 **Dindon**	 **Interdit**

 Peinture	 Intelligent	 Important	 Main
 Raisin	 Faim	 Poussin	 Magicien
 Urgence	 Uniforme	 Allumette	 Tulipe

 Jumeaux	 Laitue	 Voiture	 Chaussure
 Légumes	 Cube	 Échelle	 Élèves
 Étoile	 Zéro	 Été	 Blé
 Idée	 Pied	 Dîner	 fée

Appendix I

Task 2: Sentence reading

Instruction: Veuillez lire les phrases suivantes

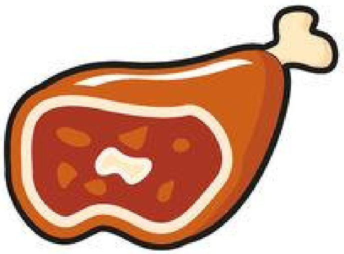
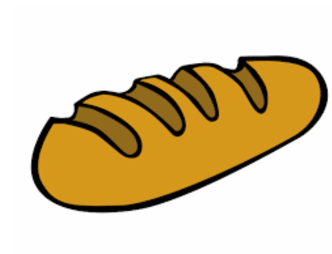
- Dans le jardin de Martin se trouve des dindons, un poussin qui mange de la confiture, un lapin qui boit du lait et un chien qui joue avec sa poupée.
- Le dimanche, Alain vient en vélo pour voir la tortue. Il dit : Lulu la tortue, au dîner, veux-tu des raisins, des caramels ou une laitue?
- Le dentiste avec l’infirmière au pantalon blanc donne un médicament aux enfants.
- Le garçon qui porte des lunettes et écoute la musique va à l’école.

Appendix J

Task 3 : Picture identification

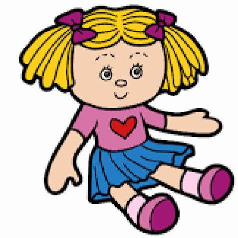
Instruction: Veuillez identifier les images suivantes

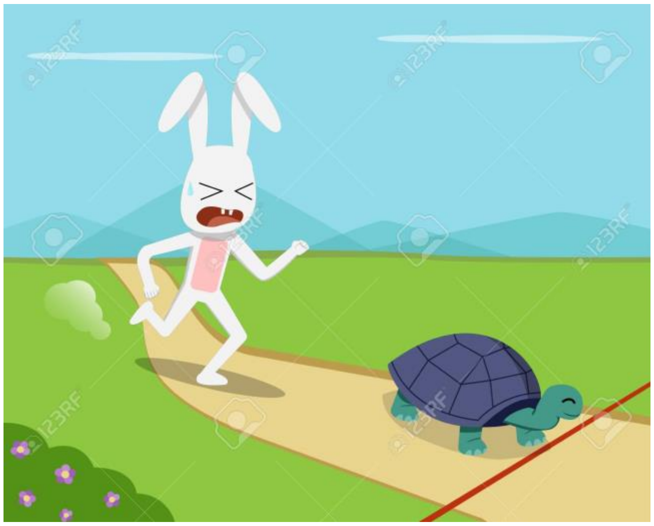
			
			
			

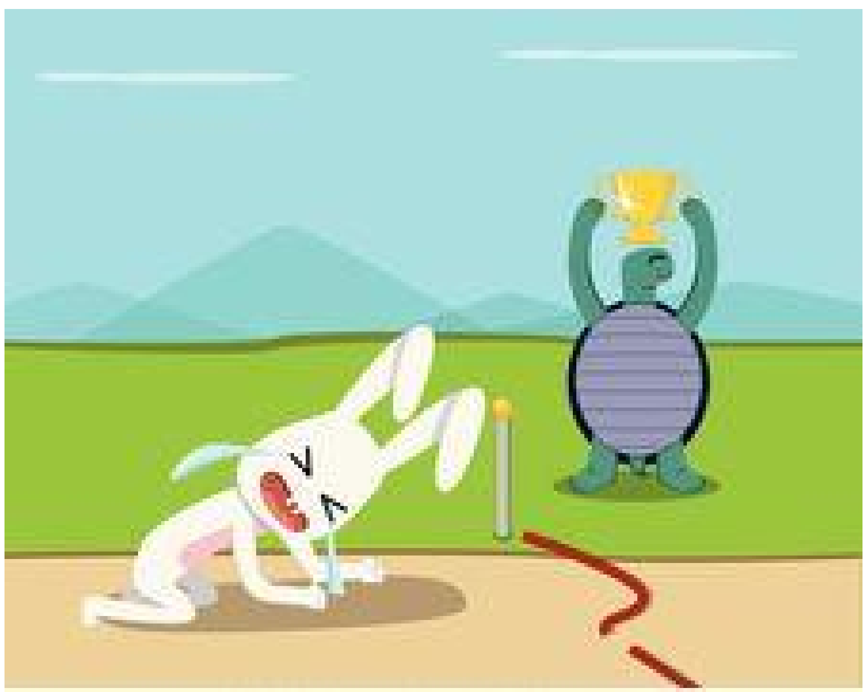
			
			
			


SCHOOL
MATH
alamy stock photo



Appendix K
Task 4: Storytelling





Appendix L

Tests of normality

Tests of Normality for the Mean Score of Segmental Pronunciation Accuracy in Task 1

Tests of Normality							
	Age Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean of correct pronunciation of the [ã] sound	Group1/4-12/1600	.222	23	.005	.857	23	.004
	Group2/8-12/1000	.152	24	.160	.908	24	.032
	Group3/4-8/1000	.170	21	.116	.950	21	.337
Mean of correct pronunciation of the [y] sound	Group1/4-12/1600	.250	23	<.001	.787	23	<.001
	Group2/8-12/1000	.176	24	.052	.894	24	.016
	Group3/4-8/1000	.194	21	.038	.898	21	.033
Mean of correct pronunciation of the [e] sound	Group1/4-12/1600	.241	23	.001	.852	23	.003
	Group2/8-12/1000	.208	24	.009	.915	24	.046
	Group3/4-8/1000	.226	21	.007	.886	21	.019
Mean of correct pronunciation of the [ɛ̃] sound	Group1/4-12/1600	.357	23	<.001	.706	23	<.001
	Group2/8-12/1000	.132	24	.200*	.943	24	.186
	Group3/4-8/1000	.198	21	.031	.860	21	.006
Mean of four sounds	Group1/4-12/1600	.159	23	.135	.939	23	.167
	Group2/8-12/1000	.118	24	.200*	.951	24	.288
	Group3/4-8/1000	.147	21	.200*	.970	21	.727

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Tests of Normality for the Mean Score of Correct Stress Placement in Task 1

	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean score of correct stress placement for [ɑ̃]	Group1	.265	23	<.001	.863	23	.005
	Group2	.274	24	<.001	.859	24	.003
	Group3	.156	21	.200*	.919	21	.082
Mean score of correct stress placement for [ɛ̃]	Group1	.254	23	<.001	.882	23	.011
	Group2	.221	24	.004	.884	24	.010
	Group3	.181	21	.070	.941	21	.225
Mean score of correct stress placement for [y]	Group1	.222	23	.004	.809	23	<.001
	Group2	.293	24	<.001	.785	24	<.001
	Group3	.236	21	.003	.874	21	.011
Mean score of correct stress placement for [e]	Group1	.340	23	<.001	.664	23	<.001
	Group2	.350	24	<.001	.733	24	<.001
	Group3	.205	21	.021	.850	21	.004
Mean score of correct stress placement of four sounds	Group1	.097	23	.200*	.970	23	.691
	Group2	.141	24	.200*	.939	24	.157
	Group3	.102	21	.200*	.981	21	.934

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Tests of Normality for the Mean Score of Segmental Pronunciation Accuracy in Task 2

	Age Group	Kolmogorov-Smirnova			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean score of correct pronunciation of [ɑ̃]	Group1	.532	23	<.001	.324	23	<.001
	Group2	.222	24	.003	.860	24	.003
	Group3	.297	21	<.001	.767	21	<.001
Mean score of correct pronunciation of [ɛ̃]	Group1	.371	23	<.001	.702	23	<.001
	Group2	.164	24	.096	.903	24	.025
	Group3	.248	21	.002	.803	21	<.001
Mean score of correct pronunciation of [y]	Group1	.499	23	<.001	.463	23	<.001
	Group2	.142	24	.200*	.931	24	.105
	Group3	.193	21	.039	.889	21	.021
Mean of correct pronunciation of [e]	Group1	.517	23	<.001	.402	23	<.001
	Group2	.162	24	.102	.874	24	.006
	Group3	.322	21	<.001	.753	21	<.001
Mean score of four sounds	Group1	.258	23	<.001	.761	23	<.001
	Group2	.171	24	.069	.913	24	.041
	Group3	.148	21	.200*	.930	21	.140

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Tests of normality for correct stress placement in task 2

	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean score of correct stress placement for [ɑ̃]	Group1	.236	23	.002	.837	23	.002
	Group2	.201	24	.014	.883	24	.010
	Group3	.268	21	<.001	.791	21	<.001
Mean score of correct stress placement for [ɛ̃]	Group1	.279	23	<.001	.792	23	<.001
	Group2	.268	24	<.001	.847	24	.002
	Group3	.315	21	<.001	.840	21	.003
Mean score of correct stress placement for [y]	Group1	.300	23	<.001	.850	23	.003
	Group2	.201	24	.014	.883	24	.010
	Group3	.156	21	.196	.949	21	**.328**
Mean score of correct stress placement for [e]	Group1	.262	23	<.001	.855	23	.003
	Group2	.232	24	.002	.887	24	.012
	Group3	.325	21	<.001	.816	21	.001
Mean score of correct stress placement of four sounds	Group1	.168	23	.091	.926	23	.089
	Group2	.106	24	.200*	.978	24	.850
	Group3	.149	21	.200*	.879	21	.014
Mean score of correct stress placement all words	Group1	.193	23	.027	.905	23	.033
	Group2	.115	24	.200*	.974	24	**.775**
	Group3	.163	21	.150	.884	21	.018

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Tests of Normality for the Mean Score of Comprehensibility in Task 2

	Age Group	Kolmogorov-Smirnova			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean score of comprehensibility	Group1	.253	23	<.001	.732	23	<.001
	Group2	.206	24	.010	.889	24	.013
	Group3	.138	21	.200*	.904	21	.041

Tests of Normality for the Mean Score of Accentedness in Task 2

	Age Group	Kolmogorov-Smirnova			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean score of accentedness	Group1	.164	23	.110	.862	23	.005
	Group2	.201	24	.013	.899	24	.021
	Group3	.183	21	.063	.901	21	.037

Tests of Normality for the Mean Score of Intelligibility in Task 2

	Age Group	Kolmogorov-Smirnova			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean score of intelligibility	Group1	.335	23	< .001	.753	23	< .001
	Group2	.295	24	< .001	.740	24	< .001
	Group3	.196	21	.035	.897	21	.031

Tests Of Normality for the Mean Score of Segmental Pronunciation Accuracy in Task 3

	Age Group	Kolmogorov-Smirnova			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean score of correct pronunciation of [ã]	Group1	.459	23	<.001	.551	23	<.001
	Group2	.321	24	<.001	.665	24	<.001
	Group3	.176	21	.087	.887	21	.020
Mean score of correct pronunciation of [ɛ̃]	Group1	.256	23	<.001	.790	23	<.001
	Group2	.232	24	.002	.905	24	.028
	Group3	.351	21	<.001	.687	21	<.001
Mean score of correct pronunciation of [y]	Group1	.316	23	<.001	.776	23	<.001
	Group2	.169	24	.076	.914	24	.042
	Group3	.186	21	.057	.916	21	.074
Mean score of correct pronunciation of [e]	Group1	.377	23	<.001	.649	23	<.001
	Group2	.159	24	.119	.919	24	.055
	Group3	.238	21	.003	.909	21	.052
Mean score of correct pronunciation of four sounds	Group1	.172	23	.076	.898	23	.024
	Group2	.162	24	.106	.942	24	.183
	Group3	.202	21	.025	.914	21	.066

a. Lilliefors Significance Correction

Tests of Normality for Correct Stress Placement in Task 3

	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean score of correct stress placement for [ɑ̃]	Group1	.290	23	<.001	.861	23	.004
	Group2	.225	24	.003	.922	24	.064
	Group3	.219	21	.010	.890	21	.022
Mean score of correct stress placement for [ɛ̃]	Group1	.217	23	.007	.917	23	.059
	Group2	.258	24	<.001	.889	24	.013
	Group3	.208	21	.018	.863	21	.007
Mean score of correct stress placement for [y]	Group1	.271	23	<.001	.871	23	.007
	Group2	.258	24	<.001	.892	24	.014
	Group3	.297	21	<.001	.889	21	.022
Mean score of correct stress placement for [e]	Group1	.196	23	.023	.914	23	.049
	Group2	.211	24	.007	.919	24	.055
	Group3	.297	21	<.001	.889	21	.022
Mean score of correct stress placement of four sounds	Group1	.187	23	.035	.930	23	.111
	Group2	.145	24	.200*	.973	24	.745
	Group3	.149	21	.200*	.953	21	.383

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Tests of Normality for the Mean Score of Segmental Pronunciation Accuracy in Task 4

	Age Group	Kolmogorov-Smirnova			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean score of correct pronunciation of [ɑ̃]	Group1	.532	23	<.001	.324	23	<.001
	Group2	.331	24	<.001	.718	24	<.001
	Group3	.355	21	<.001	.674	21	<.001
Mean score of correct pronunciation of [ɛ̃]	Group1	.479	23	<.001	.512	23	<.001
	Group2	.473	24	<.001	.531	24	<.001
	Group3	.492	21	<.001	.484	21	<.001
Mean score of correct pronunciation of [y]	Group1	.477	23	<.001	.535	23	<.001
	Group2	.161	24	.108	.923	24	.069
	Group3	.279	21	<.001	.805	21	<.001
Mean of correct pronunciation of [e]	Group1	.532	23	<.001	.324	23	<.001
	Group2	.427	24	<.001	.622	24	<.001
	Group3	.382	21	<.001	.711	21	<.001
Mean score of four sounds	Group1	.283	23	<.001	.749	23	<.001
	Group2	.158	24	.123	.901	24	.023
	Group3	.128	21	.200*	.943	21	.247

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Tests of normality for correct stress placement in task 4

	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean score of correct stress placement	Group1	.212	23	.008	.816	23	<.001
	Group2	.174	24	.057	.938	24	.151
	Group3	.112	21	.200*	.937	21	.194

Tests of Normality for the Mean Score of Comprehensibility in Task 4

	Age Group	Kolmogorov-Smirnova			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean score of comprehensibility	Group1	.148	23	.200*	.940	23	.178
	Group2	.180	24	.044	.927	24	.084
	Group3	.168	21	.123	.926	21	.117

Tests of Normality for the Mean Score of Accentedness in Task 4

	Age Group	Kolmogorov-Smirnova			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Mean score of accentedness	Group1	.166	23	.101	.950	23	.295
	Group2	.135	24	.200*	.962	24	.487
	Group3	.103	21	.200*	.983	21	.957

Appendix M

Pronunciation mistakes

Pronunciation mistakes in task 1

The target sound	Mistake	Word	Pronounced form	Number of participants who made the mistake	
[ɑ̃]	[ɑn]	kangourou	[kɑnguru]	24	76
		ange	[anʒ]	26	
		blanc	[blan]	7	
		argent	[arʒan]	19	
	[am]	ampoule	[ampul]	7	61
		pamplemousse	[pampləmus]	19	
		tambour	[tambur]	27	
		température	[tamperatyr]	3	
		jambe	[ʒamb]	5	
	[ɛ̃]	Serpent	[sɛrpɛ̃]	11	11
	[ə]	Serpent	[sɛrpə]	5	6
		Jambe	[ʒəb]	1	
	[en]	Ange	[enʒ]	5	5
	[em]	Température	[temperatyr]	2	2
	[u]	Serpent	[sɛrpu]	1	1
[ɛ̃]	[in]	dindon	[dindɔ̃]	17	49
		interdit	[intɛrdi]	4	
		intelligent	[inteliʒɑ̃]	8	
		raisin	[rɛzin]	19	
		poussin	[pusin]	1	
	[en]	dindon	[dendɔ̃]	4	27
		interdit	[entɛrdi]	6	
		peinture	[pentyr]	13	
	[em]	important	[empɔrtɑ̃]	1	16
		Faim	[fem]	15	
	[ɔ̃]	magicien	[maʒisjɔ̃]	4	13
		interdit	[ɔ̃tɛrdi]	1	
		intelligent	[ɔ̃teliʒɑ̃]	1	
		poussin	[pusɔ̃]	4	
		important	[ɔ̃pɔrtɑ̃]	1	
		peinture	[pɔ̃tyr]	2	

	[am]	faim	[fam]	5	11
		important	[ampɔrtɑ̃]	6	
	[e]	intelligent	[eteliʒɑ̃]	5	10
		main	[me]	5	
	[im]	important	[impɔrtɑ̃]		6
	[an]	interdit	[antɛrdi]	3	4
		main	[man]	1	
	[i]	dindon	[didɔ̃]	1	3
		interdit	[itɛrdi]	1	
		faim	[fi]	1	
	[wa]	peinture	[pwatyr]	3	3
[y]	[i]	urgence	[irʒɑ̃s]	16	141
		Uniforme	[inifɔrm]	12	
		Allumette	[alimɛt]	16	
		Tulipe	[tilip]	10	
		Voiture	[vwatir]	15	
		Chaussure	[ʃosir]	16	
		Legume	[legim]	16	
		Cube	[kib]	12	
		Jumeaux	[ʒimɔ]	18	
		Laitue	[lɛti]	10	
	[u]	Uniforme	[unifɔrm]	2	15
		Allumette	[alumɛt]	1	
		Tulipe	[tulip]	11	
	[a]	Urgence	[irʒɑ̃s]	6	13
		Allumette	[alamɛt]	4	
		Jumeaux	[ʒamɔ]	1	
		Laitue	[lɛta]	1	
		cube	[kab]	1	
	[an]	uniforme	[anifɔrm]	11	11
	[ə]	laitue	[lɛtə]	2	5
		jumeaux	[ʒəmɔ]	3	
	[e]	urgenece	[erʒɑ̃s]	2	5
		jumeaux	[ʒemɔ]	3	
	Deleted	allumette	[almɛt]	1	1
[e]	[ej]	Échelle	[ejʃɛl]	5	45
		Élèves	[ejlɛv]	1	

	Été	[etej]/ [ejte]	21	
	Blé	[blej]	4	
	Fée	[fej]	14	
[i]	Échelle	[iʃɛl]	6	27
	Élèves	[ilɛv]	7	
	Etoile	[itwal]	4	
	Zéro	[ziro]	4	
	Été	[ite]	3	
	Fée	[fi]	2	
	Blé	[bli]	1	
[ɛ]	Été	[ɛte]	3	10
	Blé	[blɛ]	4	
	Fée	[fɛ]	3	
[ə]	Été	[etə]	2	6
	Blé	[blə]	4	
	Fée	[fə]	1	
[u]	Été	[etu]	2	2
[aj]	Été	[etaj]	1	1
[a]	Blé	[bla]	1	1
Deleted 1	Blé	[bl]	1	1

Pronunciation mistakes in task 3

The target sound	Mistake	Correct word	Pronounced form	Number of participants who made the mistake	
[ɑ̃]	[ɑn]	Pantalon	[pantalɔ̃]	10	31
		Enfant	[anfɑ̃]	1	
		Dentiste	[dantist]	2	
		Dent	[dan]	2	
		Maman	[maman]	6	
		Orange	[ɔranʒ]	5	
		Danser	[danse]	3	
		Viande	[vjand]	2	
		Éléphant	[elefan]	5	
	[ɔ̃]	Manteau	[mɔ̃to]	6	6
	[o]	Manteau	[moto]	5	5
	[ɛ̃]	Enfant	[ɛ̃fɑ̃]	1	3
		Médicament	[medikamɛ̃]	2	
	[u]	Manteau	[muto]	1	1
[ɛ̃]	[e] 18	Cinquante	[sekɑ̃t]	18	18
	[a] 16	Cinquante	[sakɑ̃t]	13	16
		Imprimante	[aprimɑ̃t]	1	
		Pain	[pa]	2	
	[en] 13	Ceinture	[sentyr]	7	13
		Infirmière	[enfirmjɛr]	4	
		Princesse	[prensɛs]	1	
		Cinquante	[senkɑ̃t]	1	
	[an] 6	Princesse	[pransɛs]	1	6
		Imprimante	[anprimɑ̃t]	4	
		Infirmière	[anfirmjɛr]	1	
	[in] 4	Ceinture	[sintyr]	2	4
		Infirmière	[infirmjɛr]	2	
	[ɔ̃] 3	Ceinture	[sɔ̃tyr]	1	3
		Pain	[pɔ̃]	1	
		Lapin	[lapɔ̃]	1	
	[i] 3	Ceinture	[sityr]	1	3
		Cinquante	[sikɑ̃t]	2	
	[em] 1	Imprimante	[emprimɑ̃t]	1	1
	[u] 1	Lapin	[lapu]	1	1

[y]	[i]	Musée	[mize]	7	141
		Lunette	[linɛt]	24	
		Sucette	[sisɛt]	14	
		Musique	[mizik]	9	
		Écureuil	[ekirœj]	5	
		Jus	[ʒi]	7	
		Tortue	[tɔrti]	18	
		Autobus	[otɔbis]	26	
		Confiture	[kɔ̃fitir]	8	
		Sucre	[sikr]	4	
		Ceinture	[sɛ̃tir]	10	
	[u]	Musée	[muze]	3	18
		Sucette	[susɛt]	1	
		Musique	[muzik]	12	
		Écureuil	[ekurœj]	2	
[e]	[ej]	Vélo	[vejlo]	3	45
		Lego	[lejgo]	2	
		École	[ejkɔl]	3	
		Poupée	[pupej]	7	
		Nez	[nej]	14	
		Clé	[klej]	4	
	[i]	Vélo	[vilo]	6	27
		Lego	[ligo]	14	
		École	[ikɔl]	8	
		Médecin	[midsɛ̃]	4	
		Éléphant	[elifɑ̃]	12	
		Boulanger	bulɑ̃ʒi]	2	
		Café	[kafi]	1	
		Nez	[ni]	1	
	Deleted	Café	[kaf]	1	1
	[a]	Éléphant	[alefɑ̃]	1	1
	[y]	Lego	[lygo]	1	1