University of Reading

Orthography in L2 Phonology Acquisition: The role of L1/L2 phoneme similarity and phonemic coding ability

Thesis submitted in fulfilment of the requirements for the degree of PhD (Taught Track & Thesis) in Applied Linguistics

School of Literature and Languages, University of Reading

Haifa Saleh M. Alhumaid

July 2022

Supervisors: Prof. Jane Setter

Dr Daisy Powell

Proofread by: Gerard Hearne

Abstract

Orthographic input has been the focus of much investigation in the search for the second language (L2) phonological development. The focus of studies investigating this issue has been directed on to the role of experience factors in modulating this effect, e.g., congruency level between the first language (L1) and L2 orthographic systems, and level of orthographic depth of L1 and L2. However, to the best knowledge of the researcher, few studies have linked the role of internal factors with experience factors (e.g., Bassetti, Mairano, Masterson, & Cerni, 2020). This study seeks to understand the role of orthographic input in terms of availability and script familiarity in discriminating between three pairs of Arabic phonemes in word-level (phonemes embedded in words) and phoneme-level (phonemes uttered in isolation) discrimination. This role is investigated in relation to the participants' phonemic coding ability and L1 (English) /L2 (Arabic) phoneme similarity. The pairs of Arabic phonemes investigated in this study were (/θ/ and /ð/), (/s/ and /sˤ/), and (/χ/ and /ʁ/). These three pairs were selected based on Flege's (1986) equivalence classification in which the first is identical to English, the second is similar to English, and the last is novel to English.

After taking a phonemic coding ability test, participants were allocated into three orthographic groups: 1) unfamiliar script (Arabic), 2) a combination of familiar and unfamiliar scripts (Roman and Arabic), and 3) no orthographic input. Generally, easily perceived contrasts (identical and novel phonemes) were found to be discriminated better than hard-to-distinguish contrasts (similar phonemes). Notwithstanding that all groups did well overall in both phoneme discrimination levels, there were some differences in these two levels. In the word-level discrimination, findings indicated that the poorest performance was reported by those who had

Roman along with Arabic script. Groups who had unfamiliar script or had no script at all did not show any significant difference from each other in their performance. Participants performed best when they were asked to discriminate between two new phonemes, whereas there was no significant difference in their performance when they were asked to discriminate between two identical or two similar phonemes, which might be caused by impaired learning in the case of identical phonemes and perceived difficulty in the case of similar phonemes. In the phoneme-level discrimination, orthographic input type did not have any significant effect on participants' performance. On the other hand, Participants did not show any significant difference in their performance when they were asked to discriminate between two new or two identical phonemes, whereas their poorest performance was when they were asked to discriminate between two similar phonemes, which might be caused by the perceived difficulty of the two phonemes. There was no significant interaction between L1/L2 phoneme similarity level and orthographic input type in both levels of discrimination. Phonemic coding ability, however, was found to not affecting participants' performance in the following discrimination tasks.

The study contributes to our understanding of models such as the Speech Learning Model (SLM) (Flege, 1995), Perceptual Assimilation Model (PAM) (Best, 1995), and Second Language Perceptual Assimilation Model (PAM-L2) (Best & Taylor, 2007), which argue that similar phonemes tend to be more difficult to discriminate than different ones. The findings also are supported by Cognitive Load Theory (Sweller, 2011) which encourages taking the capacity and duration limits of working memory into consideration when devising instructional procedures to avoid hindering learning.

Acknowledgements

First, I would like to thank the University of Reading for providing me with all that I needed to carry out this study. I owe a debt of gratitude for the opportunities and facilities they provide for research students.

I must express my sincere and warmest appreciation to my supervisors, Prof. Jane Setter and Dr Daisy Powell, for their academic supervision and constant support and patience. They monitored my progress and offered advice and encouragement throughout the study. This thesis would not have been completed without their valuable leadership and supervision.

Also, my heartful thanks go to Dr Jacqueline Laws and Dr Parvaneh Tavakoli, who occupied the position of PhD Programme director during the time of conducting my study, and Prof. Rodney Jones, who is head of the Department of English and Applied Linguistics, for their continuing encouragement and optimism throughout my studies.

I am eternally grateful to my great parents who have been an unstinting source of support. Their care and prayers were an inspiration to carry out this study. My acknowledgements would not be complete without thanking my siblings, Ebtihal, Reema, Juri, Mohammed, and Mouath who gave me unfailing love, encouragement, advice and reassurance. My dear siblings offered me continuous valuable assistance, which was essential in helping me complete this project.

Finally, there is a multitude of individuals who helped me arrive at this point, I welcome this opportunity to thank my friends, Alhanouf, Maha, Dona, Ghaya, and Aya, who believed in me and were always there for me.

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Candidate:

Haifa Alhumaid

Table of Contents

List of Abbreviations

APPC	Arabic phonotactic probability calculator
CAH	Contrastive Analysis Hypothesis
CG	Category Goodness
CoR	Confirmation of Registration
DLAP	Defence Language Institute Battery
F1	First Formant
F2	Second Formant
F3	Third Formant
GCSE	General Certificate of Secondary Education
IPA	International Phonetic Alphabet
L1	First Language
L2	Second Language
L2LP	Second Language Linguistic Perception
MDH	Markedness Deferential Hypothesis
MLAT	Modern Language Aptitude Test
MSA	Modern Standard Arabic
NA	Non-assimilation
PAM	Perceptual Assimilation Model
PAM-L2	Second Language Perceptual Assimilation
PLAB	Pimsleur Language Aptitude Battery
SC	Single-category Assimilation
SD	Standard Deviation

SLM	Speech Learning Model
SLM-r	Revised version of Speech Learning Model
TC	Two-category Assimilation
UC	Uncategorised vs Categorised
UU	Uncategorisation

List of Tables

List of Figures

1 Introduction

1.1 Background

In recent years, researchers have shown an increased interest in the impact of orthography input on L2 phonological development. It was found that orthography plays an essential role in forming the phonology of the second language (Bassetti, Cerni, & Masterson, 2022). Determining the impact of orthographic input on L2 phonological development is important for the future of language learning (e.g., Bassetti, 2017) and language teaching (e.g., Bassetti & Atkinson, 2015; Escudero, Hayes-Harb, & Mitterer, 2008 2008; Escudero, 2015) and has theoretical implications (e.g., Bassetti, 2017; Bassetti Sokolović-Perović, Mairano, & Cerni, 2018; Showalter & Hayes-Harb, 2013; 2015). However, studies in this area have focused on a variety of different factors and used different methodologies, concluding with various and conditional results. This means that some studies have found that orthographic input is helpful in developing phonological acquisition in some conditions but not in others (Bassetti, 2008; Bassetti & Atkinson, 2015; Erdener & Burnham, 2005; Escudero, 2015; Simon, Bernard, Lalonde, & Rebaï, 2006).

1.2 Scope of the Study

This study aimed to shed light on the relationship between three factors. It aimed to investigate the role of learners' language aptitude and L1/L2 phoneme similarity in modulating the effect of orthographic input on L2 phonological development. Language aptitude is defined by Carroll (1993) as the cognitive ability which predicts certain level of future success in learning which excludes any interest or motivation of any particular activity. One aspect of language aptitude was included in this study, which is phonemic coding ability and its role in predicting L2 phonological perception. Phonemic coding ability refers to a learner's ability to analyse and retain unfamiliar sounds (Sparks, Patton, Gangschow, & Humbach, 2011), and to make links between sound strings and their corresponding graphemic symbols (Saito, 2019).

One aspect of orthographic input was considered in this study, namely, familiarity with scripts. This study set out to review in detail the available information on the effect of familiar (Romanised Arabic) and unfamiliar (Arabic) scripts, in addition to the role of a learner's phonemic coding ability in the perception of three Arabic phoneme pairs by native speakers of British English. The typology of the target phonemes was selected based on Flege's (1986) equivalence classification of L2 sounds. The first pair consisted of the two Arabic interdental phonemes, /θ/ and /ð/, which are identical to English phonemes and therefore map to two English phonemes. Identical phonemes refer to cases where learners cannot detect differences between L2 phonemes and their closest L1 phonemes at a sensory level (Flege, 2016). The second pair consisted of the two Arabic alveodental phonemes, /s/ and /sˤ/, which are similar to each other but not a separate category in English and therefore have only one counterpart in English. Similar phonemes are those which have acoustic differences from their closest L1 counterparts (Flege, 1986). The third pair consisted of the two Arabic uvular phonemes, /ʁ/ and /χ/, which are entirely new phonemes that do not exist in English. New phonemes are phonemes that do not map to direct equivalents in L1 (Flege, 1986). All participants were naïve listeners of Arabic. The data were responses to a phonemic coding ability test (Llama-E proposed by Meara, 2005; Meara & Rogers, 2019), in addition to responses to phoneme identification tasks in a perceptual experiment containing the six target fricative Arabic consonants. The audio stimuli in the perceptual experiment task were recorded by a native Arabic speaker and presented to participants via PsychoPy 3 Software developed by Peirce, et al. (2019). All the data collected in this study were analysed statistically.

1.3 Rationale of the Study

One of the difficulties faced by L1 English learners of L2 Arabic is the rich consonant inventory that Arabic has (Almahmoud, 2013; Hayes-Harb & Durham, 2016). Arabic phonology is characterised by having emphatic phonemes which do not exist in most world languages, including English (Almahmoud, 2013; Embarki, Yeou, Guilleminot, & Almaqtari, 2007). Emphatic phonemes are defined as those consonants that are produced with a coarticulation of the main place of articulation of the sound, in addition to the back of the tongue moving towards the pharyngeal wall (Amayreh & Dyson, 1998). According to Hayes-Harb and Durham (2016), studies on the perception of Arabic emphatic consonants by L1 English learners are still limited. Given that this characteristic is absent in many languages, research has consistently shown that L2 Arabic learners do not attain an adequate understanding of this characteristic in the Arabic phonology, which leads to them having several difficulties in acquiring certain emphatic phonemes, in either perception or production (Embarki et al., 2007). Being unable to discriminate such contrasts in consonants could result in losing the ability to understand what thousands of Arabic words mean (Odisho, 2005 as cited in Hayes-Harb & Durham, 2016). This is caused by the inability to establish new L2 phonetic categories for these sounds, which, according to Flege (2016), makes it impossible to understand the fundamental inter-personal variability that accounts for L2 learning. On the other hand, investigating the relationship between perceptual abilities, such as learners' phonemic coding ability, and L2 learning performance, helps in building up variable perceptual foundations, which in turn determines how ready individuals are to learn an L2 and consequently enriches the relationship between language aptitude and success in L2 learning (Sun, Saito, & Tierney, 2021).

1.4 Contribution of the Study

Recent developments in L2 perception studies, especially the reciprocal studies, have heightened the need to establish the relationship between sounds in both L1 and L2, in addition to what cues speakers of each language use. This consequently leads to making more precise predictions about how to formulate cross-language speech perception (Schmidt, 2007). The conclusions drawn from studies that have investigated the role of orthography in L2 phonological development are varied and contrast with each other to a degree that underscores the need for a deeper and wider investigation. In addition, the investigation of L2 Arabic psycholinguistics has been limited due to the complex oral system and unique orthographic system of Arabic (Aljasser & Vitevitch, 2018). This, in addition to the complexity caused by the existence of emphatic phonemes that contrast with non-emphatic phonemes, encouraged the researcher of the current study to select Arabic. Therefore, this study seeks to remedy certain problems by enhancing both theoretical and practical fields in learning Arabic phonology.

1.4.1 Theoretical Contribution of the Study

The current dominant speech perception models, such as SLM (Flege, 1995) and its revised version SLM-r (Flege & Bohn, 2021), PAM (Best, 1995), and PAM-L2 (Best & Taylor, 2007), neglect the role of orthographic input in L2 phonological development (Bassetti, 2017; Nimz & Khattab, 2020). However, Flege, (2016) argues that orthography tends to affect speech perception negatively. Therefore, orthography should not be related to speech learning. This is because, according to Flege, (2016), establishing L2 phonetic categories should depend on phonetic input that is not accessed via orthography. Flege (2016) defines the phonetic categories as language representations which specifically determine all the properties of the speech sounds that belong to a given language and differentiate these sounds from each other.

They also determine all the properties of any pair of sounds coming from different language regardless of their IPA representations. These language representations are defined either by acquired distinctiveness or acquired similarity. The former increases the sensitivity to the differences occurred in the boundary between two different categories. The latter decreases the sensitivity to the differences between members which belong to the same category. Despite the fact that the phonetic category should be accessed via phonetic input rather than orthography, the concept of the effect of orthography on L2 phonological development has recently been challenged by several studies demonstrating the positive effect of exposing L2 learners to orthography (e.g., Bassetti & Atkinson, 2015; Escudero et al., 2008; Showalter & Hayes-Harb, 2013) (more discussion about these studies will be presented in the Literature Review chapter). Flege (2016), however, built his argument based on examples of the negative effect of orthography due to the incongruency between L1 and L2 orthographic systems, and/or orthographic depth, which I will discuss in further detail in Subsection 2.4. According to Mathieu (2016), the role of orthographic input in forming the L2 phonological lexicon in the early stages needs a more comprehensive understanding, especially in the case of unfamiliar orthographic systems. Therefore, one of the objectives of the current research is to determine whether familiarity and unfamiliarity with orthographic input affects the participants' ability to discriminate phonemes.

On the other hand, Flege (2016) raised the question of whether the ability of late learners to establish phonetic categories for L2 sounds can be attributed to some special skill or aptitude. Bassetti (2008) suggests that the effect of orthography might be modulated by other factors including, but not limited to, a learner's internal factors. She suggests examining the role of learners' phonemic coding ability in their reliance on orthographic representations to discriminate phonemes. The predictability of phonemic coding ability as a subcomponent of language aptitude in L2 achievement is well evidenced in the literature (e.g., Artieda & Muñoz

2016; Carroll, 1993; Granena, 2013, 2018; Li, 2016; Meara, 2005; Reynolds, 2002; Saito, 2017, 2019; Sparks et al., 2011).

According to Kogan (2020), L2 speech perception studies have been carried out in two paradigms. The first paradigm directs its focus on individual differences without considering L1/L2 phoneme similarity. The second paradigm investigates cross-language perception by focusing on group results instead of individuals. However, to the best of the current study researcher's knowledge, there is no research that links the three aforementioned factors together. Several studies have been carried out to examine the role of orthography in relation to the perceptual difficulty caused by L1/L2 phoneme similarity (e.g., Broersma, 2005; Cutler, 2015; Escudero, 2015; Escudero Broersma, & Simon, 2013; Escudero & Wanrooij, 2010; Escudero & Williams, 2012; Showalter & Hayes-Harb, 2015; Mathieu, 2016), but there are still insufficient data on the role of learners' phonemic coding ability in this effect. According to Saito (2019), to enrich L2 speech aptitude research, it is necessary to have more systematic investigations to assess the applicability of the existing aptitude framework in L2 pronunciation proficiency. Therefore, this study aims to assess the role of phonemic coding ability with regard to the effect of orthographic input and L1/L2 phoneme similarity.

1.4.2 Practical Contribution of the Study

Investigating L2 perception is important for enhancing the pedagogical field by informing L2 instructors of the most noticeable problems faced by learners in L2 perception (Chan, 2012). Chan (2012) discusses that investigating L2 perception is important to achieve accurate L2 production. This is because inaccurate L2 perception leads to inaccurate L2 production (Chan, 2012; Flege, 1995; Kogan, 2020). Kogan (2020) assumes that the reason behind difficulties in L2 production is having difficulties in L2 perception, which is why Kogan (2020) believes that the role of perception should be prioritised in the L2 acquisition process.

On the other hand, Shariq (2015) argues that the similarities and dissimilarities between English and Arabic phonemes should be elaborated to signal the problems faced by learners. This elaboration is addressed by training learners in how English and Arabic sounds are pronounced, taking account of the place and manner of articulation of these sounds in both languages. This helps teachers to devise helpful procedures for learners to get over these problems.

Despite the fact that the current study considers not only similar sounds (emphatic vs non-emphatic) but also identical and new sounds, Chan (2012) suggests that the main focus of perception training should be on L2 sounds that are similar in perception to certain L1 sounds, instead of sounds that do not exist in L1. In this way, L2 instructors can help learners to be aware of the differences between similar sounds. Therefore, understanding the difficulty caused by Arabic emphatics and their effect on adjacent vowels (as discussed in Subsection 2.5) benefits both Arabic language instructors and learners (Hayes-Harb & Durham, 2016). This helps in directing learners' attention towards the most confusing patterns of L2 sounds, which in turn helps these learners to discriminate between these sounds (Chan, 2012). This can then lead to the establishment of new phonetic categories for those similar sounds in the learner's L2. In addition, according to Flege (2016), it helps in understanding the fundamental inter-personal variability that accounts for L2 learning. However, far too little attention has been paid to teaching Arabic as a second language, so it is still, according to Hayes-Harb and Durham (2016), in need of more guided extensive research which aims to enhance the pedagogy for native English speakers.

1.5 Objectives of the Study

This research was carried out in an attempt to reach the following objectives:

1. To gain a better understanding of the role of orthography in L2 phonological development.
2. To evaluate whether L1/L2 phoneme similarity plays a part in the ease and difficulty of discriminating L2 sounds.
3. To investigate how orthographic input and L1/L2 phoneme similarity interact in facilitating the perception of L2 sounds.
4. To assess the predictability of learners' phonemic coding ability in their reliance on orthographic input and/or L1/L2 phoneme similarity when discriminating between two L2 sounds.

1.6 Outline of the Thesis

The overall structure of the study comprises six chapters. This chapter begins by presenting an overview of the research area, touching upon the rationale of the study and the contribution this study aimed to make to the field, including both practical and theoretical aspects. It highlights the gap in the literature this study aimed to fill as well as the objectives of this study. Finally, it concludes by narrowing down the scope of the study.

The second chapter conducts a deeper discussion and presents previous studies on the main topics related to the current study. The main issues addressed in this chapter are: (a) the influence of L1/L2 phoneme similarity on the development of L2 phonology, (b) the predictability of phonemic coding ability in L2 proficiency, (c) the influence of orthographic input on L2 phonological development, and (d) an overview of the phonological and orthographic systems of Arabic. In the first section of this chapter, the influence of L1/L2 phoneme similarity in L2 phonological development explains the emergent themes influencing L2 phonological development in relation to L1 phonology by conducting a detailed discussion of the most dominant L2 speech perception models. The second section, the predictability of phonemic coding ability in L2 proficiency, highlights the key theoretical concepts of language aptitude and discusses specific methods for testing the subcomponents of language aptitude, including an evaluation of the validity of the Llama test battery. The third section of the chapter, the influence of orthographic input on L2 phonological development, draws together various findings of this effect and takes account of different factors such as the congruency level between learners' L1 and L2 orthographic systems, the orthographic depth of L1 and L2 phonological systems, the perceived difficulty of L2 sounds, and familiarity with scripts. The fourth section of this chapter conducts a descriptive discussion of Modern Standard Arabic (MSA). This section includes a more detailed discussion of how some phonemes are recognised in relation to each other and the basic rules of MSA orthography, touching upon the opaqueness of the orthographic system of Arabic. It also gives a brief overview of the English phonological system in comparison with the MSA phonological system. This chapter concludes by presenting the five research questions this study aimed to investigate after highlighting a gap in the literature.

The third chapter concerns the methodology employed for this study. It includes how the study was designed. It presents the criteria for participant recruitment in the study. It also elaborates, in detail, what types of stimuli were included in this study and how these stimuli were prepared and presented to participants. In addition, it clarifies how ethical procedures were considered. In this chapter, an explanation of how the instrument was piloted is also presented. Moreover, this chapter provides a detailed description of the instruments and procedures used to collect data. Also, it explains what methods and tools were used to analyse those data. This chapter also includes a justification for using and selecting certain procedures and methods, and some types of stimuli.

The fourth chapter conducts an analysis of the data gathered. The beginning of the chapter explains the approach followed to analyse data and what methods were used to conduct these analyses for each research question. This is followed by a baseline analysis, which precedes the main analysis, in order to test that all conditions were alike before starting the study treatment. In the following section, the results of testing whether participants were answering better than by chance or not in parts which are included in the main analysis of the study are presented. This section is followed by the main analysis section, which was conducted to answer the main research questions. This section consists of three subsections. The first subsection analyses the results of participants' performance in perceptual tasks assessing the effect of orthographic input type and L1/L2 phoneme similarity on their ability to discriminate L2 phonemes in order to provide answers to the first three research questions. The second subsection presents the results of a correlational analysis between participants' performance in two different phases to answer the fourth research question. The third subsection analyses the results of participants' performance in perceptual tasks in relation to their performance in a phonemic coding ability test to assess its role in modulating the effect of orthographic input type and L1/L2 phoneme similarity in participants' ability to discriminate L2 phonemes in order to answer the final research question. This chapter concludes by presenting a summary of the main findings of the study.

The fifth chapter discusses the significant findings of the study. It consists of two main sections. The first section summarises the principal findings of the study and conducts a brief discussion of these findings. The second section starts with a general discussion of the findings, highlighting some issues faced during the data analysis. This section is further divided into five subsections, where each subsection conducts a detailed discussion and gives an answer to one research question. The first subsection discusses the findings for the effect of orthographic input type on participants' ability to discriminate L2 phonemes to answer the first research question. The second subsection conducts a detailed discussion of the findings for the effect of L1/L2 phoneme similarity on participants' performance to answer the second research question. The third subsection discusses the relationship between orthographic input type and L1/L2 phoneme similarity and its effect on participants' performance to answer the third research question. This is followed by the fourth subsection that discusses the findings of the correlational analysis between participants' performance in two different phases to answer the fourth research question. The final subsection discusses, in detail, the findings for the role of phonemic coding ability in modulating the effect of orthographic input type and L1/L2 phoneme similarity in participants' ability to discriminate L2 phonemes, which answers the final research question. This chapter is summarised in the final section.

Finally, the sixth chapter draws a conclusion to the study, including a summary of the major findings. It also discusses some pedagogical implications of the findings of this study. It includes some limitations that the researcher faced in the process of this study. This chapter concludes by making some suggestions and recommendations for future studies.

2 Literature Review

2.1 Introduction

One of the problems faced by L2 learners is discriminating non-native contrasts that do not exist in their L1, either perceptually or productively (Escudero, 2015). Being unable to discriminate these contrasts consequently leads learners to misperceive minimal pairs that are contrasted by these phonemes; however, this is not the case when contrasting phonemes can be readily discriminated (Escudero et al. 2013)

Factors specifying the ease and difficulty of discriminating non-native contrasts and L2 phonological development are various. One of these factors is the fact that phonemic inventories differ from one language to another (Guion, Flege, Akahane-Yamada, & Pruitt, 2000). Flege (2016) argues that, according to the SLM proposed in Flege (1995), L2 learners tend to rely on two factors that determine whether to establish a new phonetic category for an L2 sound. Phonetic categories refer to language representations which specifically determine all the properties of the speech sounds that belong to a given language and differentiate these sounds from each other. They also determine all the properties of any pair of sounds coming from different language regardless of their IPA representations (Flege, 2016). SLM-r (Flege & Bohn, 2021) hypothesises that the first factor that determines the establishment of a new L2 phonetic category is that L1 categories tend to strongly attract the closest L2 sound. This in turn decreases the possibility of establishing a new category for an L2 sound. The other factor is, as the SLM-r hypothesises, when the perceived dissimilarity of an L2 sound is large, it is easier to establish a new category for an L2 sound. Therefore, according to Flege (2016), L2 speech is continuously affected by the L1 phonetic system despite the proficiency level of L2 learners. This is because the L2 phonetic system develops while it shares existing space with the L1 phonetic system. However, Flege (2016) concludes that establishing L2 phonetic categories is affected by factors other than the perceived similarity and dissimilarity between L1 and L2 phonetic systems. These include age of learning, amount of L2 and L1 use, and amount of L2 input. It is worth highlighting that the primary difference in the revised version, SLM-r (Flege & Bohn, 2021), is the fact that L2 category formation is based on three factors: to what extent the phonemes have perceived dissimilarity; to what extent L1 phonetic category is precise; and the quality and quantity of L2 input. Also, according to SLM-r, production and perception co-evolve without precedence. Moreover, SLM-r looks at individuals' learning rather than how groups differ from each other.

Moreover, one essential factor that has been investigated and found to affect L2 phonological development and word recognition is the presence and type of orthographic input during L2 phonological learning. Studies showed that orthographic input has an impact on learners' ability to discriminate non-native contrasts, which include not only perception but also production and acquisition of L2 word form learning (Bassetti et al., 2018; Escudero et al., 2008; Showalter & Hayes-Harb, 2013, 2015). This impact of orthography on word recognition and phoneme discrimination has been found to be moderated by several factors (Hayes-Harb & Barrios, 2021). These include, but are not limited to, the congruency level between learners' L1 and L2 orthographic systems (e.g., Erdener & Burnham, 2005; Escudero et al., 2008; Showalter & Hayes-Harb, 2015), the orthographic depth of L1 and L2 phonological systems (e.g., Erdener & Burnham, 2005), the perceived difficulty of L2 sounds (e.g., Cutler, 2015; Escudero, 2015), and the familiarity of the script (e.g., Mathieu, 2014, 2016; Showalter, 2012; Showalter & Hayes-Harb, 2013, 2015).

It is acknowledged in the literature that experience factors such as, the presence of orthography or how much learners practise the L2, are not the only factors that explain how adult learners differ in their success in L2 phonological acquisition. This is apparent in the fact that while some individuals may be exposed to identical or similar input in terms of amount and quality, their achievement in L2 perceptive and productive skills is different (Sun et al., 2021). Thus, another factor that might have a direct effect on the ease and difficulty of L2 phonological development and learners' ability to discriminate non-native contrasts is individual differences between L2 learners. Some individuals differ from one another in their ability to perceive and discriminate non-native contrasts accurately and rapidly. These differences are caused by several factors, for example, but not limited to, general intelligence, language aptitude, and socio-psychological factors (Hanulíkuvá, Dediu, Fang, Bašnaková, & Huettig, 2012). The literature shows that some learners have a specific cognitive ability which makes them different from one another in their L2 development (Hanulíkuvá et al., 2012). This ability is known as language aptitude which, according to Carroll (1993), refers to cognitive ability which predicts the level of future success in learning but excludes any interest in or motivation for any particular activity. Carroll (1993) emphasises that if a learner has greater cognitive ability than his/her colleagues, that does not necessarily imply that he/she will be superior in other aspects of language. These abilities differ from one learner to another. These abilities are acquired quickly and easily by some learners, but this is not the case with other learners, even if they are more highly motivated. Thus, the success of language learning is mainly influenced by language aptitude (Granena, 2018). Kogan (2020) claims that some individuals might achieve native-like phonetic and phonological proficiency even if they are late learners, whereas others maintain a heavy foreign accent even if they manage to master other factors of language. Kogan's findings support this claim as it was found that learners differ from each other in their success when dealing with sound contrasts that do not exist in their L1.

The three previously mentioned factors, namely L1/L2 phoneme similarity, presence of orthography, and individual differences, imply that the ease and difficulty of L2 phonological development differs from one context to another (depending on L1/L2 phoneme similarity, the amount and type of orthographic input and many other factors), and from one individual to another (depending on his/her general intelligence, language aptitude, socio-psychological factors and many others).

2.1.1 Chapter outline

This chapter consists of five sections. The next section (Section 2.2) presents models and approaches associated with L2 perception in relation to L1/L2 phoneme similarity along with some evidence from the literature supporting these models. The following section (Section 2.3) discusses the predictability of phonemic coding ability as an important component of language aptitude that plays a key role in the success of language proficiency. This section also reviews several studies conducted to investigate the validity of phonemic coding ability tests. Section 2.4 reviews different findings of studies that have examined the role of orthography in phonology and categorises them depending on the different methodologies they used. Section 2.5 presents a detailed description of Arabic in terms of its phonological and orthographic systems. It also briefly touches upon a comparison between the Arabic and English phonological systems. This chapter concludes by presenting the gap that this study attempted to bridge after reviewing several related studies, ending by proposing five research questions for this study in Section 2.6.

2.2 Influence of L1/L2 Phoneme Similarity on L2 Phonological Development

2.2.1 Introduction

Recent examinations of how non-native listeners categorise allophones are helpful in revealing the fact that subphonemic details in both L1 and L2 are important in making predictions about how the perception of L2 occurs. However, according to Best and Taylor (2007), the level of difficulty varies in non-native segmental contrasts as some contrasts are discriminated at near native-like levels whereas others are discriminated only moderately well. Best and Taylor argue that this relative ease or difficulty might be directly influenced by the listener's L1 phonology.

This section discusses how L1/L2 phoneme similarity plays a role in determining the level of difficulty in acquiring L2 phonology. It conducts a detailed discussion of L2 perceptual models and how these models link L2 perception to learners' L1 phonology. A review of how the literature has investigated these models is also discussed in this section. In addition, this section includes some implications of Flege (1995) and his colleagues' work (1995) on L2 speech perception.

2.2.2 L2 Phonological Acquisition in Relation to Learners' L1 Phonology

In the history of L2 learning research, L1 phonology has been thought of as a key factor in L2 phonological development. According to Almahmoud (2013), the earliest attempts to reveal the L1 phonological system in terms of filtering L2 sound acquisition were in the 1930s. This view was followed by a number of studies which were conducted in the field of L2 phonological acquisition and attributed the difficulties faced by L2 learners when learning speech sounds to the differences in structure of L1 and L2 (Chan, 2012). These studies led to the development of a considerable number of models and hypotheses. By way of illustration, the Contrastive Analysis Hypothesis (CAH) formulated by Lado (1957) and the Markedness Differential Hypothesis (MDH) proposed by Eckman (1977 as cited in Chan, 2012) both state that the difficulties faced by L2 learners can be predicted by researchers and L2 teachers by systematically highlighting the similarities and differences between L1 and L2, where difficulties arise due to the different structures and sounds from those in L1. Thus, the two hypotheses built their theoretical frameworks based on attributing the difficulties of L2 acquisition to the differences between L1 and L2 structures.

This approach appeared to be inadequate (Eckman, 1987 as cited in Almahmoud, 2013), because it was unable to provide an explanation for the reasons behind the ease in learning the contrasts between some L1 and L2 sounds and the difficulty in learning the similarities between some L1 and L2 sounds (Almahmoud, 2013). This consequently led to the formulation of speech acquisition models that shed light on areas of difficulty in learning the L2 sound system. Many of these L2 speech acquisition models disagree with the CAH and MDH hypotheses which assert that difficulty in L2 speech perception and production can occur or be predicted depending only on the existence or absence of sounds in L1 and L2 (Binasfour, 2018). In other words, these L2 speech perception models suggest that the similarities between L1 and L2 sounds may cause more difficulties than do differences between sounds. Oller and Ziahosseiny (1970, as cited in Chan, 2012) argue that L2 sounds that are similar to L1 sounds tend to be more difficult for learners to acquire than different sounds. Cutler (2015) points out that this difficulty increases when two similar phonemes in L2 have only one equivalent phoneme in L1. Hu et al. (2019) support this, arguing that, generally, distinguishing L2 phonemes might be done very easily due to the assimilation of these two phonemes into two different L1 phonetic categories, whereas it will be more difficult when they are assimilated to a single L1 phonetic category, or not assimilated to any L1 phonetic category.

Flege (2016) suggests two ways of developing phonetic categories. The first is by acquired distinctiveness which increases the sensitivity to differences occurring on the boundaries between categories. The second is by acquired similarity which decreases the sensitivity to differences between the same category members. Flege (1995) suggests that the difficulty in acquiring similar sounds occurs due to the learner's classification of these sounds as equivalent to his/her L1 sounds, which is not the case with new sounds that seem to be easier because of the learner's ability to easily notice the differences between L1 and L2. Flege (1986) refers to this process as "equivalence classification". By way of demonstration, Flege (1995) argues that L2 learners recognise the L2 phonological system by either the addition of new phonetic categories or the modification of existing ones. Thus, according to Flege, it is a commonly held view that recognising L2 phonological categories may fall into one of three possible cases. The first is when the L2 sound is recognised as an "identical" to an L1 sound; this leads to replacement of the L2 sound by the L1 sound, even if they are phonetically different. The second is when there are L2 sounds that are contrasted in L2 but do not exist in L1; this can result in learners not acquiring this contrast, which is caused by what is referred to as "interlanguage mapping between two contrasting L2 categories" (Schwartz & Kaźmierski, 2020, p. 227). The last is when there are contrasting sounds in L1 that do not exist in L2; this leads to the possibility of producing these contrasts in L2 (Weinreich, 1953 as cited in Flege, 1995). However, according to Flege (1995), this categorization neglects the role of age of learning and the amount of L2 experience. Chan (2012) elaborates on this by arguing that it is predictable that L2 learners tend to judge L2 phonemes depending on realizations of the closest learners' L1 category. Learners build new L2 phonetic categories if they can notice the difference between an L2 sound and the nearest L1 sound. However, if two L2 contrasts are distinguished in L2 but not in L1, this contrast will be poorly detected. According to SLM-r (Flege & Bohn, 2021), age is no longer regarded as a factor that plays a role in building L2 phonetic category, but rather building L2 phonetic category is affected by how precise the L1 phonetic category is formed. This means that individuals are better in discerning phonetic differences when they have precise L1 phonetic categories.

Flege's (1995) classification is in line with the claim of earlier interlanguage studies' that learning L2 phonology is merely a restructuring continuum (Corder, 1981 as cited in Chan, 2012), which means that L1 is seen as the starting point of L2 acquisition, but learners replace their L1 by L2 gradually when restructuring their interlanguage (Ellis, 1994). Schmidt (2007) claims that speech sounds are interpreted in relation to L1 phonemic categories and perceived selectively, depending on L1 requirements. This is evidenced by the fact that some items are perceived as different phonemes by listeners who have different language backgrounds (Lisker & Abramson, 1964, as cited in Schmidt, 2007). For instance, Japanese learners found the contrast between the two English phonemes /ɹ/ and /l/ difficult to discriminate, due to this contrast not existing in Japanese (Guion et al., 2000). Spanish and Portuguese learners also face difficulties in distinguishing between the two vowels /i:/ and /ɪ/, like in *beat* and *bit*, caused by the fact that these two vowels have no counterparts in Spanish or Portuguese (Escudero et al., 2008).

According to Mack (1988 as cited in Flege, 1995), learning L2 speech is found to be more analytical than in L1 acquisition, especially if L2 learners are exposed to L2 through written input in the early stages of learning. Wayland (2007) argues that adults differ from children in that adults are language-specific perceivers. Their perception of speech sounds depends largely on L1 phonological filtering. This is why the base of the perceptual discriminability of L2 speech sounds is formed by finding perceptual relationships between L1 and L2 speech sounds. Schmidt (2007) agrees with that, defining L2 speech perception study as examining the process of perceiving acoustic or gestural information depending on a pre-existing sound's individual systemization, which is the listener's L1 phonology, which indicates that this process is derived from the perception of L1 speech, at least the beginning of it.

According to Flege (1995), during the development of L1 speech, speech perception becomes automatic. This may lead learners to pay less attention to some phonetic detail when learning L2 sounds. Consequently, learners become unable to access some sensory properties related to certain L2 sounds. However, Flege argues that in some later stages of learning, L2 learners might discard some sensory information that has been processed previously as non-distinctive, or they might give different weights to some features compared to native speakers. Hence, Flege argues that as bilinguals depend for their interpretation of L2 sounds on the "grid" of their L1 phonology, this leads some L2 vowels and consonants to be perceived by non-native speakers differently from native speakers. This is because the phonetic category established by non-native speakers for those vowels and consonants might be based on different features from those of native speakers. According to Flege (1995), the possibility of this happening increases when two L2 sounds are contrasted by features that do not exist in L1. Guion et al. (2000) agree with that, suggesting that the perception of L2 sounds goes through a different process from that of native speakers, which results in perceiving target language sounds differently from how native speakers do it. Thus, Flege (1995) argues that the classification of L2 sounds should

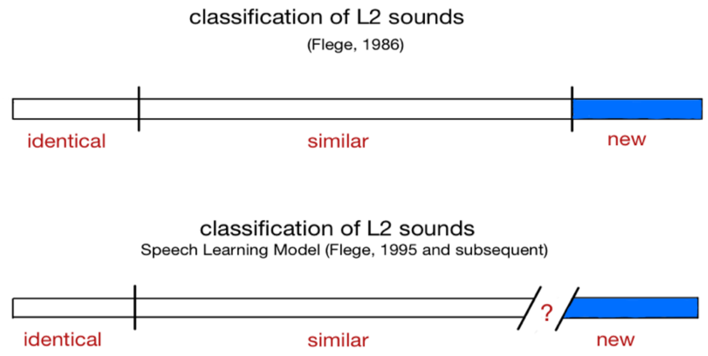


Figure 2.1 Classification of L2 sounds before and after SLM (Flege, 2016)

include "a priori metric" that occurs between similar and new sounds and distinguishes between them (Fig. 2.1).

Schwartz and Kaźmierski (2020) support the Flege's classification by providing evidence from the literature that L2 learners use acoustic features to a greater degree than native speakers. For example, L1 Spanish speakers were found to produce L2 English vowels with a greater degree of duration, even though a duration factor does not exist in their L1 (Bohn, 2017; Escudero & Broersma, 2004, both as cited in Schwartz & Kaźmierski, 2020). Schwartz and Kaźmierski (2020) argue that, in these cases, learners' performance shows less sensitivity to the position of the vowel in acoustic space than L1 English speakers. This is because they depend on duration for the production of contrast with the closest L1 vowel.

2.2.3 Models in Second Language Speech Perception

The factor of L1/L2 phoneme similarity received attention as early as in the 1970s (Chan, 2012). However, the end of the twentieth and the beginning of the twenty-first centuries witnessed the development of different speech perception models, among them the previously mentioned SLM by Flege (1995) and its revised version SLM-r (Flege & Bohn, 2021), which consists of several assumptions (discussed in detail in Subsection 2.2.3.1). The SLM-r states that phonetic differences between L1 and L2 sounds or between two L2 sounds may not be discerned because these two sounds map to one phonetic category in learners' L1 or because the features or properties forming the contrast are filtered out in the learners' L1 phonology. Inspired by Flege, Major and Kim (1999) proposed the Similarity Differential Rate Hypothesis, which claims that dissimilarity in sounds leads to faster acquisition. In addition, Best (1995) and Best and Taylor (2007) developed two hypotheses, PAM and PAM-L2, respectively, which propose that similarity in articulation causes the assimilation of L2 sounds to native language sounds. These models are discussed in more detail in the following subsections.

2.2.3.1 Speech Learning Model (SLM) and its revised version (SLM-r)

The SLM-r, proposed by Flege and Bohn (2021), is a model based on four postulates that indicate the relationship between L2 phonological development and learners' L1 phonology. The first postulate indicates that the same mechanism used to process L1 sound system is activated over the life span and used to process L2 sound system as well. The second postulate is based on the concept that long-term memory specifies some representations of language-specific aspects of speech sounds as phonetic categories. These categories which some of them were established in childhood for L1 sounds keep developing over time to include all properties of L1 and L2 sounds that are seen as realisations of any of these categories. This is the third postulate. The last postulate states that bilinguals keep trying to maintain how L1 and L2 phonetic categories, which share the same phonological space, contrast. Flege (2016) argues that the SLM (the previous version of SLM-r) is based on two concepts. First is the fact that the capacities used in the acquisition of L1 are maintained by L2 learners, even in late learning. Second is the fact that L2 speech is continuously affected by the L1 phonetic system, even with very experienced learners, which occurs because the developing L2 phonetic systems shares the L1 phonetic system in a common space. This model suggests that at the beginning of L2 phonology exposure, all L2 sounds may be identified as different allophones of L1 phonemes. However, as L2 learners gain more experience in L2, they begin to discern the phonetic differences between specific L2 sounds and their closest L1 sounds. This is the process of establishing a new phonetic category representation for a new L2 sound. This representation becomes independent of previously established representations of L1 sounds (Flege, 1995; Flege & Bohn, 2021). In other words, the L1 sound phonetic category that was established in childhood may develop when associated with an L2 sound.

According to the SLM-r, there are three important variables that may influence the development of the phonetic category: degree of L1/L2 phoneme perceived similarity; precision of L1 phonetic category; and quality and quantity of L2 input (Flege & Bohn 2021). The greater the distance between an L2 sound and its closest L1 sound, the more likely that a separate phonetic category of L2 sounds will be established. Also, the quantity and quality of L2 input which learners are exposed to, and how precise an L1 phonetic category is formed, determines how well an individual can discern phonetic differences.

Hypotheses that formulate the SLM (Flege, 1995) and SLM-r (Flege & Bohn, 2021):

- H1: There is a perceptual relationship between L1 and L2 sounds. This relationship occurs at a position-sensitive allophonic level rather than a phonemic level. In other words, learners perceptually associate L2 positional allophones to the closest L1 allophone or phoneme that is positionally defined as the closest allophone.
- H2: If bilinguals become able to discern some of the phonetic differences between L1 and L2 sounds, a new phonetic category for the L2 sound will be established which is different phonetically from the closest L1 sound.
- H3: If the perceived dissimilarity between an L2 sound and its closest L1 sound is large, it is more likely that learners will discern the phonetic differences between L1 and L2 sounds.
- H4: It is likely that as the age of learning increases, the ability to discern phonetic differences between L1 and L2 sounds, or between two L2 sounds that do not contrast in L1, decreases. In SLM-r, age is not a factor in L2 phonetic category formation as it is possible regardless of the age.
- H5: The mechanism of equivalence classification might lead to blocking the category formation of an L2 sound. This leads to one single phonetic category for both linked L1 and L2 sounds, which are named 'diaphones', a term introduced by Weinreich (1957, as cited in

Flege, 1995), which refers to linking L1 and L2 sounds perceptually. This results in using these diaphones to represent each other in production. SLM-r, however, adopts the full-access hypothesis which claims that learners can access all processes and mechanisms that they used to develop the phonetic categories of their L1, with no exception, to form L2 phonetic categories.

- H6: Bilingualism and monolingualism might lead to a difference in the phonetic category established for an L2 sound. This happens if a bilingual establishes an L2 category that is deflected away from the L1 category in order to keep the contrast in a common phonological space in L1 and L2, which refers to the case where pronunciation is affected by both L1 and L2. Also, this might happen when a bilingual and a monolingual have different representations based on different features or different feature weights.
- H7: The properties represented in a phonetic category representation of a particular sound correspond to the production of this sound.

It is worth highlighting some of the differences proposed by Flege and Bohn (2021) between SLM and SLM-r. The focus of SLM was on highly experienced L2 learners, which is abandoned by SLM-r because, according to Flege and Bohn (2021), L2 learners will never reach the same level as monolingual native speakers of the L2. This is because there will be an interaction between L1 and L2 phonetic elements that form the two phonetic subsystems in addition to the fact that the input of L2 learners and native speakers is impossible to be identical. Moreover, SLM claims that the accuracy of how L2 sounds are perceptual represented precede the accuracy of how these sounds are produced. However, SLM-r claims that perception and production co-evolve with no precedence. SLM claims that the phonetic category formation process requires a smaller phonetic distance when L2 learning starts at an earlier age. SLM-r, on the other hand, abandoned the age factor and replaced it with precision of L1 phonetic categories in which if L1 phonetic category is more precisely formed, this leads to better discern phonetic differences between L1 and L2 sounds regardless of the age. Furthermore, one of the hypotheses of SLM is the “Feature hypothesis” which proposes that if L2 phonetic category is defined by features that do not exist in learners’ L1, the new established phonetic category will be different from that formed by the native speakers. According to SLM-r, this hypothesis is replaced by the “Full-access hypothesis” which claims that all processes and mechanisms used to form L1 phonetic categories will active during L2 learning. In addition, according to the SLM, there are two important variables that may influence the development of the phonetic category: L1/L2 phoneme similarity and learners’ age. However, in SLM-r, L2 category formation is based on three factors: to what extent the phonemes have perceived dissimilarity; to what extent L1 phonetic category is precise; and the quality and quantity of L2 input. Finally, SLM-r looks at individuals’ learning rather than how groups differ from each other as SLM.

2.2.3.2 Perceptual Assimilation Model (PAM) and Second Language Perceptual Assimilation Model (PAM-L2)

PAM and PAM-L2 are two models proposed by Best (1995) and Best and Taylor (2007), respectively. They are based on the concept that naïve L2 listeners (PAM) and L2 learners (PAM-L2) tend to assimilate L2 sounds to their native language sounds depending on their articulatory similarities. This means that the phonetic similarities between non-native contrasts and the phonological categories of the listener’s native language determine how these non-native contrasts are perceived (Almahmoud, 2013; Chan, 2012). This assimilation of contrasting segments to an L1 category can be explicit or implicit and is determined by the extent to which the phonetic closeness or discrepancy of L1 and L2 sounds is shared (Best, 1995).

Best and Taylor (2007) argue that one major difference between PAM and SLM is that PAM is a non-native speech perception model and is directed at naïve listeners, whereas the SLM is an L2 speech acquisition (perception and production) model and is directed at experienced listeners. According to Best and Taylor, neither of these two models was developed for both situations. The difference in perception between these two populations occurs due to experiential influences, age at onset of L2 acquisition, and some aspects of fluency and usage of L1 and L2. Moreover, according to Best and Taylor (2007), PAM differs from the SLM in that the SLM focuses on native phonetic category formation depending on acoustic-phonetic cues, whereas PAM focuses on the articulatory gestures of native and non-native sounds. Despite these differences, Best and Taylor (2007) agree that these two models still share some common features. For instance, both PAM and the SLM extended their predictions about native language influences on the phonological contrasts to include the importance of non-contrastive phonetic similarities and dissimilarities between L1 and L2/non-native phones in terms of their phonetic goodness of fit and the relationship between phonological categories and contrasts and phonetic details. In addition, both PAM and the SLM claim that both adults learning an L2 and children learning an L1 or L2 have the same basic perceptual learning abilities.

However, despite those similarities and differences between the two models, Best and Taylor (2007) reported that there have been cases where the SLM and PAM were wrongly used interchangeably, which led them to introduce PAM-L2 which, according to them, concentrates on how these two models (SLM and PAM) are placed in terms of their commonalities and complementarities. Hence, the difference between the SLM and PAM-L2 is that PAM-L2 expands its focus to include equivalence not only at the phonetic level as the SLM does, but also at the phonological level, although both agree that L1 and L2 phonological categories are placed in a common phonological space. Best and Taylor (2007) claim that the aim of PAM-L2 is to examine whether it is possible to use the SLM as a starting point to extend PAM from a non-native speech perception framework to include L2 learners. Thus, the difference between PAM and PAM-L2 is that the former is directed at naïve listeners who are not exposed to formal instruction, whereas the latter is directed at L2 learners who are in the process of learning. PAM-L2 links the postulates proposed by the SLM and, depending on the PAM perspective, provides an evaluation of them (Best & Taylor, 2007). In this context, non-native listeners are defined by Best and Taylor as “functional monolinguals” (p. 16), which means they are not in the process of active learning and/or using L2 and are linguistically naïve to the target language; on the other hand, L2 learners are defined as people who are actively learning an L2, aiming to achieve goals that might be functional or communicative, and not only gaining some educational requirements. Therefore, the difference between naïve listeners and L2 learners occurs in that little non-L1 experience results in adults establishing a pattern of non-native speech perception, which means that functional monolinguals have clearly find it difficult to categorise and discriminate several phonetic contrasts in unfamiliar languages that do not mark contrasts in L1 lexical items.

According to both models (PAM and PAM-L2), the ability to perceive L2 sounds is governed by the perceptual similarities and differences between L1 and L2 sounds. In other words, the articulatory similarities between L1 and L2 sounds lead non-native listeners to perceptually assimilate unfamiliar L2 sounds to the closest perceptually similar sounds in their L1, which happens due to their experience of the native language. Best (1995) claims that there is a link between discriminating an L2 sound and judging its goodness of fit category (discussed below). This happens because discriminability is predicted by perceptual assimilation types which are determined by perceived category goodness. A non-native phone is heard as one of these three cases: a good or a poor exemplar of a native language phonological segment which means it is categorised, different from any single native phoneme which means it is uncategorised, or as non-linguistic non-speech sound which means it is unassimilated. This comparison between L1 and L2 sounds results in pronunciation errors in these contrasts (Best 1995; Best & Taylor, 2007). These categorizations are explained in more detail in the following paragraph. PAM and PAM-L2 suggest that L2 speech learning involves interaction between both phonetic and phonological levels based on the relationship between L1 and L2 phonological spaces. For instance, if L2 learners are able to perceptually assimilate an L2 phonological category to an L1 phoneme and the L1 and L2 phonetic versions of this phonological category are discriminable, then the listener should be able to categorise the two phones of L1 and L2 as separate phonetic realizations of that phonological category.

Best (1995) suggest that this categorization of non-native sounds for naïve listeners can take three forms. First, they are assimilated to an L1 sound and heard as either an identical exemplar of this L1 sound, an acceptable exemplar of this L1 sound, or a deviant exemplar of this L1 sound. Second, they are heard as speech sounds that exist in the phonological space but do not represent any L1 sound. Finally, they are heard as non-speech sounds. These three forms are based on five patterns (as shown in Fig. 2.2).

- Two-category (TC) assimilation in which two L2 sounds are assimilated to two different native categories because they are perceived as two acceptable exemplars of two different native sounds. In this case, listeners are predicted to have very good to excellent perception and discrimination of L2 contrasts.
- Single category (SC) assimilation in which two L2 sounds are assimilated to a single native category in which both L2 sounds are perceived equally well or are poor exemplars of the native-language sound. The perception and discrimination of these sounds are predicted to be poor.
- Category goodness (CG) difference which refers to the case where two L2 sounds are assimilated to the same native category and they are perceived differently in terms of goodness

of fit to the native-language phoneme. In this case, the perception and discrimination of these contrasts are predicted to fluctuate between moderate and very good.

- Uncategorizsation, in which one or both L2 sounds do not match any native-language phoneme. If one L2 phoneme is assimilated to a native phoneme while the other is perceived as an uncategorised speech sound (UC), the perception and discrimination of the contrast are predicted to be very good because the phonological distinction is clearly reflected. However, if both L2 sounds are perceived as uncategorised (UU), the perception and discrimination of this contrast are predicted to fluctuate between poor and moderately well.

- Non-assimilation (NA), in which neither L2 sound is perceived as a speech sound because they deviate from the articulatory properties of native phonemes. The discrimination of this contrasts is predicted to vary from good to excellent.

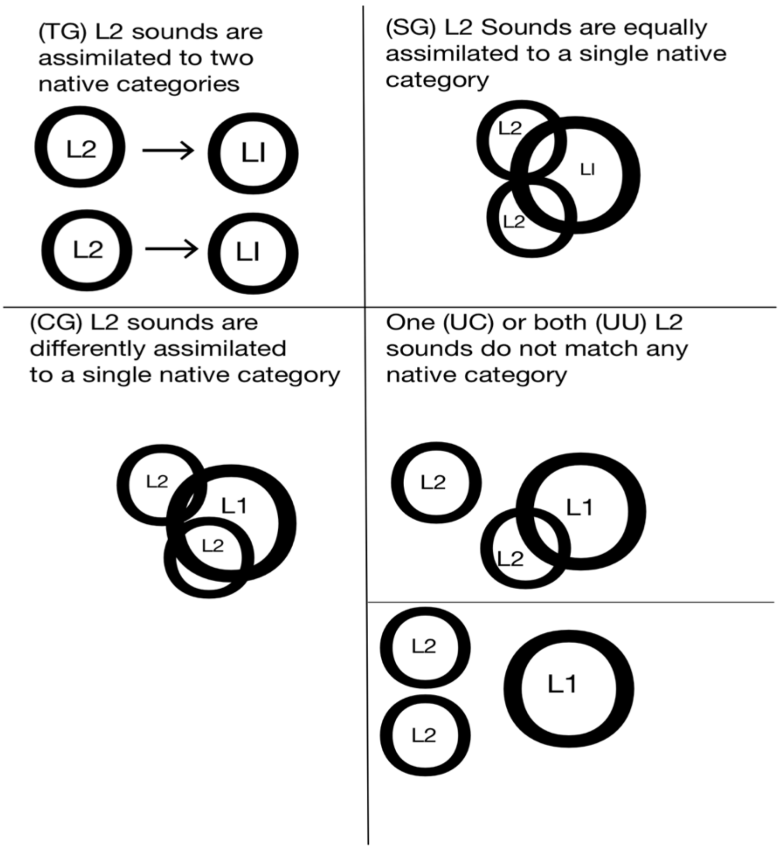


Figure 2.2 Patterns of categorisation of non-native sounds for naïve listeners

On the other hand, Best and Taylor (2007) propose four possible cases for contrasts in L2 in an attempt to extend PAM's framework to include predicting success in L2 perceptual learning. The first case is perceptually assimilating one L2 phonological category to one L1 phonological category. In this case, L2 learners are predicted to have less difficulty in perceiving and discriminating this sound compared with other contrasts in L2 categories. Also, it is predicted that learners will perceive L1 and L2 sounds not only as equivalent phonological categories, but also as equivalent phonetic categories. The second case is when two L2 phonological categories are perceptually assimilated to one L1 phonological category, with one being perceived as more dissimilar to the native phonological category than the other. The perception and discrimination of these two sounds are predicted to be good but not as good as if they are assimilated to two native categories. Therefore, it is predicted that a new phonological and phonetic category will be established for the more dissimilar one, whereas the other phoneme is perceived as equivalent to an L1 phonological and phonetic category which results in learners easily being able to recognise the contrast between the two sounds. The third case is when two L2 phonological categories are perceptually assimilated to one L1 phonological category with equally good or equally poor exemplars of the native category. Learners are predicted to initially have difficulties in discriminating this contrast. In order to establish a phonological category or categories in this case, learners should first perceptually learn to establish a phonetic category of at least one L2 sound. The last case is when learners do not have any L2 and L1 perceptual assimilation. So, learners do not assimilate either L2 contrast to any L1 phonological category, but instead have a mixture of similarities to several L1 phonological categories for both L2 sounds. The perception and discrimination of one or both sounds are predicted to be perceptually easy, which is in line with the SLM-r concept of new sounds.

Both the SLM and PAM models concentrate on the similarities between L1 and L2 phonetic and phonological features (Schwartz and Kaźmierski, 2020). Also, the Ontogeny Phylogeny Model proposed by Major (2001) agrees that the similarity between L1 and L2 phonemes plays an essential role in the ease or difficulty of phoneme acquisition. Major (2001) suggests that if a particular L2 sound is categorised as more similar to an L1 sound than to another L2 sound, then acquisition of the first one will be slower than acquisition of the second one. It is important to note that all these models mark the similarity, not difference, between L1 and L2 sounds to be the determinant of the perceptual difficulty of acquiring second language phonology.

The SLM-r and PAM models have received considerable attention. As outlined below, a number of studies support their hypotheses, whereas other studies do not show supportive evidence for these claims. Flege (1995) supports his hypotheses, that together formulate the model, with several findings that are seen as evidence for it. By way of illustration, according to Flege (1995), the fact that L2 learners succeed in producing and perceiving some allophones of English phonemes and fail in others can be seen as evidence supporting H1. For example, Flege, Takagi and Mann (1996) examined the acquisition of liquids by three groups of learners: L1 English speakers, experienced L1 Japanese learners of English, and inexperienced L1 Japanese learners of English. It is important to mention that Japanese groups alveolar median approximants ([ɹ]), alveolar lateral approximants ([l]) and similar sounds (e.g., [ɾ]) into a single phoneme category. Flege et al.'s (1996) findings show that the phoneme /ɹ/ was identified at a near-perfect rate by experienced learners and it was more likely for a new English category to be established as the English phoneme /l/ is closer than the English phoneme /ɹ/ to the Japanese phoneme /r/.

Chan (2012) also examined the first three hypotheses of the SLM by investigating the perception of English phonemes by 40 L1 Cantonese learners of English. The participants of this study completed two L2 perception tasks as well as three L1 and L2 perception tasks that asked them to categorise English phonemes into a group of similar Cantonese phonemes and rate to what extent the English and Cantonese phonemes were similar. Chan found that Cantonese /l/ and English /ɹ/ were rated and perceived as the least similar pair. They also found that the perception of English /ɹ/ was best in the case when a contrast occurs with English /l/. Chan argues that the perception of English /ɹ/ was successful because learners had established a new English phonetic category for the phoneme /ɹ/ as learners perceived less similarity between Cantonese /l/ and English /ɹ/, and this led them to discern the phonetic difference between them.

McAllister (2007) conducted a study to examine Flege's SLM in two respects: first, testing the fifth hypothesis – 'the feature hypothesis' – and second, testing the sixth hypothesis. The fifth hypothesis states that if listeners fail to discern a mismatch in a phonetic feature that signals a contrast between L1 and L2 sounds, the formation of an L2 phonetic category might be blocked. In other words, it is difficult to accurately perceive L2 sounds that have features which do not signal phonological contrasts in L1. This difficulty will be reflected in the production of these contrasts. The sixth hypothesis states that L2 learners use different strategies from those used by native speakers in order to establish a phonetic category and this difference will be reflected in the production of a contrast. McAllister's context was the acquisition of two contrasting English phonemes, /s/ and /z/, in a syllable-final position only, by L1 Swedish learners of English. It is worth highlighting that Swedish and English both have the feature of quantity distinction. Both languages involve a complex relationship in this feature between a temporal dimension (duration of the vowel and the following consonant) and a spectral dimension (the value of the formant in the vowel). While this feature is distinctive in Swedish, it is known traditionally that English phonology does not exploit this feature (McAllister, Flege, & Piske, 2002). McAllister (2007) reported that, in the early stages of investigating this problem, it was found that voice contrast in the final position of a syllable was marked by the ratio between the duration of the preceding vowel and the duration of the consonant itself (see Denes, 1955; Flege & Hillenbrand, 1986; Peterson & Lehiste, 1960, all as cited in McAllister 2007). He hypothesised that his L1 Swedish participants would be able to produce the two English contrasts /s/ and /z/. This was based on previous research (McAllister et al. 2002) that found that if an L1 user can manipulate a phonetic feature via his/her L1 phonology, this might facilitate L2 acquisition if it uses this feature. This depends on the fact that Swedish speakers have a quantity feature in their L1 that is essential in the phonetic realization of Swedish phonology. Therefore, as L1 Swedish speakers do have a quantity distinction in their L1 phonology and they can distinguish durational differences, the feature hypothesis would expect an L1 Swedish speaker to exploit this experience in contrasting /s/ and /z/ in syllable final position, because this contrast uses syllable nucleus length. McAllister (2007) also hypothesised that the participants might use the same phonetic method to differentiate two contrasts in their production, which are different durations of the preceding vowel. However, his findings did not support his first hypothesis as the participants failed to master the contrast between /s/ and /z/ in word final position. His second hypothesis, on the other hand, was successful as he found that participants who were successful in producing /s/ and /z/ (two participants who produced 8 out of 15 successful productions of /z/) mimicked the strategy of native speakers in their production of the /s/ and /z/ contrast. McAllister attributed his first finding, which was not in line with his hypothesis, to the fact that, although Swedish has a voicing distinction between some consonants, it does not have /s/ and /z/ contrast, which is why he thought that the fifth hypothesis of Flege's SLM would be applicable in his case. The difference in the feature produced by Swedish and English to mark voicing contrast might block category formation. This feature difference occurs in the use of duration as quantity contrast is important allophonically in English but prosodically in Swedish.

Hayes-Harb and Durham (2016) also contributed to the evidence for the SLM and other similar models by examining L1 English learners of L2 Arabic and their perception of the rime after a plain onset (e.g., /dæk/) versus an emphatic/ pharyngealised onset (e.g., /dˤæk/). Hayes-Harb and Durham formulated their research questions on the assumption that L1 Arabic speakers and L2 Arabic learners depend on rime in order to determine the existence or absence of the pharyngealisation feature in onset, as found by numerous studies (Alhumaid, 2019; Binasfour, 2018; Jongman et al., 2011; Holes, 2004; & Shar & Ingram, 2010) (for a more detailed discussion see Subsection 2.5.2.2.1). Hayes-Harb and Durham found that their L1 English participants were able to discriminate the plain form from its emphatic counterpart when it was followed by the vowel /æ/ more than when it was followed by the vowels /u:/ or /i:/. This was attributed to mapping the two allophones of the vowel /æ/ to two different phonetic categories in English /æ/ when following a plain consonant, and /ɑː/ when following an emphatic consonant, which is not applicable to the other two vowels, /u:/ and /i:/, as they do not have similar vowels with a more back feature in English (caused by pharyngealisation). Odisho (1981, as cited in Hayes-Harb & Durham, 2016) points out that having two different vowels in English, /æ/ and /ɑː/, helps L2 Arabic teachers to teach emphatic consonants to L1 English learners of Arabic.

Schmidt (2007) conducted a follow-up to a previous study (Schmidt, 1996, as cited in Schmidt, 2007) which examined perceptual similarity in English syllable initial consonants by native Korean learners. In the more recent study, the researcher conducted the same study in the opposite direction, i.e., perceptual similarity in Korean syllable initial consonants by native English learners. These two studies were conducted in order to identify some factors that affect similarity decisions. Thus, participants of both studies were asked to assign consonants to the closest consonants in their native language and to rate the degree of similarity of perceived and L1 consonants. Both studies concluded with the same findings, in that the acoustic cues in some phonemes were the same in both studies, such as /p^h, t^h, k^h, j, h/. Schmidt (2007) found that acoustic cues affected the strength of domain-initial articulation. For example, Korean listeners in Schmidt (1996, as cited in Schmidt, 2007) tended to pay more attention when labelling English /b/ in individual tokens to intensity rise time and vowel onset. It was also found that the following vowel affected how participants rated perceived similarity. When the following vowel was /u/, the similarity rating of the preceding consonant was negatively affected, in that it was rated as less similar to the native consonant than when it preceded /i/ or/ /ɑ/. The researcher attributed this to the fact that the Korean vowel /u/ is more lip-rounded than the English vowel /u/, which resulted in perceived acoustic differences. Schmidt (2007) concluded that L1-acceptable variation in the production of acoustic cues introduces systematic different acoustic cues that may significantly help L2 listeners. Therefore, it was found that L2 listeners perceive L2 sounds as sounds within L1 categories based on the acoustic cues that these categories are related to.

On the other hand, Harnsberger (2001) examined the five assimilation types of Best's PAM in the discrimination of six types of nasal consonants from the Malayalam language by seven listener groups. These groups have systematic differences in their native sound inventories. The six nasals (bilabial /m/, interdental /n̪/, alveolar /n/, retroflex /ɳ/, palatal /ɲ/, and velar /ŋ/) were put in different places of articulation. These consonants in particular were selected because they were expected to be perceptually challenging consonants for all non-native listener groups. Participants were asked to categorise and rate the category goodness of these consonants in an AXB task in which they were presented with two sounds and asked which one of them was more similar to a sound that existed in their L1. Participants were native speakers of Malayalam, Marathi, Oriya, Punjabi, Tamil, Bengali, and American English. After completing the AXB discrimination task, each nasal contrast was identified as one of the five PAM assimilation types depending on the participants' categorisations and ratings. The findings of this study did not support PAM's predictions. This was because some effects of linguistic experience, which cannot be attributed to abstract units such as phonemes and allophones, were found. By way of illustration, it was expected for the alveolar and the dental-retroflex nasal phoneme groups to easily classify these consonants due to their nasal consonant inventory. However, this was not the case as these consonants were realised by suballophonic aspects of these consonants' perceptual categories that could not be described by the researcher. Harnsberger (2001) attributed his findings to: being independent of his participants' phonemic or allophonic inventory, including the overall dispersion of nasals; clustering nasals to different patterns such as interdental-alveolar-retroflex and palatal-velar; and finally there being some similarity between bilabial and retroflex nasals.

Lababidi and Park (2014, as cited in Hayes-Harb & Durham, 2016) conducted a similar study to Chan's (2012) in which L1 English participants were asked to label Arabic consonants which were presented to the participants orally by classifying them to the closest corresponding English phonemes. Following the "category goodness" of Best (1995), participants labelled both Arabic phonemes /d/ and /dˤ/ as good and fair exemplars of the English phoneme /d/, respectively. Lababidi and Park (2014 as cited in Hayes-Harb & Durham, 2016) concluded that their participants' inability to perceive a contrast between the two Arabic phonemes /d/ and /dˤ/ and their categorization of these two phonemes into a single English phonetic category is evidence of the fact that their perception was hindered by their L1 phonology.

Wayland (2007) also conducted a study examining the influence of the methodological variable on the fit between predicted and actual discrimination scores in an attempt to explain the relationship between the identification and discrimination of L2 contrasts. Wayland examined the perception of the nine stop Korean consonants (/p^h, p, p', t^h, t, t', k^h, k, k'/) by L1 Thai listeners, and the perception of the six Thai stop consonants (/b, p, p^h, d, t, t^h/) by L1 Korean listeners. Both groups of participants were able to discriminate stop consonant contrasts easily, despite the fact that these consonants have different phonetic properties. These results indicate that perceived assimilation patterns are not necessarily predicted by acoustic and/or articulatory-phonetic properties, as claimed by Best's PAM. In addition, the findings suggest that it is possible to have various phonetic cues used by listeners as a tool to gauge perceptual similarities between speech sounds. This is because these speech sounds are weighted and integrated differently, and they serve as a function of the context wherein these sounds occur. This leads the listener to perceive non-native sounds as belonging to different native categories which have different degrees of fit in different contexts, either when perceived or produced.

Almahmoud (2013) examined four assimilation types of PAM, namely, Two-category (TC), Category goodness (CG), Uncategorised speech sound (UC), and Uncategorisation (UU) in the acquisition of L2 Arabic phonemes by L1 American English speakers. His study included 12 Arabic phonemes compared to seven English phonemes. The Arabic phonemes are (/θ, ð, k, h, t, d, tˤ, ðˤ, q, ħ, χ, ʁ/), whereas the English ones are (/θ, ð, k, h, t, d, g/). Based on PAM assimilation types, it was predicted that the discrimination between the phonemes (/θ/ and /ð/) and (/t/ and /d/) would be excellent because they might be perceived as identical exemplars of English phonemes and each pair was predicted to be assimilated into two separate English categories (TC). The contrasts between the two emphatics and non-emphatics (/tˤ/ and /t/) and (/ðˤ/ and /ð/) were expected to be discriminated moderately to very well, because the emphatics might be heard as poor exemplars of English (/t/ and /d/), respectively, and, therefore, take the (CG) type. The first phoneme of the pharyngeal-glottal (/ħ/ and /h/) and uvular-velar (/q/ and /k/) pairs were predicted to resemble poor exemplars of the English phonemes /h/ and /k/, respectively, and therefore take the (UC) type where the discriminability might be moderate to very good. Finally, the contrasts between the three pairs (/χ/ and /ʁ/), (/χ/ and /ħ/), and (/χ/ and /q/) were predicted to be non-assimilable to any English category and therefore take the UU type. Findings for the (TC) types were in line with the researcher's prediction based on PAM. (CG) aligned with the prediction in PAM only in the case of (/ðˤ/ and /ð/), but not (/tˤ/ and /t/). This was attributed to the possibility that the emphatic phoneme in the former pair was more deviant from its English counterpart than the emphatic phoneme in the latter pair, and therefore discriminating the first pair was easier. (UC) type did not align with the researcher's prediction as discrimination of the contrast between (ħ/ and /h/) and (/q/ and /k/) was poor. This might have happened, according to the researcher, due to the phonetic proximity of the segments. Finally, the (UU) type showed various findings. For the contrast between (/χ/ and /ʁ/) and (/χ/ and /ħ/), the discrimination was very poor. However, the contrast between (/χ/ and /q/) was discriminated very well. This supports the prediction of PAM, which claims that the discrimination of (UU) fluctuates between poor to very good. Almahmoud (2013) concluded that the ease of discrimination in his study followed the scheme: (TC) followed by (CG) followed by (UC and UU).

2.2.4 Implications and Limitations of Flege and his colleagues' Work on L2 Speech Perception

There is no doubt that Flege and his colleagues' investigations into second language performance have contributed to forming important frameworks of speech perception models. However, much of the criticism that their work has attracted relates to the weak link between that and foreign language teaching theories (Piske, 2007). This, according to Piske (2007), might be why the findings of their work did not receive much attention in the foreign language teaching literature. This might be due to one major reason, which is most of the investigations in their work being directed to examine immigrants, not foreign language students. This has established a gap because the two populations differ in many aspects. Immigrants, for example, are surrounded by L2 as a predominant language, which is not the case with foreign language students who are not necessarily surrounded by L2 as a predominant language. In addition, the conditions of learning L2 are different between immigrants and foreign language students because immigrants have more exposure to L2 with more high-quality input such as native-speaker input, and more frequent use of the language than foreign language students (Piske, 2007).

However, despite these differences between the two populations, Piske (2007) identified several common features shared by immigrants and foreign language students which indicate that the factors that influence immigrants' learning might also have an influence on foreign language learning. This supports the idea of employing Flege and his colleagues' work in language teaching contexts. The first common feature shared by both populations is the age of learners; both immigrants and foreign language students can start learning in either early or late stages of life. In addition, immigrants do not necessarily use the language more than students. Some immigrants rarely use the language, even if they lived in an L2 speaking country for several years. On the other hand, students might have the opportunity to use the language more frequently in classrooms. Moreover, Piske (2007) argues that both students and immigrants might be exposed to ungrammatical sentences and incorrect or foreign-accented pronunciation – students from their classmates and sometimes their teachers, and immigrants from other non-native speakers. Finally, both populations may or may not be exposed to explicit instructional input in L2 grammar, vocabulary, or phonology.

Furthermore, Flege (1995) reported that one major obstacle to testing the hypotheses proposed by the SLM is that it lacks an objective means to gauge how great is the degree of perceived L1/L2 phoneme similarity. This means that the metric used by bilinguals to gauge this is unclear. Best and Taylor (2007) criticise some L2 speech perception models, including the SLM, by highlighting how listeners identify non-native phones as equivalent to L1 phones, and the level(s) where this identification occurs. They claim that some models, including the SLM, implicitly involve a process that consists of inactive reception of proximal stimulus details that have no intrinsic meaning, and which are acoustic features. The statistical distribution of these meaningless proximal stimulus details in the input are eventually calculated in these models. On the other hand, PAM involves a direct-realist situation in which the process consists of the active reception of intrinsically meaningful distal event information. Moreover, Schwartz and Kaźmierski (2020) criticise current models, including the SLM, in that they do not account for all the phonetic properties of speech sounds, i.e., it is not merely the equivalence and similarities between L1 and L2 phonological segments that determine the success of L2 sound perception and production. Vowels, for instance, are characterised by other properties, rather than just the duration and position of the first formant (F1) and the second formant (F2) space. These properties include changes in the quality of duration over time, which is referred to as vowel inherent spectral change. This change, according to Schwartz and Kaźmierski (2020), makes it more difficult to decide which vowels are similar to L1 vowels and which ones are new. Thus, Schwartz and Kaźmierski (2020) believe that the SLM lacks any explanation of such changes, whereas PAM and PAM-L2 are better at explaining these changes as linking L2 perception to articulatory gestures, which in turn governs production.

Flege (1995) introduced two possibilities used by bilinguals to gauge the distance between L1 and L2 phonologies. The first one might be assessed according to the sensory properties associated with L1 and L2 sounds. The second possibility is that cross-language distance is assessed according to differences in perceived gestures. However, Flege (1995) believes that even this metric might not be easy to apply. By way of illustration, Harnsberger (2001), as mentioned above, found that cross-language similarity in neither phoneme-representation nor allophone-representation helped to predict accurate results in a task involving seven language nasal labelling. Moreover, Best and Taylor (2007) criticise the SLM in that perception models should take into account not only the phonetic level when differentiating properties of language-relevant speech, but also the phonological level in addition to the gestural level. They suggest that, according to PAM, contrasts occur at the functional linguistic level in listeners' L1 phonology and how these contrasts are related to the phonological contrasts in L2 have the same degree of importance to perceptual learning as phonetic categories in both languages, which is not the case with the SLM. According to Best (1995) and Best and Taylor (2007), these three levels, namely phonetic, phonological, and gestural levels, are all important because they influence L2 learners' discrimination ability. James (1984, as cited in Flege, 1995) states that, depending on syllable position, three different metrics can be used, which are gestural, acoustic phonetic and abstract phonology.

More recently, Tsukada, Cox, Hajek, and Hirata (2018) have criticised PAM(-L2) and the SLM in that these models do not account for the effect of cross-language distance between L2 and another unfamiliar language; and in this case, sometimes, it is not clear how L2 phonetic categories are established. Tsukada et al. (2018) suggest that further investigations should be carried out to assess the boundaries of these models in order to include or exclude the effect of cross-language distance between L2 and another unfamiliar language. Tsukada et al. (2018) carried out a study to examine the effect of L2 Japanese learners' ability to acquire contrastive consonant lengthening in L2 Japanese. In addition, they aimed to examine the cross-language transfer of this phonetic feature from L2 learning to processing an unknown language, which in this case was Italian. Furthermore, the researchers aimed to determine if non-native speakers who have this contrastive feature in their L1 rely on their L1 knowledge and consequently outperform others in the perception of Japanese consonant length, even if they do not know

Japanese. As both Japanese and Italian languages have consonant length contrast and neither Korean nor English has this contrast, they examined how the perception of this feature can be affected in Italian as a foreign language by L1 Korean and L1 Australian English learners of L2 Japanese. It is important to mention that Korean language has lengthening of tense obstruents which correspond to shorter preceding vowels, but this feature, however, is not contrastive. When comparing this feature with Japanese and Italian, Japanese, on the one hand, has longer vowels when preceding long consonants than when preceding short consonants; Italian, on the other hand, has shorter vowels before long consonants and longer vowels before short consonants. Tsukada et al.'s (2018) findings can be summarised as follows. The first finding was that L1 Japanese listeners were significantly better in their accurate perception of Italian consonant length than L1 English, but not L1 Korean. This indicates that having an L1 consonant lengthening background helps in processing consonant length in an unknown language. The second finding was that L1 Italian did not significantly differ from L1 English and L1 Korean in the perception of Japanese consonant length. This indicates that L1 experience equates to L2 learning experience. Third, L1 Korean misperceived short consonants as long more frequently than long consonants as short in Japanese, which was the opposite in Italian, indicating that consonant length was not equal in Japanese and Italian, and categorizing consonant length by the same individuals might be different from one language to another. Finally, L1 English misperceived short consonants as long more frequently than long consonants as short in both Japanese and Italian, indicating that consonant length is equal in Japanese and Italian. The perceived similarity between Korean and Italian in preceding vowel duration provided L1 Korean speakers with a cue that helped them to perform better in Italian than in Japanese when perceiving consonant length contrast. The researchers concluded that foreign language experience as well as L1 phonetic similarity may play an essential role in the effects of cross-language transfer.

2.2.5 Summary of the Section

This section has presented a detailed overview of perceptual similarity theories between L1 and L2 phonologies, and how this similarity is seen in the literature as playing an essential role in affecting L2 perception. It has elaborated on how evidence from the literature shows that L1 phonology is seen as a starting point for learning L2 phonology. This was done by conducting a review of some studies which were conducted to investigate this role that perceptual similarity plays in the development of L2 phonology. In addition, this section has highlighted the theoretical framework of L2 speech perception by discussing some basic L2 speech perception models, namely the SLM (Flege, 1995) and its revised version SLM-r (Flege & Bohn, 2021), PAM (Best, 1995), and PAM-L2 (Best & Taylor, 2007), and offering similar and dissimilar views of these models. It has also touched upon a brief view of some implications and criticisms of these models.

2.3 Influence of Phonemic Coding Ability on L2 Phonological Development

2.3.1 Introduction

Language aptitude is confirmed by a number of studies as the strongest predictor of language proficiency (e.g., Abrahamsson & Hyltenstam, 2008; DeKeyser & Koeth, 2011; Granena & Long, 2013; Li, 2016; Saito, 2019; Sun et al., 2021). Looking at the literature, it appears that language aptitude has passed through different conceptualizations in terms of its components. Despite the fact that a number of researchers disagree on what language aptitude consists of, they all agree that language aptitude is componential in nature (Artida & Muñoz, 2016), i.e., it is made up of different components. Phonological acquisition is related in one way or another to some of these components of language aptitude. By way of illustration, one of these components is phonemic coding ability.

This section presents a brief history of language aptitude as a concept. It also discusses how this concept went through different stages and how its conceptualization contributed to dividing language aptitude into different components. In addition, this section elaborates on how phonemic coding ability, as a component of language aptitude, was found to predict language proficiency. This section also touches upon types of validity in educational assessment and provides some evidence for the construct validity of some language aptitude tests including the Llama test battery (Meara, 2005), which is used in the current study to measure phonemic coding ability.

2.3.2 Concept of Language Aptitude

Dörnyei (2005) defines language aptitude as a synonym of language ability. However, Dörnyei's definition is too general according to Carroll (1993), who is the so-called "father of language aptitude" (Moskovsky, Alshahrani, Ratcheva, & Paolini, 2015) and argues that the term ‘ability’ is neutral and uninformative because it might mean either ‘aptitude’ or ‘achievement’. Thus, Carroll (1993) introduced and narrowed down the ‘aptitude’ term to “variations in individuals’ potential for present performance on a defined class of tasks” (Carroll, 1993, p. 16). However, as Carroll (1993) argues, prior learning and experience affect, to some extent, the performance in tasks, which leads to measuring this performance as achievement. The level of performance, on the other hand, is also affected by the level of constitutional maturation of the learner or by his/her central nervous system health. In addition, it might be affected by the individual’s genetic constitution (Plomin, DeFries & McClearn, 1990, as cited in Carroll, 1993). Therefore, Carroll defines the term ‘aptitude’ as the cognitive ability which predicts a certain level of future success in learning, which excludes any interest in or motivation of any particular activity. Carroll’s extensive investigation of language aptitude, which took place in 1959, is considered the first investigation in this area. Because this era witnessed the appearance of two important aptitude batteries – the Modern Language Aptitude Test (MLAT; Carroll & Sapon 1959, as cited in Carroll, 1993) and the Pimsleur Language Aptitude Battery (PLAB), developed by Pimsleur (1966, as cited in Granena, 2018) – it is known in aptitude research as the golden era of language aptitude (Moskovsky et al., 2015). According to Granena (2018), Carroll was the first to propose a methodology to examine the nature of language aptitude, which eventually led to the identification of four components of language aptitude: phonemic coding ability, grammatical sensitivity, rote learning ability, and inductive learning ability (Sasaki, 2012). Phonemic coding ability, as mentioned above, refers to learners’ ability to analyse and retain unfamiliar sounds (Sparks et al., 2011), and to make links between sound strings and their corresponding graphemic symbols (Saito, 2019). Grammatical sensitivity, on the other hand, is being aware of the native language’s grammatical structures (Carroll, 1993). Rote learning ability refers, according to Moskovsky et al. (2015), to the ability to draw and retain connections between meanings and sounds.

Finally, inductive learning ability refers to learner's ability to generalise and work out rules (Carroll, 1962).

Another conceptualization of language aptitude is presented by Skehan (1998). Skehan claims that language aptitude is divided into three main components. Skehan agrees with Carroll that language aptitude consists of phonemic coding ability, grammatical sensitivity and inductive learning ability and he adds memory as an additional language aptitude component. Skehan's memory refers to learners' ability to efficiently retrieve information to handle natural conversational needs. This definition is close to what is known nowadays as working memory (Artieda & Muñoz, 2016). However, Skehan, (1998, 2016) disagrees with Carroll on how language aptitude components are categorised. Skehan (1998) categorises language aptitude into three main components: phonemic coding ability, memory, and analytic ability, which in turn is divided into two subcomponents: grammatical sensitivity and inductive learning ability. In other words, Skehan categorises similar language abilities into one major component, language analytic ability, which includes a learner's ability to make morphological classifications (grammatical sensitivity) and grammatical inferences (inductive learning ability). This is not the case with Carroll (1962, 1993), who separates these abilities into two different language aptitude components despite their relationship. In terms of the stages of L2 acquisition, Skehan (2016) introduced different mechanisms of language aptitude. These include input processing, noticing, pattern identification, automatization and lexicalization. Siato (2019) elaborates on this mechanism in that, in the first place, learning occurs by analysing incoming input, which is followed by automatizing parts of acquired knowledge, and finally attainment use of L2 is achieved. Skehan (2016) argues that phonemic coding ability takes place in input processing, noticing, and pattern identification, as learners usually attempt to analyse unfamiliar sounds efficiently. This is why phonemic coding ability is more active in the initial stages of learning (Skehan, 2016). However, because analytic ability involves learners to focus on more complex grammatical and lexical structures, which are related to pattern identification as well as restructuring the existing system, analytic ability is more active in the mid-stage of learning. Finally, as the final mechanisms are automatization and lexicalization, they require existing knowledge, which is why they occur in the later stages of learning and involve memory and implicit learning aptitude (Saito, 2017, 2019).

Moreover, Robinson (2005) introduced the Aptitude Complexes / Ability Differential framework, which also contributed to reconceptualizing the concept of language aptitude. This framework consists of two main hypotheses: the Aptitude Complexes Hypothesis and the Ability Differentiation Hypothesis. The first hypothesis claims that higher order aptitude complexes consist of a combination of several basic cognitive abilities. These cognitive abilities include phonological working memory capacity, processing speed, and pattern recognition. Some of the abilities that these complexes include are memory for contingent speed and noticing gaps. The second hypothesis states that L2 learners are not equal in their cognitive abilities, as some are stronger than others in these abilities, leading them to have different aptitude complexes from each other. This aligns with Carroll's (1981) previously mentioned claim that learners have different levels of ability to learn or acquire a second language despite how much they are motivated. The different aptitude complexes, according to Robinson (2005), are four: focus on form aptitude, incidental learning via oral content aptitude, incidental learning via written content aptitude, and explicit rule learning aptitude. Therefore, according to Wen, Bierdroń, and Skehan (2017), Robinson's framework consists of a number of abilities, learning processes, tasks, and pragmatic/interactional abilities, which led Robinson to introduce aptitude-treatment interaction. This refers to the relationship between the importance of specific aptitudes and the context, which includes some teaching methods or techniques (DeKeyser & Koeth, 2011). This is why Robinson's classification is considered a turning point in the conceptualization of language aptitude (Wen et al., 2017). This can be seen when he suggests including a focus on form, oral content, written content and explicit rule learning as the main complexes of language aptitude because these are basic components of the learning process (Wen et al., 2017).

Furthermore, Granena (2013) suggests that language aptitude can be conceptualised as two essential dimensions: explicit language aptitude and implicit language aptitude. Explicit language aptitude involves analytic learning ability and requires cognitive aptitude. Analytic learning ability involves explicit language learning, such as tasks that include problem-solving strategies. This dimension, according to Granena, consists of components such as phonemic coding ability, grammatical sensitivity, associative memory, and inductive learning ability. Implicit language aptitude, on the other hand, is related to sequence learning ability and requires cognitive aptitude which involves implicit language learning. Sound recognition is a major component of implicit language aptitude. Granena's conceptualization of language aptitude aligns with Robinson's (2005) classification involving categorizing language aptitude into incidental and explicit learning, just as Granena suggests. Saito, Sun, and Tierney (2019) differentiate between the two types of aptitude, in that explicit learning aptitude is aptitude that is assessed by tasks that require both practice and testing phases, whereas implicit learning aptitude is assessed during tasks that have no practice phase and participants are not aware of what is being learned.

To sum up, these four conceptualizations by Carroll, Skehan, Robinson, and Granena show that there is a consensus that phonetic sensitivity (DeKeyser & Koeth, 2011; Reylonds, 2002), along with analytic ability and memory, is a major component of language aptitude paradigms (DeKeyser & Koeth, 2011). They do, however, disagree on how these components are classified. Sun et al. (2021) argue that the complex nature of naturalistic L2 learning leads scholars to stress the importance of investigating the relationship between language aptitude and L2 learning in both types of aptitude (explicit and implicit) with both types of context (classroom and naturalistic settings).

The following subsections shed light on the predictive validity of language aptitude in general and phonemic coding ability in particular, including a definition of the Llama test battery (Meara, 2005; Meara & Rogers, 2019) and what it consists of, because it was the tool used to measure phonemic coding ability in the current study.

2.3.3 Types of Validity in Educational Assessment

A number of studies in the language aptitude field have directed their attention to examine the construct validity of aptitude test batteries (e.g., Abrahamsson & Hyltenstam, 2008; Granena, 2013, 2018; Granena & Long, 2013; Li, 2016; Saito, 2017). Li (2016) defines the term "construct" as a specific measurement which may include different factors used to measure the different dimensions of a variable which is assumed to exist but cannot be directly observed. The validity of a construct refers to associating the construct conceptual definition of a particular variable with the way this variable is operationally measured or manipulated (Schwab, 1980, as cited in O'Leary-Kelly & Vokurka, 1998).

Li (2016) discusses the concept of validity in educational assessments by presenting two major divisions of this validity; first, the traditional concept of validity, consisting of a combination of content validity, criterion validity, and construct validity, and second, the current concept of validity which considers construct validity as an overarching term that refers to all three types of validity. In other words, content validity and criterion validity are merely aspects of construct validity. Content validity refers to how test items measure skills and how the tasks within a test reflect real world situations. Criterion validity refers to the correlation between test scores and criterion measures. This type of validity is divided into two types:

concurrent validity, which means to what extent a predictor variable and a criterion variable are interchangeable, and predictive validity, which refers to whether the performance of an individual in a test relates to his/her performance in another later test (Li, 2016). The latter type of validity has attracted a great deal of interest in the language aptitude domain; a number of studies have been conducted to examine the predictive nature of aptitude scores for course outcomes, but not the other way round (Carroll & Sapon, 1959, as cited in Carroll, 1993). Figure 2.3, below, shows a summary of how validity is divided according to Li (2016).

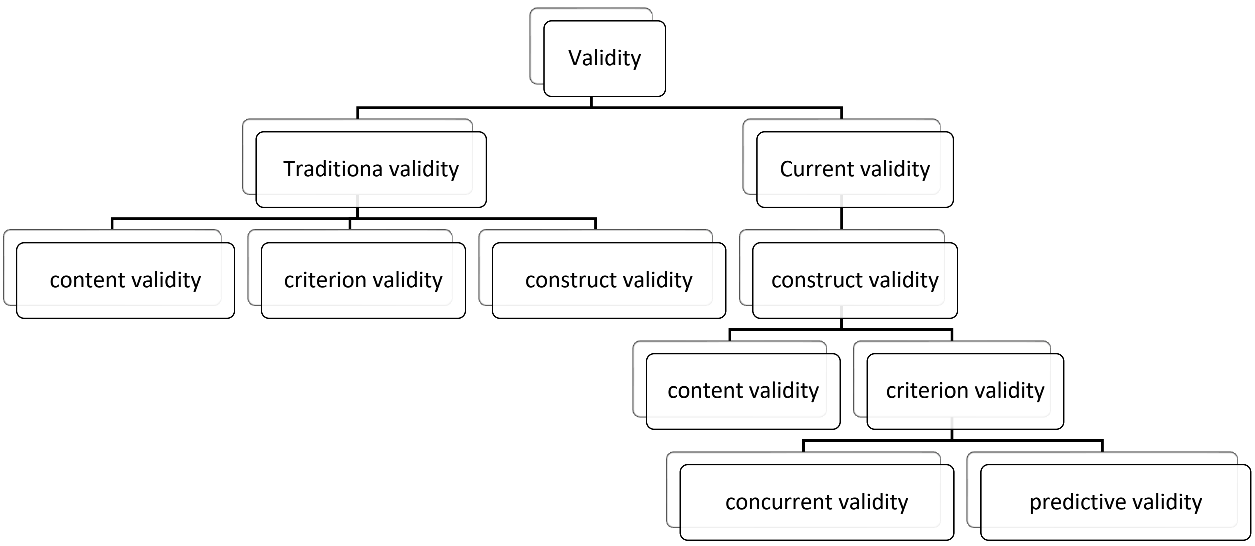


Figure 2.3 Division of validity according to Li (2016)

2.3.3.1 Phonemic Coding Ability as a Predictor of Language Proficiency.

One of the principal factors that determine the success of language learning is language aptitude (Granena, 2018; Sun et al., 2021). Most studies conducted in this area have confirmed that language aptitude is the strongest predictor of language proficiency, by using different aptitude tests, such as MLAT (Li, 2016), the Llama test battery (Abrahamsson & Hyltenstam, 2008; Granena & Long, 2013; Saito, 2019), and Hi-LAB (Linck et al., 2013 as cited in Granena, 2018). On reviewing the literature, it becomes obvious that language aptitude is seen nowadays as a multifaceted construct that reveals various aspects, both explicit and implicit abilities, as discussed in the previous subsection. This diversity has led to the development and validation of new and different tests (Saito, 2017). According to Granena (2018), most studies conducted to examine the predictability of language aptitude have used MLAT and followed the traditional concept of language aptitude as defined by Carroll (1993), while a very limited number of studies have used other tests.

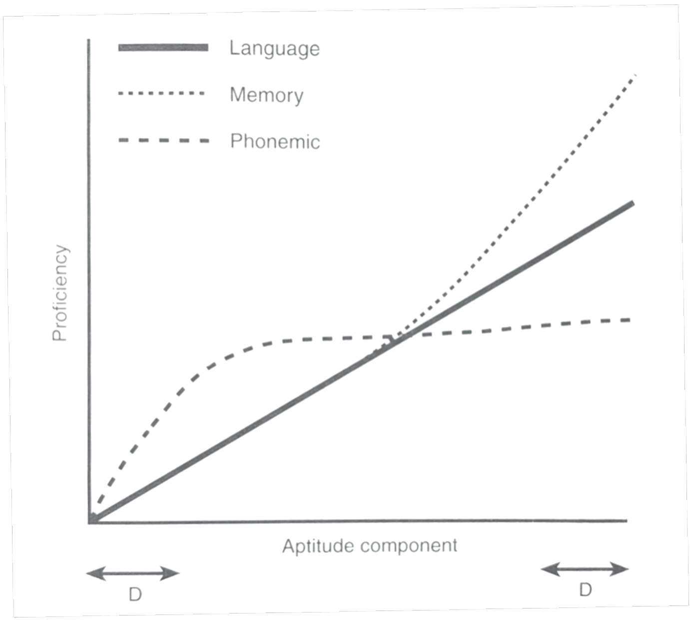
Carroll (1993) and Meara (2005) argue that learners who have high phonemic coding ability can easily structure words into smaller phonetic units. This allows them to analyse words and divide them into smaller forms, including pronunciation and morphological rules and meaning aspects. Therefore, having high phonemic coding ability is found to help L2 learners develop input processing strategies and consequentially be able to recognise and integrate new linguistic units, which in turn increases the significance of phonemic coding ability for success of language learning (Reynolds, 2002). In contrast, Carroll (1962) claims that learners who have low phonemic coding ability have difficulties in remembering phonetic form as well as in mimicking speech sounds. These claims are supported by the findings of Saito (2019), who found that the ability of L1 Japanese participants to acquire and pronounce English /ɹ/ was affected positively by the results for their phonemic coding ability and associative memory.

After correlating participants' scores in the Llama tests (refer to subsection 2.3.3.2.1 for details on the Llama test) as a measure of language aptitude with participants' performance in the pronunciation of words and sentences, as well as their fluency, vocabulary richness, and accuracy and complexity of lexicogrammar usage, Saito (2017) found that learners who have higher phonemic coding ability (measured by Llama-E) are more accurate grammatically in their speech, as having high phonemic coding ability enhances learners' capacity to retain and analyse novel sound patterns, and this in turn predicts pronunciation and grammar accuracy in their speech production. The findings of Saito (2017) are supported by a later study by the same researcher (Saito, 2019) in which the role of phonemic coding ability had a positive impact in the easy dimension of English /ɹ/ pronunciation, which is characterised by lower F2, but had no impact in the difficult dimension which is characterised by phonemic lengthening and third formant (F3) reduction. This variability in the role of phonemic coding ability led Saito (2019) to suggest that the different stages of L2 pronunciation development and different constructs of aptitude have a multifaceted relationship which supports Skehan's (2016) model of the language aptitude mechanism. These findings of Saito (2017; 2019) support the claim of Reynolds (2002) that having high phonemic coding ability helps L2 learners to develop input processing strategies. Furthermore, Saito's (2017) findings show that language analytic ability, including grammatical inferencing and sensitivity (measured by Llama-F), increases L2 learners' speed to grasp the grammatical information encoded in words. This helps learners to know how sentences and phrases are formed by smaller words. Lastly, Saito (2017) found that having a large rote and associative memory (measured by Llama-B) enhances learners' ability to express multiple phrases at a faster speed and greater fluency using larger amounts of lexical information. Granena (2013) tested 186 participants, aiming to validate and assess the reliability of the Llama test, and concluded that, in naturalistic learning settings, phonemic coding ability predicts L2 pronunciation and grammatical accuracy. Artieda and Muñoz (2016) support this finding and extend it to include formal learning settings after examining 140 participants who were allocated into two different proficiency groups and compared depending on their aptitude scores. Their results showed that there is a statistically significant and equal relationship between language proficiency and language aptitude at two different proficiency levels. However, after breaking aptitude down into multiple components (including phonemic coding ability, rote learning ability, and language analytic ability) the impact of language aptitude on their performance was different for two levels of proficiency. This difference occurs because phonemic coding ability has a greater impact on learners who are in the early stages of L2 proficiency, whereas language analytic ability has an impact at all levels but with a more robust impact on intermediate learners. Artieda and Muñoz's (2016) findings support the previously mentioned mechanisms of language aptitude proposed by Skehan (2016) (see Subsection 2.3.2). Moreover, Saito (2017) argues that a high level of phonemic coding ability assists adult L2 learners to have the ability to remember and analyse unfamiliar sounds and rely on this to predict the pronunciation of words that have refined phonological and morphological accuracy. This may also be applicable to reading, as learners who have low phonemic coding ability struggle with decoding the phonemes in a reading text (Reynolds, 2002). This eventually distracts them away from comprehending reading text as they are so focused on the decoding process, which may lead them to ignore the meaning of the context (Pikulski & Chard 2005). As a result, this could draw their attention to the words themselves instead of the meaning of the text, leading to insufficient comprehension (Ehri 2005).

Sparks et al.'s (2011) discussion of phonemic coding ability aligns with Saito's (2017) argument. They highlight the importance of phonemic coding ability and phonological awareness in learners' ability to decode words. Phonological awareness refers to being conscious that words are divided into smaller phonological units which have no semantic value, and this consciousness enables learners to manipulate these phonological units (Reynolds, 2002). Sparks et al. (2011) claim that weak phonemic coding ability and/or phonological awareness leads to the inability to discriminate sounds and decode words efficiently and rapidly. This was evidenced after conducting a factor analysis of a test that included L1 skills, academic aptitude, L2 aptitude, and effective measures (anxiety and motivation) to measure the predictivity of L2 proficiency. Their conclusion was that learners are likely to have low proficiency in written and oral L2 learning aspects if they have weak phonological skills. In contrast, L1 phonemic coding ability and phonological processing skills can be the best predictors for L2 spelling and word decoding skills. Research shows that reading skills (including decoding and comprehension) largely depend on developing decoding skills in addition to cognitive and linguistic abilities (Ibrahim 2011). Geva and Zadeh (2006), for instance, state that highly efficient readers usually have better oral language proficiency and score higher in phonological awareness tasks than those who have lower proficiency.

This review shows that phonemic coding ability plays an important role in predicting L2 proficiency. However, the literature reveals that the relationship between phonemic coding and L2 proficiency occurs in different language aspects. These include accurate pronunciation and morphology (Artieda & Muñoz, 2016; Granena, 2013; Saito, 2017, 2019, Saito et al., 2019), decoding skills, reading and writing (Sparks et al., 2011), and vocabulary learning (Li, 2016). However, Li (2016) concludes that the role of phonemic coding ability in predicting listening comprehension and speaking proficiency remains weak. Li points to the factor of listening, in which learners tend to focus on meaning rather than form, as a reason for the weak role of phonemic coding ability because it focuses on form rather than meaning, as it asks learners to relate the written form to its corresponding sound. Speaking, on the other hand, involves sound production rather than sound comprehension, which shows that it is not related to phonemic coding ability. Li (2016) concludes that L2 writing cannot be predicted by phonemic coding ability because the aptitude tests involved in recent studies that have been conducted do not provide measures that evaluate the different skills involved in writing. Moreover, Saito (2017) reported that vocabulary appropriateness is not associated with phonemic coding ability.

Notwithstanding the evidence for phonemic coding ability as a predictor of L2 proficiency, Skehan (1998, 2016) claims that it only plays a role in the early stages of learning, as it starts to decrease when it reaches a particular level. This is because learners face difficulties in transferring acoustic input and processing it into language chunks, which causes phonemic coding ability to be seen as a threshold because phonemic coding ability makes a minimal contribution to language proficiency when it is at a higher level. Skehan (1998) built his argument based on case study findings, in which he reported that exceptional learners did not have any exceptional level in either phonemic coding ability or language analytic ability. In addition, he presented a diagram that shows that phonemic coding ability has the highest curve in the early stages of learning and eventually starts to decrease, showing that as language improves, its role decreases (see Fig. 2.4). According to Skehan (1998), this is not the case with other aptitude components. The importance of language analytic ability, for instance, is equal at all language proficiency levels. Also, the importance of memory is enhanced after a learner reaches an advanced level of proficiency and then becomes a fundamental factor in helping learners to achieve native-like proficiency. Saito (2019) and Artida and Muñoz (2016) supported Skehan's claim after investigating the relationship between language aptitude and L2 proficiency at two language proficiency levels (beginners and intermediate). Artida and Muñoz (2016) found that there was a correlational relationship between language proficiency and language aptitude at the two levels investigated. However, this finding was only applicable to beginner learners for whom phonemic coding ability had the largest impact on learning, while grammatical inferencing was the ability most affected in the intermediate group. After conducting a meta-analysis of language aptitude that included various studies conducted to investigate different aptitude tests which took place in the last five decades, Li (2016) also supported Skehan's (2016) argument by finding that phonemic coding ability offers the strongest predictability of accurate L2 speech production for learners who have little language experience.



Language = language analytic ability
Phonemic = phonemic coding ability
D = unusual neurological conditions

Figure 2.4 Relationship between aptitude components and proficiency level (Skehan, 1998, p.217)

2.3.3.2 Construct Validity of Language Aptitude Tests

As this thesis makes use of one subtest of the Llama test battery, this section concentrates on the Llama test battery in particular and touches upon a brief comparison of this test with other well-known tests. This section starts with a short introduction about the Llama test battery and what it consists of.

2.3.3.2.1 What is the Llama Test Battery and What Does it Consist of?

The Llama test battery (Meara, 2005) is a test that started as a project for MA students' research at the University of Swansea. It is a revised version of the LAT aptitude test introduced by Meara, Milton, and Lorenzo-Dus (2003, as cited in Granena, 2013). This test battery was designed based on the components of language aptitude introduced by Carroll and Sapon's MLAT (1959, as cited in Carrol, 1993), with the idea of digitising it to make it easier to access

(Rogers, Meara, Branett-Legh, Curry, & Davie, 2017). The components introduced by Carroll (1993) are rote memory, inductive learning ability, grammatical sensitivity and phonemic coding ability (Sasaki, 2012). The Llama test battery consists of four different subtests, each one dedicated to measuring one of the aptitude subcomponents. The subtests are known as Llama-B, Llama-D, Llama-E and Llama-F (Granena, 2018).

Llama-B: This subtest is designed to evaluate associative memory, defined by Moskovsky et al. (2015) and Saito (2019) as the ability to associate sounds with meanings and to retain these associations in memory. This part of the battery tests the ability to learn new words (See Appendix A, which shows a screenshot of a Llama-B test).

Llama-D: This test is not based on MLAT but rather on the assumption that a key factor in mastering a language is the ability to recognise repeated sounds in speech and eventually being able to isolate those sounds within words, leading to the acquisition of morphological competence (Granena, 2018). It is seen as a means of measuring implicit learning (Rogers et al., 2017), which is why there is no learning phase in this part of the battery. Meara (2005) claims that being able to recognise sequences of sounds helps in making the learning process easier and faster (See Appendix B which shows a screenshot of a Llama-D test).

Llama-E: Llama-E is based on the MLAT phonetic script subtest which examines participants' ability to associate sounds with their corresponding symbols, that is, their phonemic coding ability (Granena, 2018; Saito, 2019) (See Appendix C which shows a screenshot of a Llama-E test). A further explanation of how this subtest is run is presented in subsection 3.3.5.2.

Llama-F: This part of the test is designed to assess learners' ability to infer and work out the grammatical rules of an unknown language. In other words, it examines analytical ability combining both inductive learning ability and grammatical sensitivity (Skehan, 2002) (See Appendix D which shows a screenshot of a Llama-F test.).

2.3.3.2.2 Criterion Validity – Predictive Validity of the Language Aptitude Tests

DeKeyser and Koeth (2011) criticised the old aptitude tests, including MLAT, PLAB and the Defense Language Institute Battery (DLAB), as lacking construct validity and concentrating only on the predictive validity of language aptitude. Furthermore, these three traditional aptitude tests do not address the theoretical constructs that psychologists are now able to provide. DeKeyser and Koeth also claimed that current alternative tests, like the CANAL-F and the Llama, still lack higher predictive validity. However, one major difference between the Llama test battery and other language aptitude tests is that most of these tests were designed for specific purposes and in specific languages, which enhances their intended predictive validity, while the Llama test battery is independent of participants' first language and not directed at a specific audience. For instance, Carroll developed the MLAT after the Second World War to help the US army train foreign language learners and reach a satisfactory level easily and in a short time (Granena, 2018). On the other hand, DLAB, introduced by Petersen and Al-Haik (1976, as cited in Rogers et al., 2017), was developed after the MLAT; its aim was to determine in what language or language family a person would perform better. It focuses on syntactic rules and the ability to learn them systematically. Similar to MLAT, PLAB was an alternative to MLAT to direct young learners by examining their knowledge of vocabulary in English. A deficiency of these tests is that they neglected construct validity and concentrated only on predictive validity (Granena, 2018). Research on these three tests has shown the powerful predictive validity of aptitude as an important predictor of foreign language proficiency. It is reported that they yielded correlations in the range of 0.4 to 0.6 (Granena, 2018).

Granena (2013) argues that language neutrality is very important in cognitive measures because it helps to avoid the potentially confounding content inherent in tests that depend on participants' L1 or L2, such as language proficiency, linguistic background, literacy level and many others. In addition, it may help test administrators conduct the test without the need for translation, which may influence the reliability and validity of the test. In an attempt to measure the language neutrality of the Llama test battery, Rogers et al. (2017) hypothesised in their study that participants' first language would affect their performance on the Llama test battery. Based on previous studies (Gholamain & Geva, 1999; Green & Meara, 1987; Hamada & Koda, 2008 as all cited in Rogers et al. 2017) holding that the degree of distance between L1 and L2 has an important impact on word processing and retention, Rogers et al. (2017) believed that L1 English participants would outperform their L1 Arabic and Chinese counterparts, as English participants use Roman script and the test uses Roman script as well. They further expected that L1 Arabic participants would outperform Chinese participants, as Arabic is a consonant alphabetic language. However, the findings of their study did not support these hypotheses, as all three groups performed equally well, regardless of their first language. They concluded that the Llama test battery is language-neutral, thus providing strong evidence for its predictive validity. Granena (2013) attempted to validate the predictive nature of the Llama test battery by examining 11 studies that had used it. She found that there was a correlation between scores on the Llama test battery and the use of analytic and metalinguistic abilities in an explicitly instructional second language learning environment, emphasising that only Llama-D (sound recognition test) was related to implicit language learning and processing. Artieda and Muñoz (2016) extended these findings by Granena (2013) to include a formal context effect, in addition to a naturalistic context effect. They found that beginner learners performed more accurately than their intermediate counterparts. In addition, Granena's (2018) findings support the prediction made by Skehan (2016), by discussing the predictive nature of Llama-D in second language speaking proficiency, which was measured by analysing a factor labelled "Implicit Memory Ability". She found that learners with greater implicit memory ability are more likely to recall foreign language words and relate their meanings, highlighting the role of the interaction between implicit memory and implicit language ability in predicting lexical complexity. In addition, Li (2016), in his investigation of the construct validity of aptitude batteries, including the Llama test battery, concluded that there is a strong association between aptitude as a construct and second language proficiency in general, depending on the participants' course grades or their scores on standardised tests, such as TOEFL. He also highlighted the greater effect of language aptitude at the beginner level than that shown at an advanced level. He attributed that difference to the possible sensitivity to language aptitude in the initial stages more than in the later stages of second language proficiency. Li (2016) presents a detailed conclusion in his paper of how language aptitude predicts second language proficiency, in which he claims the following: language analytic ability, as measured by different tests, among them the Llama test battery, is a good predictor of learning grammar. On the other hand, phonetic coding is not enough to predict speaking and listening comprehension. Finally, rote memory has the least predictive nature among all the aptitude components. However, it is important to acknowledge that most of the studies investigated by Li (2016) used MLAT test, with only a few exceptions that used the Llama and other aptitude tests, notwithstanding that the Llama test battery is designed in accordance with MLAT's theoretical framework (Granena, 2018, Rogers, et al., 2017, & Li, 2016; Saito, 2019).

Rogers et al. (2017) criticise the Llama test battery in that it might be affected by prior experience or training, although they acknowledge that it is still not proven that language aptitude is trainable. This is supported by Artieda and Muñoz (2016), who also argue that the Llama test battery is not designed to measure advanced second language learners' cognitive ability. They acknowledge that if their study, which was designed to measure beginner and intermediate level performance, had been conducted on learners at a more advanced level, the effect of the aptitude components might be different.

2.3.3.2.3 Content Validity of the Language Aptitude Tests

The aforementioned studies in this section shed light on the content validity of the Llama test battery by providing some evidence relating items in the Llama test battery to the skills being measured. In the first instance, Rogers et al. (2017) criticise MLAT for being designed on the basis of an assumption that test-takers, supposedly native English speakers, are unfamiliar with the Kurdish language used in the test. They state that, because the Llama test battery is designed to be independent of the participant's first language, they add additional variables. For example, Llama-B provides test-takers with non-real-object pictures, thus increasing the possibility of testing what is intended to be tested, regardless of the participant's first language. Similarly, Artieda and Muñoz (2016) discuss this issue and they clarify to what extent Llama-F supports the concept of first language independence by presenting visual stimuli to the participants. Moreover, Rogers et al. (2017) point to the fact that Llama-B and Llama-E are presented using Roman script, raising the question of whether the first language script of the test-taker affects his/her performance on the test as, according to them, many researchers have found that first language script may affect second language acquisition. As mentioned above, they concluded that the Llama test battery is language-neutral, which can be seen as good evidence of its content validity.

Furthermore, Granena (2013) hypothesises that there are two aptitude subcomponents which are linguistic in nature measured by the Llama test battery. The first subcomponent is analytical learning ability, which is measured by three subtests (Llama-B, E, and F). These subtests involve using strategies and problem-solving techniques and include a learning phase. The ability being measured by these three subtests is developed in the learners' L1 linguistic knowledge and requires a degree of analysis in order for learning to occur. In other words, it involves explicit language aptitude. The second subcomponent is sequence learning ability and is measured by Llama-D, which does not have a learning phase. This subtest involves discovering language structures through statistical characteristics of the input, which is seen by Granena (2013) as an implicit cognitive process. Granena's hypothesis is supported by a series of later studies. For instance, Granena (2018) found that learners perform differently in the Llama tests that involve their explicit aptitude, and in serial reaction time tests that reflect their implicit aptitude. Granena (2018) states that two types of aptitude are correlated, respectively, with the rationale-analytical, which represents explicit learning, and the experiential-intuitive, which is a sign of implicit learning. Meara (2005), the founder of the Llama test battery, supports this idea by highlighting the role of the Llama-D test in that it helps learners to discriminate morphological distinctions implicitly, as it does not have a learning phase and involves automatic processing.

Artieda and Muñoz (2016) criticise the content validity of the Llama test, in that Llama-B, which is designed to measure memory, does not have any effect on the other dimensions of language, despite the fact that memory is, according to Robinson (2005), one of the cognitive abilities that plays a major role for beginner learners. However, they acknowledge that it is not claimed that the Llama test battery is designed to measure the full range of cognitive abilities of advanced second language learners. They found that level of proficiency had an impact on their intermediate learners' performance in three tests, Llama-B, E, and F, which was not the case with advanced learners. This led them to suggest that using measures other than the Llama test battery may reveal different aptitude components.

In terms of test reliability, based on two different indexes that measure the internal consistency and stability of the Llama test, Granena (2018) argues that all the four Llama subtests have acceptable to good reliability. However, Granena (2018) attributed having test-retest reliability below .70 in her study to the long interval between the two test administrations, which was two years. Moreover, Granena points out the poor reliability (according to George and Mallery, 2003, as cited in Granena, 2018) recorded for the Llama-D test (which was .50) to the nature of Llama-D measurement, as it is designed to measure implicit learning. She states that implicit learning tasks have weaker reliability than explicit learning tasks. Thus, she suggests that researchers should be cautious when generalising from the results of Llama-D.

2.3.4 Summary of the Section

This section gave a brief presentation of how language aptitude has gone through different processes of conceptualization. It also presented a detailed description of phonemic coding ability as a subcomponent of language aptitude. It elaborated on how this subcomponent predicts second language proficiency in many different aspects including structuring language into smaller phonetic forms (Carroll, 1993; Meara, 2005), accuracy in grammar (Saito, 2017), input processing strategies (Reynolds, 2002), L2 pronunciation (Artieda & Muñoz, 2016; Granena, 2013; Saito, 2017, 2019; Saito et al., 2019), decoding skills (Sparks et al., 2011) and vocabulary (Li, 2016). This section also demonstrated how language aptitude were designed to measure different language aptitude aspects, in addition to how their validity was evaluated in the literature.

2.4 Influence of Orthography on L2 Phonological Development

2.4.1 Introduction

Although it has long been acknowledged by L2 teachers that orthography plays an essential role in developing phonology, systematic empirical research in this area is still a relatively recent enterprise (Hayes-Harb & Barrios, 2021). Even though a number of leading studies were conducted in the 1990s to investigate the role of orthography in L2 phonological development (e.g., Massaro, Cohen, & Thompson, 1990; Young-Scholten, 2002), earnest investigation of this issue has only been seen in the last two decades (Bassetti Escudero, & Hayes-Harb, 2015). In the 1990s, Massaro et al. (1990) conducted an important study to compare the performance of participants in the case of audiovisual speech input and orthographic input. They found that the group who had orthographic input performed better than the group who had audiovisual input. However, despite these studies, Bassetti et al. (2015) demonstrate that the first scientific meeting in which this topic was discussed was EuroSLA, 2013[1], and before that there was a lack of edited collections of papers or monographs in this area. Bassetti et al. attribute this to many possible reasons. Among these is the lack of any theoretical justification which is due to the common interest of linguistic research at that time; linguistic research was dominated by the universals of language, where spoken language had primacy and the communicative approach was dominating language teaching. That is why the effect of orthography on L2

[1] EuroSLA stands for the European Second Language Association which is an annual conference that presents studies on second language research which are multi-lingual and cross-cultural based on wide-ranging theoretical perspectives.

phonological development was seen as "irrelevant and inconsequential" (Bassetti, et al., 2015, p. 2).

This section discusses the importance of the effect of orthography on L2 phonological development in the current literature and how some researchers encourage L2 teachers to provide orthographic input during phonological development while others encourage teachers to avoid it. In addition, this section presents the different factors that have been investigated in this area, with a detailed review of each factor presented separately.

2.4.2 Importance of the Topic

Bassetti et al. (2015) emphasise the importance of the effect of orthography on L2 phonological development. They argue that this issue should be acknowledged nowadays and not ignored anymore. This is because it can have an effect on different aspects of L2 phonological development, such as production and perception, as well as L2 word form learning (Bassetti, et al., 2018; Cerni, Bassetti, & Masterson, 2019; Escudero et al., 2008; Showalter & Hayes-Harb, 2013, 2015). Moreover, the effect of orthography is also expanded in a number of different studies to include having an influence on language learning (e.g., Bassetti, 2017) and language teaching (e.g., Bassetti & Atkinson, 2015; Escudero, 2015; Escudero et al., 2008), and to have theoretical implications (e.g., Bassetti, 2017; Bassetti et al., 2018; Showalter & Hayes-Harb, 2013, 2015). By way of illustration, Nimz and Khattab (2020) argue that it is assumed by L2 pronunciation researchers and teachers that orthographic cues provide additional help to L2 German learners in establishing new phonetic categories for vowels in German, which consequently results in having better and more native-like pronunciation. However, Nimz and Khattab believe that this issue still needs further investigation. Furthermore, the findings of Bassetti (2017) and Bassetti et al. (2020) show that orthographic representations resulted in L1 Italian learners of L2 English establishing two phonological categories for some consonants, both of which were different from the native English category based on the existence of consonant digraphs.

These findings raise the demand for more systematic evidence that deals with orthography as an effective factor in L2 phonology as it is needed to enhance L2 phonological development models (Hayes-Harb & Barrios, 2021). Bassetti (2017), depending on her findings, suggests that the current dominant models of L2 phonological development, such as the SLM (Flege, 1995) and PAM-L2 (Best & Taylor, 2007), can be associated with orthographic input during L2 phonological development as it is found that L2 speakers produce phonological contrasts depending on the orthographic representation. Nimz and Khattab (2020) support this, highlighting that none of these models demonstrate the role played by orthography, even though its role is acknowledged in teaching pronunciation and L2 speech research. In addition, Bassetti's (2017) findings suggest expanding L2 speech perception models to include situations where L1 two categories are mapped to only one corresponding L2 category. This will eventually help in providing theoretical frameworks in this area (Bassetti, 2017; Bassetti et al., 2015; Nimz & Khattab, 2020). Looking at the literature (e.g., Bassetti, et al., 2018; Erdener & Burnham, 2005; Escudero, 2015; Escudero et al., 2008; Hayes-Harb, Nicol, & Barker, 2010; Showalter & Hayes-Harb, 2013, 2015), the results and conclusions drawn from studies that have investigated the role of orthography in L2 phonological development are contrasted in a way that makes them contradict each other (these results are presented later in this section with further details). Therefore, there is a need for systematic data that answer questions like: In what circumstances does orthography help L2 phonological development or hinder it? What are the factors that play a role in this effect? (Hayes-Harb & Barrios, 2021).

2.4.3 Reasons for the Different Findings Reported in the Literature on the Effect of Orthography on L2 Phonological Development

As mentioned above, numerous studies have investigated the role of orthography availability in L2 phonological development in the last two decades (e.g., Bassetti, 2006; Bassetti, 2017; Bassetti et al., 2018; Bassetti et al., 2020; Bassetti et al., 2022; Cerni et al., 2019; Erdener & Burnham, 2005; Escudero, 2015; Escudero et al., 2008; Escudero, Simon, & Mulak, 2014; Hayes-Harb et al., 2010; Nimz & Khattab, 2020; Showalter & Hayes-Harb, 2013, 2015). The findings of these studies, however, are not consistent with each other. Bassetti et al. (2015) demonstrate that some studies have found that orthography helps in facilitating speech production, perception or in some cases word form learning (Cerni et al., 2019; Escudero et al., 2008; Showalter & Hayes-Harb, 2013), while others have concluded that it may prevent phonological acquisition from reaching a target-like level (Bassetti, 2007; Bassetti et al., 2020; Bassetti et al., 2022; Hayes-Harb et al., 2010), and some have found that it had no effect or a mixed effect (Cerni et al., 2019; Escudero & Wanrooij, 2010; Nimz & Khattab, 2020; Simon, Chambless, & Kickhöfel Alves, 2010; Showalter & Hayes Harb, 2015). These findings vary as they touch upon different factors that are believed to play a role in the magnitude of the effect that orthography has on acquiring L2 phonology. For instance, some researchers have investigated the role of the orthographic depth of L1, L2 or both (Bassetti, 2008; Bassett et al., 2020; Bassetti, et al. 2022; Erdener & Burnham, 2005). Orthographic depth refers to the extent to which the orthographic system allows one-to-one phoneme-to-grapheme correspondences. Therefore, languages can be seen as varying from transparent-to-opaque on a continuum of orthographic depth (Erdener & Burnham, 2005) (more explanations and examples are provided below in Subsection 2.5.4.1). Other studies have focused their investigation on the extent to which L1 and L2 are congruent, which is defined by Showalter and Hayes-Harb (2015) as the case where L1 and L2 have the same graphemic symbols that correspond to the same phonemes (Bassetti, 2006; Escudero et al., 2008; Escudero et al., 2014; Hayes-Harb et al., 2010; Nimz & Khattab, 2020) (more explanations and examples are provided below in Subsection 2.5.4.2). A number of studies focus their investigation on the degree of familiarity of the script, some provide a totally novel script (e.g., Alhumaid, 2019), partially novel script (e.g., Showalter & Hayes-Harb 2013), a familiar script (e.g., Bessetti & Atkinson, 2015; Erdener & Burnham, 2005; Escudero et al., 2008; Simon et al., 2010;) or both partial and total novel scripts (e.g., Mathieu, 2016; Showalter & Hayes-Harb, 2015).

Some researchers have attempted to identify the predictors of the effect of orthography on L2 phonological development. Some studies, for instance, touched upon different factors of the orthographic systems that may modulate the effect of orthographic input. These include systematicity, which refers to whether or not the phonological contrast is represented in orthography in a systematic basis, familiarity, congruency, and perceptibility (Hayes-Harb & Barrios, 2021). By way of illustration, Bassetti et al. (2020) investigated the effect of the number of letters on the pronunciation of L2 English words by L1 Italian learners. The researchers also shed light on some factors that might have an impact on the magnitude of the orthographic effect. These include, but are not limited to, external factors such as how old the participant is when exposed to L2, how proficient he/she is in English, how long he/she has stayed in an L2 environment, and how much written input out of total input he/she is exposed to. In addition, the study also included internal factors, such as how important it is to have native-like pronunciation for the participant, how motivated he/she is, as well as his/her phonological short-term memory and mimicry ability for dialects and foreign languages. Researchers have concluded that the magnitude of the orthographic effect is positively influenced by (1) English language proficiency in the case of consonants, (2) the desire to learn English in the case of vowels, (3) the type of exposure (naturalistic or instructional) in which naturalistic exposure reduces the effect of orthography, and (4) short-term memory. However, a negative effect was found for the proportion of written input out of total input in the case of naturalistic exposure. Bassetti (2008) also argues that there is a possibility that the effects of L2 orthographic input on acoustic input are influenced by: (1) L1 grapheme-phoneme conversion rules, (2) the mental representation of L2 phonology which is affected by L2 phonological input and orthography input, or (3) merely L1 phonology. Bassetti et al. (2018), investigating L1 Italian learners of L2 English, also support this by relating the effect of orthography to L1 phonology, L2 orthography, and L1 orthography-phonology correspondences. The researchers built their argument based on their finding that the orthographic effect on the vowel duration of L1 Italian learners of L2 English was weaker than the effect on consonant duration. This was attributed to the fact that vowel duration is not phonemic in the participants' L1, in addition to the fewer correspondences of vowel length with the number of letters in English than those corresponding to consonant length in Italian. Showalter and Hayes-Harb (2015) demonstrate two possible reasons for having mixed findings reported in the literature for the effect of orthography on L2 phonological development. The first reason is the different levels of perceptual difficulty of target auditory forms. For instance, they assume that the findings of Simon et al. (2010), who found that orthography had no role in their experiment, were due to the novel vowel contrasts that exist in L2 but not in L1, because participants found these contrasts difficult to discriminate and therefore orthography was found to be unhelpful. Moreover, Showalter and Hayes-Harb (2015), based on their study conducted on L1 English naïve listeners of Arabic, also attribute their own finding that there was no effect of Arabic orthographic input on L1 English learners to the fact that the auditory contrast (/k/ and /q/) is too difficult for L1 English learners of L2 Arabic, as evidenced by Almahmoud (2013). Escudero et al. (2008) argue that the lack of L2 lexical representations of similar-sounding words results in having difficulty in distinguishing these words and building them up phonologically, and this eventually leads learners to deal with these words as homophones.

The second reason for the diversity of findings demonstrated by Showalter and Hayes-Harb (2015) is the level of congruency between L1 and L2 orthographic systems, which means the relationship between relevant grapheme-phoneme correspondence. This aspect is discussed in more detail in Subsection 2.5.4.2. Escudero (2015), on the other hand, argues that orthography can only help if there is a congruency between L1 and L2 grapheme-phoneme correspondences. She supports her claim with the finding on her naïve listeners of Dutch that orthography was only helpful for two perceptually difficult pairs out of 66 pairs divided into perceptually difficult minimal pairs (e.g., "pag" and "paag"), perceptually easy minimal pairs (e.g., "pag" and "pieg"), and non-minimal pairs (e.g., "beeptoe" and "pag"). This division was based on the performance of Spanish listeners' vowel discrimination in Escudero and Wanrooij (2010), in addition to the possibility that the perception of Dutch vowel contrasts is done through mapping these contrasts to a single Spanish vowel (perceptually difficult) or to two different Spanish vowels (perceptually easy).

Bassetti and Atkinson (2015) highlight the role of task type in the impact of orthography on L2 phonological development. They built their argument on their finding after they provided their participants, who were L1 Italian learners of L2 English, with two different tasks: reading aloud and word repetition. They found that the performance of their participants was affected better by removing orthographic input and providing an immediate native model to imitate. Their findings suggest that by depending only on a reading-aloud task, orthographic effects may be overestimated, whereas by depending only in a word repetition task, orthographic effects may be underestimated. Similarly, Escudero (2015) argues that having more options to choose from, such as the case when several native phonemes are perceived as one single L2 phoneme, might have led to inaccuracy in sound perception. According to Escudero, having more than one option to choose from might be the reason for her unexpected finding that Spanish-speaking listeners' performance was better than Australian English-speaking listeners' performance in the perception of vowels, even though English speakers have a larger vowel inventory than that in Spanish.

Frost, Repp and Katz (1998) argue that if an utterance is accompanied by its printed version, it is perceived much more clearly. They attribute this to the fact that the written form of an utterance helps in decoding it into an internal speech-like representation, so that an internal phonetic structure is built by linking the printed word to its auditory input. However, opponents of providing learners with orthography when acquiring phonology disagree with this, claiming that orthography leads to establishing a discontinuity between lexical (words) and prelexical (phonological) representations (Cutler, 2015), which is the case when learners are able to distinguish words in the lexicon without the ability to notice the phonological contrasts in these words (Escudero, 2015). Cutler (2015) argues that the main goal of providing orthography is to help learners to distinguish phonological contrasts, especially those that are believed to be difficult to perceive such as those that have only one native phonemic category. However, according to Cutler, orthography is supposed to help build a lexical representation, but eventually it fails to help learners to perceive non-native phonemic distinctions, which results in these words being harder to recognise perceptually and then, in this case, orthography offers more disadvantages than benefits. For example, Escudero et al. (2008) found that, although their Dutch-English bilingual participants were able to learn new English words without being exposed to orthographic input, they failed to use the contrasts that occurred in some familiar words that they had already learned, such as /pændɑ/ or /pɛnsl/, to distinguish the first syllable of unfamiliar English words that have phonemes in the same syllable position, such as /tændək/ and /tɛnzə/. Escudero (2015) agrees with Cutler (2015) in that orthography may not help accuracy and it is better to avoid it because its role is restricted to only when it has a congruency in the grapheme-phoneme correspondences between L1 and L2 orthographic systems.

By highlighting the role of orthographic depth and linguistic background after examining two different linguistic groups, L1 Turkish and L1 English naïve listeners of Spanish and Irish languages, Erdener and Burnham (2005) agree under certain circumstances with Cutler and Escudero's suggestion of the benefits of avoiding orthography. The Turkish and Spanish languages have phonologically transparent orthographic systems, whereas English and Irish languages have phonological opaque orthographic systems. They found that the performance of the Turkish participants in their learning of Spanish was affected positively by the availability of written forms in Spanish words while it was affected negatively by the availability of Irish written forms in their learning of Irish. On the other hand, English learners of Spanish were affected negatively by written Spanish words and positively by written Irish words. Erdener and Burnham (2005) concluded that if learners have prior experience of an opaque orthographic system in their L1 or the target language has an opaque orthographic system with learners having a transparent orthographic system in their L1, orthography should be avoided, especially in the initial stages of foreign language exposure. On the other hand, they argue that when the target language has a phonologically transparent orthographic system, orthography is more helpful, especially for pronunciation. The role of orthographic depth (transparent vs opaque) is discussed in more detail in the following subsection.

2.4.4 Different Factors Investigated in this Area

2.4.4.1 Level of Orthographic Depth of L1 and L2

The perception of native and non-native contrasts is found to be impacted upon greatly by the extent to which the learner is experienced with phoneme-to-grapheme conversion rules (Erdener & Burnham, 2005). Bassetti (2005) states that the type of language writing system can play an essential role in segmenting the language into different phonological units. Bassetti et al. (2022) emphasize on the role of how L2 sounds and words are represented orthographically in that it may prevent native-like pronunciation. This is because L2 speakers used to recode L2 orthographic forms based on the orthography-phonology conversion rules of their L1. Bassetti et al. (2022) point out that the orthographic effects on L2 phonology is not even influenced by lengthy naturalistic exposures because studies showed that bilinguals continue to produce, perceive, and judge L2 words depending on how these words were spelled in spite of how long they stay in an L2-speaking environment. As mentioned above, Erdener and Burnham (2005) claim that orthography helps in developing L2 phonology only if learners have prior experience of a phonologically transparent L1 orthographic system. On the other hand, if learners have prior experience of a phonologically opaque L1 orthographic system or they are learning an L2 with a phonologically opaque orthographic system, then orthography is found to have a negative effect. Bassetti (2008) and Simon et al. (2006) agree with that, claiming that learners tend to rely more on orthography if they have a phonologically transparent orthographic system in their L1. Bassetti (2008) extended this to the situation where the L2 orthographic system is also transparent. In this case, learners are found to rely on orthography more than those who are learning L2 with an opaque orthographic system. Therefore, a considerable number of studies have directed their investigations onto the role of phonological transparency level in the effect of orthographic input on L2 phonological development.

Based on the orthographic depth continuum, Erdener and Burnham (2005) argue that the ideal degree of orthographic depth is when there is one single phoneme corresponding to one single grapheme, as for example in the Turkish and Spanish languages. Alhumaid (2019) provides an example of how Turkish and English differ in their orthographic depth. For example, the cognate word *cancer* has two different orthographic representations in Turkish and English. In Turkish, the word *kanser* has one grapheme that corresponds to the phoneme

/k/ and another grapheme that corresponds to the phoneme /s/, which is not the case in English as it has the same grapheme <c> corresponding to both /k/ and /s/.

Based on the claims of how transparency level affects the level of learners' reliance on orthography, Bassetti and Atkinson (2015) hypothesised that the homophones in their study, e.g., *son* and *sun,* would be pronounced differently by their Italian participants as L1 Italian speakers have been trained to rely on their transparent L1 orthographic system. This hypothesis, according to Bassetti and Atkinson, was based on two different attributions. First, when two different words are spelt differently in Italian, they are pronounced differently. Second, it is possible that Italian participants will depend on the rules or conventions of other words to pronounce target words. They might, for example, rely on how the words *run* and *gun* are pronounced in order to predict how to pronounce the word *sun*, and they might pronounce the word *son* differently as they know that the grapheme <o> is pronounced differently in other words. Bassetti and Atkinson (2015) hypothesised that this case might occur only with homophones but not with homonyms (words that have the same spelling but different meanings), due to the fact that homonyms have an identical orthographic representation, leading learners to pronounce them identically. In addition, based on previous evidence from Browning (2004, as cited in Bassetti & Atkinson, 2015) on primary school L1 Italian children pronouncing English silent letters, Bassetti and Atkinson hypothesised that not only L1 Italian L2 English beginning learners but also experienced instructed learners would pronounce English silent letters. Moreover, because of the phonologically transparent orthographic system of Italian, their hypotheses also included the pronunciation of English words that have vowels represented by digraphs with a longer duration than those which have vowels represented by a singleton letter, e.g., the vowel in *seen* will be pronounced longer than that in *scene*. All their findings were in line with their hypotheses. In the first hypothesis, which predicts pronouncing homophonic words differently, participants were found to map two phonological forms of homophonic words, which was not the case with homonyms. For the second hypothesis of both beginners and experienced learners pronouncing silent letters, they found that almost all the participants pronounced words that end with the cluster <mb> and <bt> as /mb/ and /bt/, despite the fact that these clusters do not have phonetic representations in English. The last hypothesis was also in line with their findings, as the participants were found to pronounce vowels longer when they are spelt with vowel digraphs than when they are spelt with a single vowel, even if the word ends with a silent <e> which indicates length of the vowel in English. They concluded that orthography plays an essential role in determining the phonological representations of target words. They argue that the reason behind pronouncing vowels longer than English native speakers in the case of vowel digraphs, adding phonemes to words that have silent letters, or pronouncing homophones differently, is their orthographic representation, not the transfer of L1 phonology. They built this claim in the vowel digraph case upon the fact that vowel length is not contrastive in the participants' native language, although they produced English words with different lengths of vowels depending on the way a word is spelt.

The findings of Bassetti and Atkinson (2015) are in line with those of Bassetti et al. (2020), who found that L1 Italian learners of English pronounced consonants and vowels longer or shorter based on whether they are spelt with a singleton letter or with a digraph. Bassetti et al. (2022), who also conducted their study on L1 Italian learners of English, also extended that to the case where students were explicitly instructed to the fact that consonant digraphs do not map to longer consonants, students continued to produce them as longer. In Bassetti et al's (2020) study, participants who were high school students studying English in Italy were affected more strongly in the case of consonants by the number of letters than bilinguals who were immigrants in the UK. So, they pronounced the /t/ in words like *kitty* longer than that in the word *city* because the phoneme /t/ in these two words is represented orthographically in two different ways, a double consonant in the former and a singleton consonant in the latter.

However, in the case of vowels, bilinguals were affected by the number of representing letters slightly more than learners. Notwithstanding that the effect was different between the two groups, both groups were affected by how the word was spelt. Going back to Bassetti and Atkinson's argument, the second case of their study was that of silent letters. They claimed that Italians pronouncing silent letters in English words was based on the effect of orthography rather than L1 phonology. This was evidenced by the case of cognate words (debito /debito/ 'debt' and salmone /salmone/ 'salmon') which had no effect and behaved no differently from noncognates, which in turn decreases the role of L1 phonology as both types of words (cognates vs noncognates) were pronounced in the same way. Lastly, Bassetti and Atkinson argue that the evidence for the effect of orthographic representation rather than L1 phonology for the last case, which is the case of homophones, is the fact that the pronunciation of homophones is affected by orthographic input, unlike homonyms, which were not affected because they have the same orthographic representation. Bassetti and Atkinson's argument is supported by Escudero (2015) who argue that language background plays no role with regard to the effect of orthographic availability on phonological development. She built her argument on her finding that the performances of her participants who were L1 Spanish and L1 Australian English learners of L2 Dutch were comparable, although the Dutch language is phonetically closer to Spanish in its written vowels.

2.4.4.2 Level of Congruency Between L1 and L2 Orthographic Systems

The extent to which L1 and L2 share the same graphemes that represent the same phonemes is referred to as congruency between the two languages (Showalter and Hayes-Harb, 2015). By way of illustration, there is a high congruency between English and French. For example, the grapheme <c> represents both /s/ and /k/ in both languages, like the two cognates *certifié* /sɜʁtɪfje/ "certificated" /sɜ:tɪfikeɪtɪd/ and *cabine* /kabi:n/ "cabin" /kabɪn/. This is not the case in

Turkish, for example, as the grapheme <c> represents the phoneme /dʒ/ like in the word *önce* /ondʒe/ "before". Erdener and Burnham (2005) demonstrated the effect of the congruency level of L1 and L2 orthographic systems on phonological development, finding that their L1 Turkish participants confused the two Spanish phonemes /x/ and /ʒ/, which was caused by the fact that the grapheme <j> represents the phoneme /ʒ/ in Turkish while it represents the phoneme /x/ in Spanish. The incongruency between the two languages confused the learners and resulted in negative effects of orthographic availability while learning L2 phonology. However, this was the only case in this study when the presence of orthography negatively affected learning. In other cases, it was found to be helpful. This was attributed to the transparency of the Turkish (L1) and Spanish (L2) orthographic systems which was discussed in detail in the previous subsection (Subsection 2.5.4.1). The level of L1 and L2 congruency in phonological and orthographic representations was also investigated by Hayes-Harb et al. (2010). Their participants, who were L1 English speakers, were divided into three groups: congruent orthography group, congruent/ incongruent orthography group, and auditory only group. The difference between the congruent and congruent/ incongruent groups was that the stimuli for the congruent group included words that share the same orthographic conventions as those in English, like, for example, *kamad* corresponding to /kɒməd/, whereas the congruent/ incongruent group were exposed to three types of words: words with the same spelling as the congruent group; words with an extra letter in their spelling such as *kamand* corresponding to /kɒməd/; and words with an incorrect letter in their spelling, such as *faza* corresponding to /fɑʃə/. Participants in this study went through a learning phase of some pseudowords including auditory stimuli and pictures with or without (depending on the group) written stimuli. The learning phase was later followed by a testing phase via a word-picture matching test. A comparison of the performance of the two groups having written stimuli showed that the participants' phonological representations were affected by orthographic representations. The performance of the participants in the congruent/ incongruent group was less accurate than those in the congruent group, especially when the grapheme-phoneme correspondence was different from their L1 conventions. Looking at the performance of the auditory group, it was found that their performance was the most accurate among the three groups, pointing to the fact that orthography has a negative role in L2 phonological development.

In addition, Nimz and Khattab (2020) investigated how the performance of L1 Polish learners of L2 German in producing vowels was related to the (in)congruency between these two languages. The German language has 15 vowels, whereas Polish has only six, which makes the German vocalic system difficult to learn for Polish learners. Moreover, there are a number of signals of incongruency between German and Polish, the German vowel /e:/, for example, which is represented in German as <e> is mapped to two Polish vowels /i/ and /ɨ/ represented in Polish as <i> and <y>, respectively. Although orthographic cues were found to be helpful in Nimz and Khattab's study on vowel length, vowel quality was affected negatively by the incongruency between German and Polish grapheme-to-phoneme correspondences. Nimz and Khattab concluded that orthography can have a two-sided effect on L2 phonological development. Escudero et al. (2008) conducted a study on L1 Dutch learners of L2 English. They concluded that their participants were able to differentiate between the two vowels /æ/ and /e/ depending on their spelling if they are spelt as <a> and <e>, respectively. By looking at the congruency between English and Dutch, it is seen that the two graphemes have different representations in the two languages. In English, both <a> and <e> represent front vowels, whereas in Dutch, the grapheme <a> has a low back vowel representation while the grapheme <e> has a front central vowel representation. The participants' task was to attempt to build up their lexicon by linking a presented auditory form with the picture it represents, so they heard one auditory form and saw two pictures and had to choose which one of the two pictures represents the auditory form. Escudero et al.'s (2008) conclusion was not in line with those of

Bassetti (2005, 2006) and Hayes-Harb et al. (2010). Escudero et al. (2008) points out that the orthographic differences representing phonemic contrasts that exist in learners' L1 can be transferred to help learners learn new L2 words. However, there might be some other factors that affected the word recognition of Escudero et al.'s (2008) participants. For example, other differences that exist in the lexical stimulus, which are supposed to present the two target phonemes /æ/ and /e/, may have impacted on the participants' word recognition. The list of lexical items included nonwords targeting the two phonemes, but they were not minimal pairs, like *tenzer* and *tandek* or *meskle* and *mastik*. These other differences in the words, rather than /æ/ and /e/, may have facilitated word recognition, as participants may have depended on these other phonemes to discriminate the lexical representations of the presented pictures. It is argued that a listener's ability to discriminate between lexical items including these vowels in particular is possible even in the case of their inability to distinguish between these vowels auditorily (Weber & Cutler, 2004). Thus, it might have been more reliable if the study consisted of minimal pairs contrasting only in the target phonemes.

Hayes-Harb and Cheng (2016) tested the role of the congruency, along with familiarity, on L1 English naïve listeners of Chinese in two different orthographic systems, Pinyin and Zhuyin, where Pinyin consists of Roman letters and Zhuyin consists of entirely unfamiliar graphemes. The researchers conducted three experiments aiming to test the orthographic effect in the participants' ability to learn grapheme-phoneme correspondences, word learning, and phoneme discrimination. Their experiment included words that were either congruent to English represented in Pinyin, incongruent to English represented in Pinyin, or words represented in Zhuyin. Findings showed that, in the case of incongruent words, the Zhuyin group outperformed Pinyin group in grapheme-phoneme correspondences learning and word learning, indicating the negative effect of incongruency. In the case of congruent words, on the other hand, the two groups did not significantly differ. However, as for the phoneme discrimination, the two groups did not significantly differ in their performance.

2.4.4.3 Level of Familiarity of L2 Script

Apart from the role of L1 and L2 congruency, other factors may have a key impact on the effect of orthographic input on L2 phonological development. A number of researchers have attempted to investigate the role of script familiarity in L2 phonological development. In other words, whether being exposed to familiar or unfamiliar script play different roles in helping L2 learners to develop their phonological learning (e.g., Hayes-Harb & Cheng, 2016; Hayes-Harb & Hacking, 2015; Jackson; 2016; Mathieu, 2014, 2016; Showalter, 2012; Showalter & Hayes-Harb, 2013, 2015). By way of illustration, Showalter and Hayes-Harb (2013) conducted a study to investigate the role of a partially unfamiliar orthographic system, which means only some aspects of the writing system were novel to the participants. They targeted L1 English naïve listeners of Mandarin using Pinyin in their study. Pinyin is an orthographic system that uses Roman letters with diacritic marks. Thus, the graphemes were familiar to the learners but diacritic marks were not. Showalter and Hayes-Harb found that the existence of orthography helped the learners to differentiate between phonemes despite the fact that diacritic marks were unfamiliar to them. So, Showalter and Hayes-Harb concluded that orthography facilitates phonological development even in the case of partially unfamiliar items. This study led them to develop their research method and investigate entirely novel orthographic systems in their following study (Showalter & Hayes-Harb, 2015).

Showalter and Hayes-Harb (2015) conducted a study that investigated the role of a totally novel orthographic system in the development of L2 phonology. Their participants were L1 English learners of L2 Arabic, which has a totally different orthographic system from that in English in terms of its graphemic symbols, its writing system, and its directionality (Arabic is written from right to left). The methodology was identical to their (2013) study. Participants were allocated into two different groups, an orthography group and a no-orthography group, and both groups went through the same three phases. The first phase was dedicated to learning words by presenting pictures along with Arabic non-words. The orthography group had additional input which was words written in Arabic, whereas the no-orthography group had the symbol [ططط] corresponding to [xxxx] in English. In the second phase, a Criterion Test Phase, the researchers aimed to measure their participants' acquisition of meaning. The participants were provided with a picture while listening to one of the non-words they had learned. Their task was to decide whether the picture represented the word or not. In this phase, if there was no match between the picture and the auditory form, the two words not only contrasted in one phoneme, but rather they were totally different, e.g., participants heard the word /kubu/ and saw the picture for the word /qaʃu/. This is because the goal of this phase was only to enhance their learning and test the acquisition of meaning apart from phonological representation, which was tested in a later phase. The last phase, a Final Test Phase, was very similar to the Criterion Test Phase. The only difference between the two phases was that the participants were required to differentiate between two target phonemes, /q/ and /k/, in which if the picture and auditory form do not match, words are minimal pairs that contrast only in these two target phonemes, e.g., participants heard the word /kubu/ and saw the picture for the word /qubu/. The results showed that Arabic orthography did not play any role in learners' ability to distinguish between the two target phonemes. The researchers attributed their findings to two possible reasons. The first reason is the perceptual difficulty of the two target phonemes as both of them map to only one English L1 phoneme, namely /k/. This is discussed in more detail in the following subsection (Subsection 2.5.4.4). The second reason might be the difficulty and unfamiliarity of Arabic script. The participants might find it difficult to interpret Arabic script because it has different shapes of graphemes as well as different directionality. This led the researchers to provide the learners with a version of Arabic script transliterated into a

Romanised version. This procedure caused them to deviate from their intended method, as it was planned to investigate a totally novel script. However, they concluded that the Roman script affected the learners negatively because they used the grapheme <k> to represent the phoneme /k/ and the grapheme <q> to represent the phoneme /q/. This might have led the learners to think that they are both pronounced the same as these two graphemes represent the same phoneme in English, even though <q>, however, tends to be associated with /kw/.

Jackson (2016) conducted a follow-up study to that of Showalter and Hayes-Harb (2015) using the same auditory and visual stimuli. Jackson provided the participants, who were also L1 English naïve listeners of Arabic, with two types of orthographic input for the novel phoneme /q/: novel grapheme or diacritic dot under the grapheme <k> to differentiate it from the grapheme <k> representing /k/. The rest of graphemes in the word were presented in a Roman script. Their findings indicated that participants' performance was more accurate when they were exposed to novel grapheme than when they were given a diacritic dot. They attributed this disadvantage of the diacritic dot to the similarity between the two grapheme forms (with or without a diacritic dot). This similarity may have led participants to think that the difference between the two phonemes was of no importance. Thus, having novel grapheme was facilitative, given that the novel grapheme was more noticeable, and it does not map to any existing L1 forms.

Based on a methodology similar to these two studies, Mathieu (2016) investigated the extent to which familiarity of script affects phonological category formation by examining the perception of two Arabic phonemes /ħ/ and /χ/ by L1 English naïve listeners of Arabic. Three types of script were included in this study. The first was Arabic script which was entirely unfamiliar to participants of the study. The second was Cyrillic script, that of Kazakh. Cyrillic script, unlike Arabic, is characterised by having independent characters, i.e., not adjoined to adjacent characters in writing. This feature made it easy for participants to parse. However, it consists of many superficially Roman-like characters including the target phonemes represented by <h> and <x>. This may have had a negative effect due to the incongruency effect, as discussed in the previous subsection. The third script was a Hybrid script, which was a mix of familiar and unfamiliar characters. In an attempt to avoid an incongruency effect and simultaneously retain the unfamiliarity factor, the target phonemes were changed to unfamiliar Cyrillic characters, whereas the rest of the phonemes were kept as Roman script. The researcher hypothesised that participants who were exposed to Arabic script would have the lowest performance of all, followed by those who were exposed to Cyrillic script. Participants who were exposed to Hybrid script were expected to have better performance than participants seeing Arabic and Cyrillic scripts, but equal to or lower than those who did not see any script. Findings did not support these hypotheses as all three conditions of unfamiliar scripts were not significantly different, suggesting that acquisition of non-native contrasts is not affected by the degree of script familiarity. By comparing these conditions with those who were not exposed to orthography, it was found that the latter had significantly better performance than the Arabic-script and Cyrillic-script groups. As script familiarity had no effect, the researcher attributed the lower performance by the Arabic-script group to the interaction between the perceived difficulty and unfamiliarity of the script. However, a possible reason for the lower performance of the Cyrillic-script group was the interaction between the perceived difficulty and incongruency of the graphemic representation of the target phonemes. Finally, the Hybrid-script group was found to have lower performance than those who were not exposed to orthography, but this difference was not significant. According to the researcher, the reason for this insignificant difference was because having familiar letters prevents the establishment of phonetic categories of contrasts due to the activation of L1 phonological units.

Hayes-Harb and Hacking (2015) also conducted a study with a very similar methodology investigating the role of script familiarity in the usefulness of inserting lexical stress marks. They examined L1 English naïve listeners of Russian in their Russian word learning. Two groups of participants were included in this study, participants who had no prior experience with Russian language, and participants who were experienced learners of Russian. These participants were allocated into two different cases. The first case was consisting of experienced learners who were only exposed to Cyrillic script (unfamiliar script) with two conditions: stress marks and no stress marks. The second case was consisting of inexperienced learners who were exposed to both Cyrillic (unfamiliar script) and Latin scripts (familiar script) with two conditions: stress marks and no stress marks. Their findings showed that, for the inexperienced participants, neither familiarity/unfamiliarity nor the presence or lack of stress marks affected their performance. There was also no significant effect for experienced participants in the condition of stress marks from that of no stress marks. Moreover, the performance of the two groups of participants did not significantly differ in the condition of stress marks from that in the no stress marks condition. All these findings led to the conclusion that orthographic input did not have any effective role in participants' performance regardless of their experience level.

The role of familiarity of L2 script in the development of L2 phonology is also supported by the findings of Bassetti (2006). Her study was conducted on L1 English learners of Chinese. She hypothesised that the phonological representation of learners of L2 Chinese would be negatively influenced by Pinyin orthographic representation, resulting in non-target-like pronunciation caused by the orthographic input's interference in the mental phonological representation of learners. This hypothesis is supported by Bassetti (2005), as she argues that the ability to segment language into phonological units for L2 learners who are learning a language that has a different writing system from their L1 is not the same as that of those who are familiar with only one writing system. This ability is expanded to improve different ways that learners can use to segment different linguistic units. This ability is enhanced by knowing two different writing systems. Hayes-Harb & Barrios (2021) argue that the characteristics of writing systems affect how reading is processed. This is because knowing two languages that have two different writing systems leads to an interaction between these two systems, in addition to the requirement of adjusting the input based on what each language imposes. These adjustments, according to Hayes-Harb & Barrios (2021), include considering the script type (e.g., logographic or phonographic), transparency (e.g., transparent/shallow orthography or opaque/deep orthography), script direction (e.g., left-to-right, right-to-left, or top-down), and congruency (employing the same or different grapheme-phoneme correspondences). Bassetti (2005), who conducted her study on L1 English learners of L2 Chinese, built her argument based on the findings of Tsai, McConkie and Zheng (1998), who conducted a study on participants who were originally two different groups. The first group included L1 Chinese participants who knew Pinyin, which is, as mentioned above, a writing system for Standard Chinese that uses Roman letters with diacritic marks. The second group included Chinese who used only hanzi, which means using Chinese characters in the writing system. The participants' task was to segment Chinese texts that consisted of 300 characters; they were asked to put marks between the words depending on their subjective evaluation. The performance of Tsai et al.'s participants in the two groups differed, as those who knew Pinyin outperformed their counterparts on dividing the characters into words. Besides, according to Bassetti (2005), learners of a second language that has a different orthographic system than that of their L1 are more likely to develop different ways to segment the linguistic units from those used by native speakers of this language. This is enhanced by knowing two different orthographic systems, as mentioned above. Bassetti (2006) found that her participants depended on how the words were spelt in Pinyin rather than on how they were pronounced. This was obvious in the participants' interpretation of some written vowels which were omitted and thus not counted in a phoneme-counting task, this impacted on the conventions of the English language. Bassetti's (2006) study included two different experiments conducted on two different groups. The stimuli in this study included different Chinese syllables written in Pinyin that were either phonology-orthography consistent (high phonological transparency) or phonology-orthography inconsistent (low phonological transparency). The tasks for both types of stimuli (phonology-orthography consistent and phonology-orthography inconsistent) required the participants to count the number of phonemes in the first experiment and pronounce each single phoneme separately in the second experiment. Bassetti (2006) found that phonological-orthography consistency had a significant main effect in which participants counted more phonemes in syllables where its pinyin representation consists of a vowel than those where its pinyin representation does not consist of a main vowel. This indicated that the orthographic representation in pinyin impacted on the mental representations of Chinese syllables. Bassetti concluded that the interpretation of orthographic representation in terms of either phoneme counting or phoneme segmentation was found to be influenced by L1 grapheme-phoneme conversion rules rather than the conventions of L2 orthography. The participants built a mapping between phonological input and orthographic input that was interpreted according to L1 grapheme-phoneme conversion rules, resulting in creating a mental representation of non-target-like syllables that affected their phonological awareness level as well as their L2 production. However, the number of participants in the group was small, which may have led to generalizability problems, as she had 18 participants in the first group and only five in the second. In addition, Bassetti used different measures with different groups, which may have affected the reliability of her measures, because having different measures with different groups might have added more confounding variables that impacted on the outcome.

The findings of Bassetti (2006) are also supported by the findings of Vaid, Chen and Rao (2022) who examined the performance of two groups knowing two writing systems in a phonological awareness task. The first group included native speakers of Hindi who were either experienced learners of English or immigrants to the USA who rarely read or write in Hindi. The second group included native speakers of English who had had formal instruction in Hindi for no more than two years. The Hindi orthographic system differs from English in that its basic writing unit is the syllable (CV), whereas the basic writing unit in English is the grapheme (C or V). Participants in both groups were presented orally with cross-language homophones that were used in both languages (e.g., *hum* “we” in Hindi and “music sense” in English) and were asked to delete the first segment of the word. The study was designed to help participants to build their decision on which list the word was presented in: either Hindi or English word lists. Findings showed that the English speakers’ decision was always based on phoneme segmentation regardless of which group the word belonged to. However, the Hindi speakers made their decision based on the group the word belonged to, i.e., phoneme segmentation if the word belonged to the English list, and syllable segmentation if the word belonged to the Hindi list. Vaid et al. (2022) attributed the fact that the native Hindi speakers outperformed the native English speakers to either their longer experience with the L2 or the strong impact of knowing an alphabetical writing system as an L1 writing system, which might have had an impact on the native English speakers.

2.4.4.4 Level of Difficulty of Perceived Auditory Forms

Some researchers have investigated the role of the level of difficulty of perceived sounds with regard to the effect of orthography availability in L2 phonological development. One reason for difficulty is the case when two contrasting phonemes in L2 are mapped to one L1 phonemic category in L1 (Best, 1995; Best & Taylor, 2007; Cutler, 2015; Flege, 1995, 2016;

Flege & Bohn, 2021). Learners are likely to be unable to discriminate these phonemic contrasts either perceptively or productively because these contrasts do not exist in their L1 (Best, 1995; Best & Taylor, 2007; Escudero, 2015, Flege, 1995, 2016; Flege & Bohn, 2021). This problem leads to the misperception of minimal pairs that include these contrasts, as in the case of Showalter and Hayes-Harb's (2015) study where participants failed to discriminate the contrast between the two Arabic phonemes /k/ and /q/. This idea is supported by a number of speech perception models such as the SLM-r (Flege & Bohn, 2021), PAM (Best, 1995), PAM-L2 (Best & Taylor, 2007) (discussed in detail in Section 2.2) and the second language linguistic perception (L2LP) model (Escudero, 2005). The concept of the L2LP model agrees with the SLM-r in that it highlights the role of L1 perceptual behaviour in the initial stages of L2 acquisition. It states that learners tend to rely on their L1 perceptual behaviour in that stage. In other words, L2LP is based on the Full Copying Hypothesis, where learners start their L2 learning by copying the grammar of their L1 perception; y getting more exposure to L2, these connections between L1 and L2 are shifted to better suit L2 perception and recognition (Escudero, 2005). This is indeed related to the acoustic similarities shared by L1 and L2 sounds that affect how non-native sounds are perceived, which is also in line with Flege's SLM (Escudero, 2005). It is worth pointing here that L2LP (Escudero, 2005) has a revised version (van Leussen & Escudero, 2015) that includes the case where one L2 phoneme maps to more than one L1 phonemes. According to the revised version of L2LP, the process of the perception of phonemes in this scenario is driven by the meaning of the lexical items. This differs from the previous version in that the process of perception is driven by the learners' awareness of the number and types of phonemes in relation to their new language. In other words, the difficulty of L2 phonological development can be predicted by the acoustic similarities between L1 and L2 sounds, especially those that correspond to only one L1 category rather than those that correspond to two different ones. A number of studies provide some examples of how this difficulty occurs when two L2 sounds are perceived as one single L1 sound. For instance, it was found that L1 Spanish learners of L2 Dutch found it difficult to distinguish vowel contrasts in Dutch that have only one counterpart in Spanish (Escudero & Wanrooij, 2010; Escudero & Williams, 2012). For example, Escudero and Williams' (2012) findings show that the Dutch contrasts /a/ vs /ɑ/ and /y/ vs /ʏ/ are categorised as one single Spanish vowel for each pair: /a/ for the former and /u/ for the latter. On the other hand, Escudero et al. (2008) found that their participants, who were L1 Dutch learners of L2 English, when presented with only auditory forms along with pictures, could not distinguish between initial syllables containing the two pairs /æ/ and /e/ because these two pairs were reported to be difficult to discriminate by them (Broersma, 2005), which was not the case when orthographic input was available. This can be evidenced by the findings of Escudero et al., (2013), who found that in some cases L1 Spanish learners of L2 Dutch can easily discriminate the contrasts in other minimal pairs, either familiar or novel ones, and this is because these phonemes contrasting minimal pairs are already distinguished in their L1. Escudero (2015) argues that the orthography availability of perceptually difficult contrasts can play a positive role in developing L2 learners' perception. She elaborates on that by stating that perceptually difficult contrasts can help to create separate lexical representations for minimal pairs carrying these contrasts, which results in developing new phonological categories in L2 acquisition. Escudero believes that having separate lexical representations is important for learning novel minimal pairs, in addition to the development of recognition and production.

In her study, Escudero (2015) targeted the role of difficulty in discriminating L2 contrasts to investigate how orthography affects the phonological development of Spanish and Australian learners of Dutch. The vowel inventories in the three languages differ in many aspects. Spanish, for example, has five monophthongs /a/, /e/, /i/, /o/ and /u/ where vowel duration does not have any phonemic differences. Dutch, on the other hand, has a bigger vowel inventory that consists of 15 vowels divided into monophthongs and diphthongs, where vowel duration plays an essential role in determining vowel identity (Escudero & Wanrooij, 2010). According to Escudero (2015), English and Dutch have some shared vowel contrasts that Spanish does not have, in addition to the fact that vowel duration is phonemic in English. Escudero (2015) included three types of lexical stimuli which consisted of 66 lexical items divided into eight perceptually easy minimal pairs that included these phonemes (/ɪ-ɑ/, /ɪ-a/, /i-ɑ/, /i-a/, /ɑ-ʏ/, /ɑ-y/, /a-ʏ/, /a-y/), seven perceptually difficult minimal pairs that included these phonemes (/ɑ-a/, /i-ɪ/, /y-ʏ/, /ɪ-ʏ/, /ɪ-y/, /i-ʏ/, /i-y/), and 51 non-minimal pairs. Escudero depended in her classification on Spanish listeners' vowel discrimination performance in her previous study (Escudero & Wanrooij, 2010), as well as the possibility of perceiving vowel contrasts in Dutch that lead to two different Spanish vowels (perceptually easy) or that lead to one single Spanish vowel (perceptually difficult). Escudero's study showed that orthography was found to be helpful only in two pairs, those which were classified as perceptually difficult, namely /ɪ/ vs /y/ and /ɪ/ vs /ʏ/, whereas there was no role for orthography in perceptual easy and non-minimal pairs. This indicates that, as mentioned above, orthography is just a redundant cue to highlight differences already perceived. In addition, she found that the Spanish participants outperformed their Australian participants in perceptual easy minimal pairs in the case of vowels that have a large acoustic distance in Spanish and Dutch. This was not expected by the researcher as she thought that Australian participants would be likely to find these contrasts more perceptually salient because English has a larger vowel inventory. Escudero (2015) attributed her findings to the possibility that enlarging the range from which participants could choose, the case where they have several native phonemes that might correspond to one single L2 phoneme, may have led the participants to be less accurate in their recognition of minimal pairs. Escudero built her attribution on some previous studies that reported that the number of response options has an effect on perceptual abilities (Benders, Escudero, & Sjerps, 2012 as cited in Escudero 2015). This, according to Escudero (2015), might apply to the case where a number of native vowel categories lead to a single non-native vowel. This result led Escudero to suggest to avoid orthography during L2 phonological learning as it did not show any effect, either positive or negative, except for just two pairs.

2.4.5 Summary of the Section

This section presented a detailed overview of studies conducted to investigate the role of orthographic availability in the development of L2 phonology. It highlighted how this topic is important in enhancing the theoretical and practical backgrounds of L2 learning and teaching. It also took a brief look at the history of this topic in the literature. It elaborated on how this topic is presented in the literature, including different methods and different factors that affect the findings of these studies. This section also included a deeper discussion of the reasons behind having contrasting findings, pointing out the factors that have been investigated in this area.

2.5 The Arabic Language

2.5.1 Introduction

Arabic is a language that is spoken by 300 million native speakers in 27 states (Saiegh-Haddad & Henkin-Roitfarb, 2014). Arabic has three main varieties: Classical Arabic, Modern Standard Arabic (MSA), and colloquial Arabic. The first one is the variety used in the pre-Islamic and early post-Islamic eras. Thus, it is the variety used in the Holy Quran as well as old literary forms. MSA follows the same grammatical rules but, in addition to a large number of lexical items from Classical Arabic, it has added some modernised vocabulary. Classical Arabic and MSA are known as the literary varieties of Arabic as they are considered the primary language of literacy (Saiegh-Haddad & Henkin-Roitfarb, 2014). MSA is used to serve educational and media purposes, including formal written and spoken contexts. Thus, the variety that is used in the pedagogical context is MSA, as children are taught how to read and write in schools using this variety (Almahmoud, 2013). This variety is not the native variety of any L1 Arabic speaker as it is learned not acquired. Colloquial Arabic consists of many regional dialects which differ from one country to another and from one region to another. Colloquial Arabic refers to a group of Arabic vernaculars in different local dialects. All native speakers of Arabic acquire these varieties as their mother tongue. Because the written form is usually presented in MSA, colloquial dialects occur only in spoken contexts, except for informal texting where Arabs tend to use their regional dialects when texting to each other. (Saiegh-Haddad & Henkin-Roitfarb, 2014). However, because MSA is the variety used in education, and it is the official variety of Arabic, in addition to being the variety taught to foreign learners (Almahmoud, 2013), this study focuses on this variety only. Therefore, the following subsections present a detailed description of this variety.

This section has two main subsections. The first subsection describes the phonology of MSA, including vowels, consonants, stress, and syllabic structure. This touches upon some aspects of Arabic consonants such as emphatic consonants and their effects on surrounding vowels. The second subsection describes the main characteristics of MSA orthography. This presents how diacritics in the Arabic orthographic system affect meaning. Moreover, this subsection discusses some examples of the opaqueness of the MSA orthographic system.

2.5.2 Phonology of MSA

MSA phonemes are divided into consonants and vowels. There are six vowels in MSA and 28 consonants (Almahmoud, 2013; Tibi & Kirby, 2018).

2.5.2.1 Vowels in MSA.

Vowels in MSA are characterised by two features. These are quality and quantity (Watson, 2002). Quality refers to three parameters: height of the tongue (the extent to which the jaw is opened during the production of a vowel), the position of the tongue in the front-to-back dimension (whether the tongue is pushed forwards, left in its normal position, or pulled backwards during the production of a vowel), and the shape of the lips (whether they are rounded or unrounded during the production of a vowel) (Maddieson, 2013). According to Ladefoged and Ferrari (2012), vowel quality differences involve more complicated acoustic correlates which can be shown in the differences in the shape of the soundwave caused by the repetition rate and size. As for tongue height, only two vowels in MSA are open: /a/ and /a:/, while the other four are closed. In terms of tongue position and lips shape, MSA has only two rounded and back vowels; /u:/ and /u/, the rest are unrounded and front vowels. Quantity, on the other hand, refers to whether a vowel is long or short, meaning that the articulators remain longer in one position when producing long vowels than when producing short vowels

(Alghamdi, 2015). Thus, there are two types of vowels in MSA in terms of quantity: long vowels and short vowels. There are three long vowels, /i:/, /u:/, and /a:/, and also three short vowels, /i/, /u/, and /a/ (Saiegh-Haddad & Henkin-Roitfarb, 2014). These six Arabic vowels are all monophthongs (vowels that are constant and do not move or glide to another vowel) (Roach, 2000). Thus, as shown in Figure 2.5, which is adopted from the International Phonetic association framework, the vowel system in MSA is triangular (Binasfour, 2018).

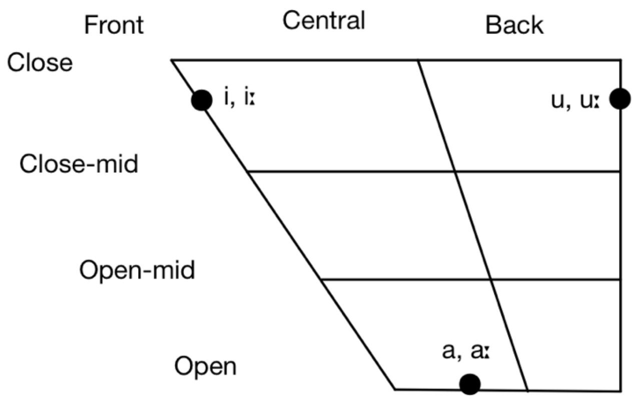


Figure 2.5 Inventory of MSA vowels (Salameh & Abu-Melhim, 2014)

These vowels differ phonemically, i.e., replacing one of them with another leads to a change in meaning (Aljasser & Vitevitch, 2018). For example, the word /dafaʕ/ 'he paid' is different from the word /da:faʕ/ 'he defended', and the word /qatal/ 'he killed' is different from the word /qa:tal/ 'he fought'.

2.5.2.2 Consonants in MSA

Tibi et al. (2021) states that some consonants in Arabic share their place of articulation, manner of articulation, and are distinguished by either voicing or pharyngealisation. Other consonants share their place of articulation, manner of articulation, and also voicing, and are distinguished only by pharyngealistion. Further discussion of pharyngealisation in Arabic phonemes is presented in Subsection 2.5.2.2.1. Table 2.1 presents MSA phonemes in terms of the four features of each MSA consonant: place of articulation, manner of articulation, voicing, and pharyngealisation. This table is adapted from Alghamdi (2015) with some modifications. Some terminology of the Distinctive Feature Theory is used in this table, such as [±voice] and [±pharyngealisation] (Jakobson, Fant, & Halle, 1952, as cited in Skidmore & Gutkin, 2020). This theory states that every speech sound is characterised by distinctive features that are either present or absent, for example +voice if it is voiced and -voice if it is not. Looking at Table 2.1, it appears that the phonemes of Arabic are distinguished from those of English language (the native language of the participants in the current study) by nine consonants. These are the two pharyngeal fricatives, /ħ/ and /ʕ/, a uvular plosive, /q/, two uvular fricatives, /ʁ/ and /χ/, and four emphatic consonants, /sˤ/, /tˤ/, /dˤ/ and /ðˤ/.

2.5.2.2.1 Emphatic Consonants.

One essential feature of Arabic is the presence of emphatic consonants (Embarki et al., 2007). The production of these consonants involves a coarticulation of the main place of articulation of the sound in addition to the back of the tongue moving towards the pharyngeal wall (Amayreh & Dyson, 1998), which makes them known as emphatic consonants. Jongman et al. (2011) and Hayes-Harb and Durham (2016) point out that the articulation involved in producing emphatic consonants is either uvularised or pharyngealised. However, this can be clearly observed with the uvular plosive /q/ more than the two uvular fricatives /ʁ/ and /χ/, because /q/ is usually followed by a pharyngealised vowel such as in the word قال /qa:l/ 'he said', but /ʁ/ and /χ/ are not.

Several studies have reported some distinctive characteristics marking the Arabic sound system, such as gemination, pharyngealisation and vowel duration, that are considered to be difficulties faced by L2 MSA learners (Alghamdi, 2015). Gemination refers to linking two consonant slots to a single consonant (Davis & Ragheb, 2014). For instance, /s/ in the word كسّر/kas.sar/ 'he smashed'. Gemination is phonemic in Arabic; for example /kasar/, 'he broke', is different in meaning from /kas.sar/, 'he smashed' (Davis & Ragheb, 2014). This difference is indicated by the use of an intervocalic geminate consonant. MSA and a large number of Arabic dialects do not allow gemination in the initial position, only in the middle and final positions (Davis & Ragheb, 2014). However, as gemination is not what this study investigates, there will be no further discussion of this topic.

Embarki et al. (2007) argue that in that the existence of emphatic consonants the Arabic, and the absence of this characteristic in most other languages of the world, causes difficulties for L2 Arabic learners in both the perception and production of these four consonants. Abu Rabia and Sammour (2013) attribute this difficulty to the extent to which these sounds (emphatics) are similar acoustically and auditorily to those that exist in their L1 (non-emphatics), bearing in mind that non-emphatic consonants exist in most world languages, whereas emphatic consonants do not.

Table 2.1: *Inventory of MSA Consonants (Alghamdi, 2015)*

		Bilabial		Labiodental		Interdental		Alveodental		Alveopalatal		Palatal		Velar		Uvular		Pharyngeal		Glottal	
		+v	-v	+v	-v	+v	-v	+v	-v	+v	-v	+v	-v	+v	-v	+v	-v	+v	-v	+v	-v
Nasal	+pharyngealisation																				
	-pharyngealisation	/m/						/n/													
Plosive	+pharyngealisation							/dˤ/	/tˤ/												
	-pharyngealisation	/b/						/d/	/t/						/k/		/ q/				/ʔ/
Fricative	+pharyngealisation					/ðˤ/			/sˤ/												
	-pharyngealisation				/f/	/ð/	/θ/	/z/	/s/		/ʃ/					/ʁ/	/χ/	/ʕ/	/ħ/		/h/
Affricate	+pharyngealisation																				
	-pharyngealisation									/dʒ/											
Glide	+pharyngealisation																				
	-pharyngealisation	/w/										/j/									
Lateral	+pharyngealisation																				
	-pharyngealisation							/l/													
Trill	+pharyngealisation																				
	-pharyngealisation							/r/													

Influence of some consonants on producing vowels in Arabic

According to Jongman et al. (2011), a considerable number of previous pharyngealisation acoustic analyses have concentrated on surrounding vowels' properties rather than emphatic consonants themselves, even though they consider that the main position of pharyngealisation is on the consonant. One essential characteristic of Arabic is that the phonetic representation of vowels is usually determined by surrounding consonants (Holes, 2004). Holes (2004) points out that emphatic consonants affect the production of the vowels preceding or following these consonants, leading to having different allophones of these vowels from those following their non-emphatic counterparts. This is obvious in having vowels with a more pronounced back feature surrounding emphatic consonants than those which are surrounded by non-emphatic consonants. This is known as emphasis or pharyngealisation spread and leads to more than one allophone for each Arabic vowel (Shar & Ingram, 2010). This feature is what makes the contrast between emphatic and non-emphatic sounds different because, when producing emphatic sounds, it is necessary to move the back of the tongue towards the rear of the pharyngeal wall (Shar & Ingram, 2010), resulting in moving the tongue during production of the vowel towards the rear of the vowel space (Alkhatib, 2008).

As the vocal cords produce vibration through the glottis causing resonance to occur, the formants are produced by raising and lowering the degree of these vibrations. These formants are not stable in the degree of vibration depending on the position of the tongue and other articulators in the vocal tract. F1 is related to the closeness between the tongue and the roof of the mouth, so that the closer the tongue is to the roof of the mouth, the lower is the F1. F2, on the other hand, is related to the back of the tongue, so the higher the back of the tongue is, the lower is F2 (Alghamdi, 2015). Alghamdi (2015) elaborates MSA short and long vowels in terms of their acoustic characteristics. He points out that the acoustic difference between the three short vowels occurs in F1 and F2. F1 is found to be lower in the two high vowels /i/ and /u/, and higher in the low vowel /a/. F2, on the other hand, is found to be higher in the front vowel /i/, and lower in the two vowels /u/ and /a/. Because the long vowels differ mainly from the short ones only in their quantity, they are very similar to their short counterparts in terms of F1 and F2, as shown in the Figures 2.6, 2.7, and 2.8.

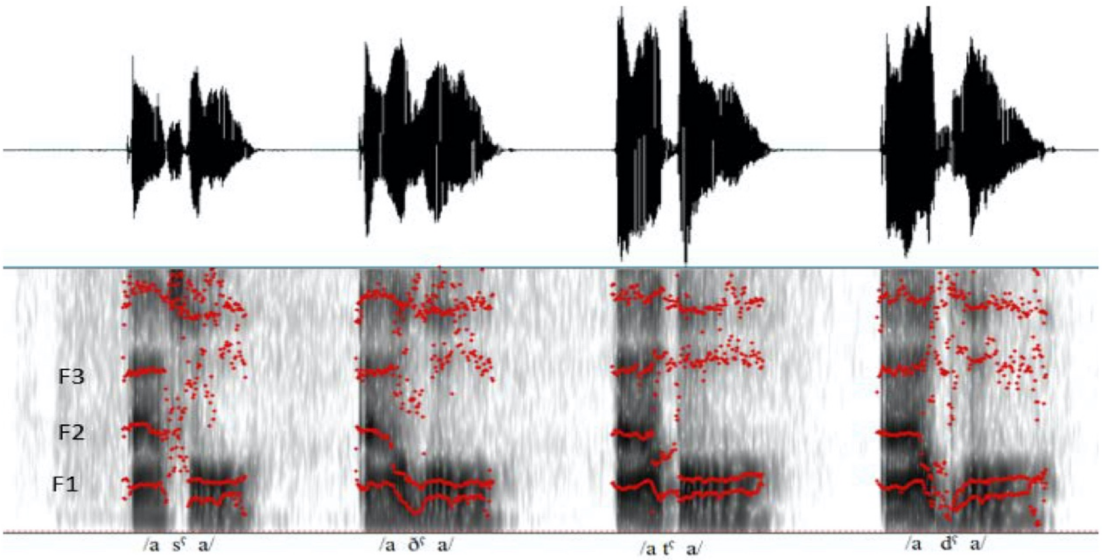


Figure 2.6 Spectrogram of the pharyngealised vowel /a/ (Binasfour, 2018)

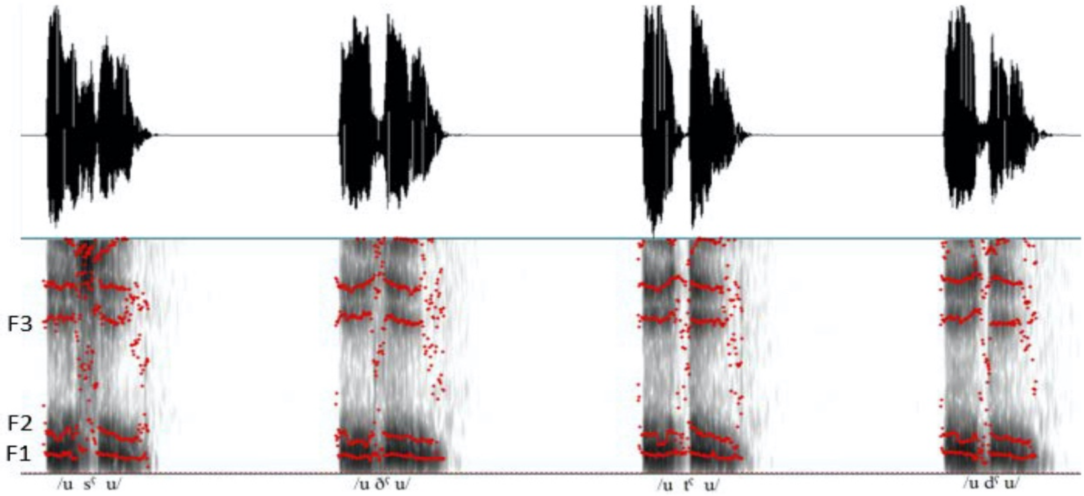


Figure 2.7 Spectrogram of the pharyngealised vowel /u/ (Binasfour, 2018)

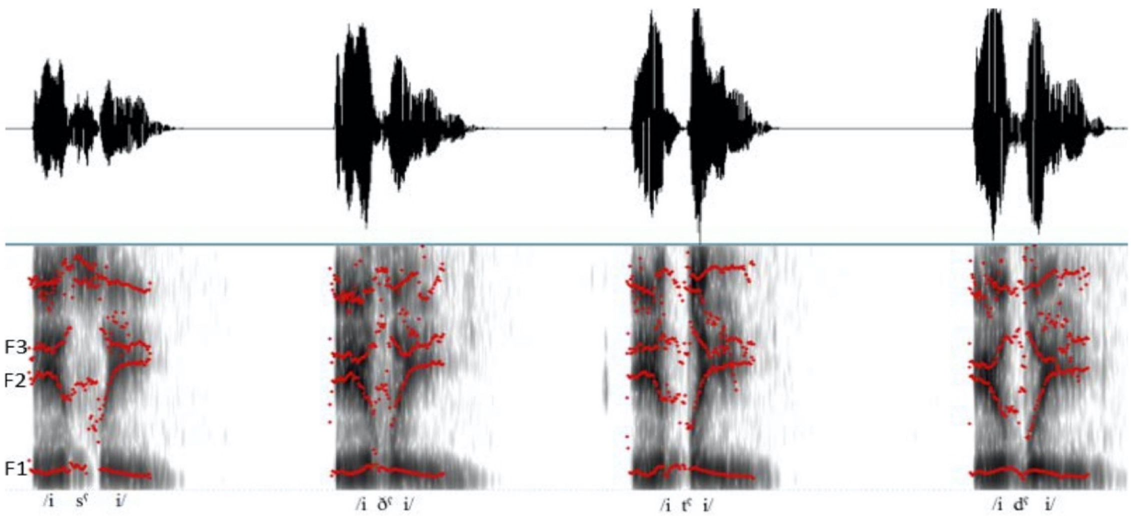


Figure 2.8 Spectrogram of the pharyngealised vowel /i/ (Binasfour, 2018)

The effect of pharyngealisation spread, or emphasis spread, can be seen in having a lower degree of F2 in the following or preceding vowels than those that follow or precede non-emphatic consonants (Binasfour, 2018). Shar and Ingram (2010) demonstrate this by stating that this effect can be seen in following or preceding vowels as well as in other surrounding sounds. Binasfour (2018) also argues that emphatic and non-emphatic sounds share the same primary articulation, but differences occur in their secondary articulation, which occurs at the back of the tongue. Figures 2.9 and 2.10 show how vowels change when surrounded by emphatic or non-emphatic consonants in two Arabic minimal pairs that contrast in /s/ and /sˤ/ in one case, and /t/ and /tˤ/ in another case (Binasfour, 2018). Therefore, these differences caused by the surrounding emphatic consonants result in having different allophones for MSA vowels.

Al-Ani (1970, as cited in Binasfour, 2018) describes how the six allophones of pharyngealised vowels are caused by following or preceding emphatic consonants. These are: the short high front unrounded vowel /i/ [ɨ] and its long counterpart /iː/ [ɨː]; the short high back rounded vowel /u/ [ʊ]; and its long counterpart /uː/ [ʊː]; and the short low central unrounded vowel /a/ [ɑ] and its long counterpart /aː/ [ɑː].

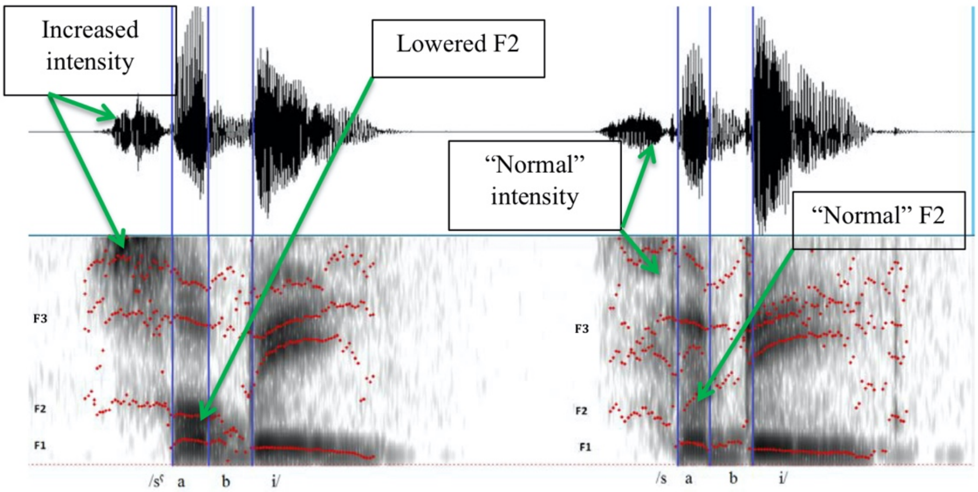


Figure 2.9 Acoustic representation of the Arabic minimal pairs /sˤabi/ and /sabi/ (Binasfour, 2018)

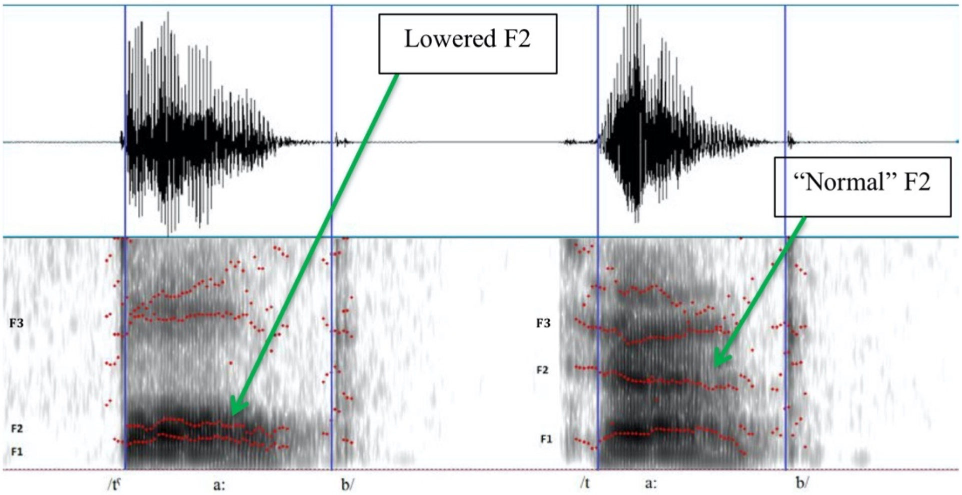


Figure 2.10 Acoustic representation of the Arabic minimal pairs /tˤa:b/ and /ta:b/ (Binasfour, 2018)

Jongman et al. (2011) also mention the acoustic effect of pharyngealisation on the surrounding vowels. They state that the F2 frequency of vowels surrounding emphatic plosives in either initial or final word positions are found to be lower than those surrounding non-emphatic consonants. According to Binasfour (2018), Card (1983), and Jongman et al. (2011), F2 which is lowered by surrounding emphatic consonants, is different from one vowel to another. They demonstrate that pharyngealisation has the weakest effect on the vowel /i/, while it has the greatest effect on /a/, as shown in Figures 2.6, 2.7, and 2.8 This results in making the emphatic consonants surrounding the vowel /a/ easier to discriminate than those which surround the vowel /i/. In other words, it is expected that L2 Arabic learners are more capable of producing or perceiving emphatic consonants which are preceded or followed by the two vowels /a/ or /u/ than those that are preceded or followed by the vowel /i/ (Binasfour, 2018). This claim is supported by the findings of Hayes-Harb and Durham (2016) who built their study on the basis of the findings of Walley and Carrell (1983, as cited in Hayes-Harb & Durham, 2016). On the other hand, Jongman et al. (2011) reported that vowels preceded or followed by emphatic consonants were found to display higher F1, which was greatest with the vowel /a/, just like F2.

Walley and Carrell (1983, as cited in Hayes-Harb & Durham, 2016) found that, in order to identify consonants, listeners usually depend on the acoustic information of surrounding vowels. Hayes-Harb and Durham (2016) tested this hypothesis by examining English L2 Arabic learners' ability to discriminate emphatic consonants from their non-emphatic counterparts. Participants completed a vowel identification task as one of many tasks in this study. The vowel identification task included all six Arabic vowels in which each vowel was presented in three tokens. Participants were asked to listen to these tokens and select which English vowel the word contained. Participants also had another discrimination task in which they were required to differentiate between L2 sounds and how close they were to participants' L1 sounds. In other words, participants were presented with three stimuli A, B, and X and were asked to decide which among A or B (L2 sounds) was closer to X (L1 sound). They found that their participants relied on the surrounding vowels in order to discriminate contrasted consonants (emphatic vs non-emphatic). Hayes-Harb and Durham state that their participants were able to detect the Arabic allophone [ɑ] because it overlaps with the two English vowels /ɑ/ and /æ/. However, this was not the case with the Arabic allophones [ɨ] and [ʊ], as it was not found that participants could distinguish these. The existence of the two Arabic allophones /ɑ/ and /æ/ with English vowel qualities is supported by Odisho (1981, as cited in Hayes-Harb & Durham, 2016) as a source of help to teach Arabic emphatics to L1 English learners of Arabic. This results in their claim that the allophones of /i/ and /u/ are identified as indistinguishable when surrounding either emphatic or non-emphatic consonants (Hayes-Harb & Durham, 2016). The reliance on surrounding vowels as a clue for discriminating the pharyngealisation feature of consonants has also been reported as being used by L1 Arabic speakers (Jongman et al., 2011). Jongman et al. found that Urban Jordanian Arabic speakers relied on the information provided by rime, the string of phonemes that includes a vowel and the ones following it in a syllable, more than their reliance on information provided by the onset consonant itself, the phonological unit that precedes the vowel in a syllable. Jongman et al.'s stimulus was recordings of the phrase "say ----- once more" uttered in Jordanian Arabic, as well as Arabic script printed notecards that consisted of Arabic monosyllabic words and non-words. These words included minimal pairs that contrast in the emphatic and its non-emphatic counterpart, including all four emphatic consonants and their non-emphatic counterparts. Participants were asked to repeat target words. Jongman et al.'s finding is also supported by my previous study (Alhumaid, 2019) which found that Saudi Arabic speakers rely on the following vowels in order to discriminate the contrast between non-word minimal pairs that include the two consonants /s/ and /sˤ/, such as /su:la/ and /sˤu:la/, or the two consonants /ð/ and /ðˤ/, such as /ða:ni/ and /ðˤa:ni/. This was done by manipulating the following vowels by switching the pharyngealised vowel to follow the non-emphatic consonant and the non-pharyngealised vowels to follow emphatic consonants. This manipulation led Arabic L1 participants to confuse emphatic and non-emphatic consonants and wrongly select the opposite one. Moreover, Hayes-Harb and Durham (2016) agree with this, reporting that a considerable number of studies have found that native speakers of Arabic usually rely on the rime information in CVC syllables, a type of syllable that of a consonant followed by a vowel followed by another consonant, in order to determine the onset consonant status in terms of emphatic and non-emphatic contrasts.

2.5.2.2.2 Description of the MSA Phonemes Investigated in this Study

Tibi et al. (2021) elaborates on the fact that a number of Arabic consonants share their place of articulation, manner of articulation, and are distinguished by either voicing or pharyngealization, while other consonants share their place of articulation, manner of articulation, and also voicing, and are distinguished only by pharyngealistion. All emphatic consonants have non-emphatic counterparts which are the same in all features except for the feature of pharyngealisation. This is also the case with voicing where some voiced consonants have their voiceless counterparts that are the same in all features except for the feature of voicing. However, in the case of voicing, not all voiced consonants have voiceless counterparts and not all voiceless consonants have voiced counterparts. Table 2.2 shows the categorisation of emphatic and their non- emphatic counterparts, as well as voiced and voiceless counterparts.

The current research investigates the perception of six MSA phonemes. These are /ð/ and /θ/, /sˤ/ and /s/, and /ʁ/ and /χ/. L1/L2 phoneme similarity (Arabic and English) was taken in consideration as to which phonemes were categorised into three groups: identical phonemes, similar phonemes, and new phonemes. Identical phonemes are /ð/ and /θ/, meaning that they have identical counterparts in English (participants' L1).

Table 2.2: *Voiced and Voiceless, Emphatic and Non-emphatic Phonemes in MSA*

Place/ manner of articulation	Voiced phoneme	Voiceless counterpart	emphatic phoneme	Non-emphatic counterpart
Fricative	/z/	/s/	/sˤ/	/s/
Alveodental	/ð/	/θ/	/ðˤ/	/ð/
Plosive	/d/	/t/	/tˤ/	/t/
Alveodental	/dˤ/	/tˤ/	/dˤ/	/d/
Fricative Pharyngeal	/ʕ/	/ħ/		
Fricative Uvular	/ʁ/	/χ/		

According to Almahmoud (2013), these two phonemes are produced with a slight difference from those in English. In MSA, they are produced by placing the tip of the tongue between the upper and lower teeth and are, therefore, referred to as inter-dental phonemes. In English, however, they are produced by placing the tip of the tongue at the back of the upper front incisors and are, therefore, referred to as dentals. However, this difference is very slight which is why they are categorised in this study, as well as many other studies (e.g., Almahmoud, 2013), as identical. Similar phonemes, on the other hand, are /sˤ/ and /s/ in which /sˤ/ has a similar counterpart in English which is /s/. New phonemes are /ʁ/ and /χ/ which are novel to L1 English speakers. By looking at Table 2.2, it is apparent that both members of each pair are identical in all features except for one: voicing in both /ð/ and /θ/, and /ʁ/ and /χ/, and pharyngealisation in /sˤ/ and /s/. However, there is another consonant that is identical to /ð/ in voicing, but is different in pharyngealisation. This consonant is /ðˤ/. Therefore, these two phonemes are presented in the same figure (Fig. 2.11). Alghamdi (2015) provides three figures for the six MSA phonemes. Figure 2.11 shows how the interdental fricative phonemes /ð/, /θ/ and /ðˤ/ are articulated. Figure 2.12 shows how the alveodental fricative empahtic /sˤ/ and its non-emphatic counterpart /s/ are articulated. Figure 2.13 shows how the uvular fricatives /ʁ/ and /χ/ are articulated. In Figures 2.11 and 2.12, the dotted line indicates the position of the tongue in the emphatic phonemes /ðˤ/ and /sˤ/, respectively.

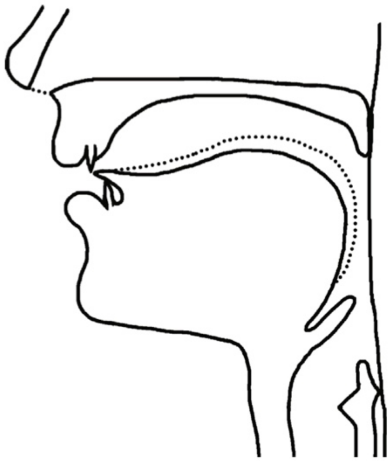


Figure 2.11 Articulation of /ð/, /θ/ and /ðˤ/
(Alghamdi, 2015)

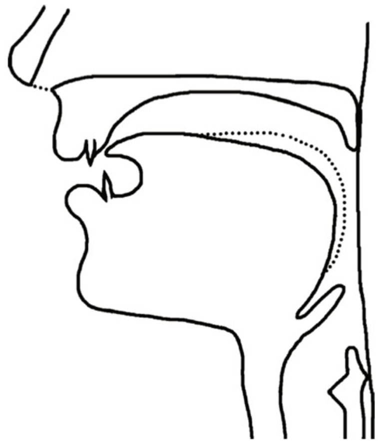


Figure 2.12 Articulation of /sˤ/ and /s/
(Alghamdi, 2015)

Figure 2.13 Articulation of /ʁ/ and /χ/
(Alghamdi, 2015)

2.5.2.3 Syllabic Structure in MSA.

MSA allows a certain syllabic structure including the following clusters: onset and nucleus such as CV /fi:/ 'in'; onset, nucleus and coda such as CVC /dʒad/ 'grandfather' and CVCC /baħr/ 'sea'. In MSA syllables, all nuclei must have a vowel and the vowel can be either short or long (Saiegh-Haddad & Henkin-Roitfarb, 2014). These examples indicate that vowels are not allowed to take initial positions in MSA syllables. In addition, this indicates that onset consonant clusters are also not allowed in MSA. This can be summarised as follows:

- Every syllable must have a vowel.
- There is only one consonant in the onset.
- MSA syllables can be open (end with a vowel) or closed with one or two consonants.

Table 2.3 shows some examples of final consonant clusters occurring in the coda (Bani Salameh, 2015).

Table 2.3: *Final Consonant Cluster in MSA (Bani Salameh, 2015)*

Symbol	Arabic example	English meaning
/tl/	/qatl/	‘killing’
/lb/	/qalb/	‘heart’
/xl/	/naχl/	‘tree dates’ pl.
/rb/	/dˤarb/	‘beating’
/hb/	/nahb/	‘robbery’
/ħr/	/siħr/	‘sorcery’
/hr/	/sˤihr/	‘son-in-law’, ‘brother-in-law’
/lm/	/silm/	‘peace’
/sm/	/dˤaχm/	‘big’, ‘large’, ‘magnificent’
/rθ/	/ħarθ/	‘ploughing’, ‘tilling’

2.5.2.4 Stress in MSA

Stress in Arabic is not phonemic, meaning that the stress placement does not lead to a change in meaning (Holes, 2004). It has a stable placement in MSA. The rules of stress placement can be summarised as follows: if a word has more than one short vowel, stress occurs on the first one; if a word has more than one long vowel, stress occurs on the last vowel; and if it has only one long vowel, this vowel is stressed (Saiegh-Haddad & Henkin-Roitfarb, 2014).

2.5.2.5 Consonants and Vowels in MSA and English

English and MSA belong to two different language families, which in turn enlarges the number of differences between the two languages. English is a member of the Germanic group which belongs to the Indo-European languages. Arabic belongs to the Semitic language family (Shariq, 2015).

The difference between the English and MSA phonological systems in terms of their phonemes occurs mainly in the number of vowels. British English, which is the participants' L1 in this study, consists of 20 vowels including six short vowels, five long vowels and nine diphthongs (Ladefoged & Ferrari, 2012). MSA has only six monophthongs including three short vowels and three long vowels (Saiegh-Haddad & Henkin-Roitfarb, 2014). Figure 2.14 shows the vowel inventory of British English (Received Pronunciation, (RP)), which can be compared with Figure 2.5. which presents the vowel inventory of MSA. Another difference between the vowels in English and MSA occurs on the boundaries of the formant space (Ladefoged & Ferrari, 2012). As discussed in subsection 2.5.2.2.1, F2 is found to be lower in the vowel /u/ in MSA. However, in most dialects of English, F2 is found to be higher in this vowel (Ladefoged & Ferrari, 2012). Furthermore, the English vowel /ɑ/ is very similar to the pharyngealised Arabic allophone [ɑ], in that both are low back vowels where a similar F2 position is occupied and F2 is found to be lower than the other vowels (Binasfour, 2018; Hayes-Harb & Durham, 2016). On the other hand, MSA and English have a common feature in vowels which can be seen in the fact that they may both depend on the vowel quantity and not only formant frequencies as an essential way of distinguishing vowels (Ladefoged & Ferrari, 2012).

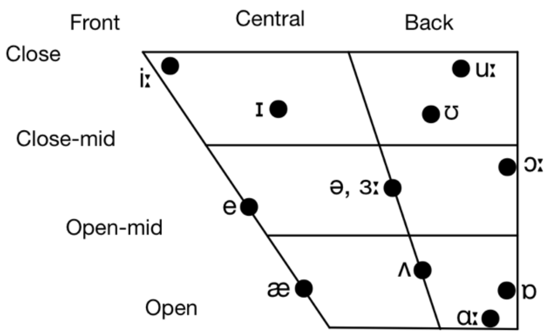


Figure 2.14 Inventory of British English *vowels (Roach, 2004)*

Consonants in English, on the other hand, have articulatory and acoustic similarities to and differences from MSA consonants. English has 25 consonants and MSA has 28. Table 2.4 presents a summarised list of English consonants with their corresponding consonants in MSA in terms of their place and manner of articulation. By comparing the two languages in terms of position occupation, one of the similarities that these two languages share is that most of the consonants that exist in both languages share the same positions whereby most of them occur initially, medially, and finally except /j/, which does not occur in final position in either English or MSA. One of the dissimilarities, however, occurs in /r/ which does not take final position in Standard British English while it takes all three positions in MSA. /ʔ/, /h/, and /w/, also, do not occur in all three positions in Standard British English in which the final position cannot be occupied by these three phonemes in Standard British English but can in MSA (Shariq, 2015). It is worth noting that /ʔ/ is an allophone of /t/ in Standard British English but not a phoneme in its own. Moreover, MSA differs from English in the presence of pharyngealisation that does not exist in English except in the allophone /ɫ/ (Recasens, 2004 as cited in Binasfour, 2018) which has a different manner of articulation from those pharyngealised in MSA (Binasfour, 2018). Furthermore, the phoneme /r/ has different realizations in the two languages. The Arabic /r/ is realised as a dental-alveolar tap (Shariq, 2015; Watson, 2002), but the English /r/ as a non-rhotic post-alveolar approximant (Roach, 2000).

2.5.3 Orthography of MSA

Arabic graphemes were augmented from the Aramaic alphabet to represent additional Arabic consonants (Saiegh-Haddad & Henkin-Roitfarb, 2014). Arabic has an Abjad writing system (Tibi, Edwards, Kim, Schatschneider & Boudelaa, 2022). Abjad refers to a writing system where every consonant is represented by a single symbol and vowels are supplied by the reader depending on the context (Saiegh-Haddad & Henkin-Roitfarb, 2014). Despite short vowels not always being present, long vowels are non-omittable and must be written as letters, making Arabic an impure abjad (Zitouni, 2014). Arabic graphemes are shown in Table 2.5.

Saiegh-Haddad and Henkin-Roitfarb (2014) elaborate on the Arabic orthographic system in that it follows three main rules. First, unlike many languages, Arabic is written from right to left. Second, Arabic graphemes are written joined to preceding and sometimes following graphemes. However, this rule has a number of exceptions which are presented in detail in the following

Table 2.4: *MSA and English Consonants (Alghamdi, 2015; Roach, 2000)*

Manner of Articulation	Place of Articulation	MSA consonants	English consonants
Plosives	Bilabial	/b/	/b/, /p/
	Alveodental	/d/, /t/, /dˤ/, /tˤ/	/d/, /t/
	Velar	/k/	/k/, /g/
	Uvular	/q/	
	Glottal	/ʔ/	/ʔ/
Fricatives	Labiodental	/f/	/f/, /v/
	Interdental	/θ/, /ð/, /ðˤ/	/θ/, /ð/
	Alveodental	/s/, /sˤ/, /z/	/s/, /z/
	Alveopalatal	/ʃ/	/ʃ/, /ʒ/
	Uvular	/χ/, /ʁ/	
	Pharyngeal	/ħ/, /ʕ/	
	Glottal	/h/	/h/
Affricates	Alveopalatal	/dʒ/	/dʒ/, /tʃ/
Nasals	Bilabial	/m/	/m/
	Alveodental	/n/	/n/
	Velar		/ŋ/
Approximants	Bilabial	/w/	/w/
	Palatal	/j/	/j/
	Alveopalatal		/r/
Laterals	Alveodental	/l/	/l/
Trills	Alveodental	/r/	

Table 2.5: *Arabic Graphemes*

Arabic grapheme	IPA symbol	Arabic grapheme	IPA symbol
ا	/aː/	ز	/z/
آ	/aː/	س	/s/
أ	/ʔ/	ش	/ʃ/
إ	/ʔ/	ص	/sˤ/
ئ	/ʔ/	ض	/dˤ/
ؤ	/ʔ/	ط	/tˤ/
ء	/ʔ/	ظ	/ðˤ/
ى	/aː/	ع	/ʕ/
ة	/t/	غ	/ʁ/
ب	/b/	ف	/f/
ت	/t/	ق	/q/
ث	/θ/	ك	/k/
ج	/dʒ/	ل	/l/
ح	/ħ/	م	/m/
خ	/χ/	ن	/n/
د	/d/	ه	/h/
ذ	/ð/	و	/w/
ر	/r/	ي	/j/

paragraph. Third, only long vowels are written as graphemes. Short vowels are not, but are sometimes represented as diacritics. These short vowels are three, as mentioned above. The first one is *fatħa* [◌َ] which stands for opening the lips relatively wide, it represents the short vowel /a/, and is written above the letter. The second one is *kasra* [◌ِ] which stands for a slight spreading of the lips, it represents the short vowel /i/, and is written below the letter. The third one is *dˤamma* [◌ُ] which stands for a slight rounding of the lips, it represents the short vowel /u/, and is written above the letter (Saiegh-Haddad & Henkin-Roitfarb, 2014).

Arabic graphemes are divided into three groups in terms of their form in orthography. One grapheme is always written in the base form and is not joined to any other grapheme, regardless of its position in the word. This grapheme is ء /ʔ/. Some graphemes are joined only to preceding graphemes, not to following ones. These graphemes are ا , آ , أ , إ , ؤ , ى , ة , ـد , ـذ , ـر , ـز, and ـو. Thus, these graphemes have only two forms: the base form and joined-to-the-previous-grapheme form. The third group consists of graphemes that are joined to preceding and/or following graphemes. These graphemes are ئ, ـب, ـت, ـث, ج, ح, خ, ـس, ـش, ـص, ـض, ـط, ـظ, ع, غ, ـف, ق, ك, ل, م, ن, ه, and ي. Therefore, these graphemes have four different forms; the base form, joined-to-the-preceding-grapheme form, joined-to-the-following-grapheme form, and between-two-graphemes form (Saiegh-Haddad & Henkin-Roitfarb, 2014; Tibi & Kirby, 2018). Table 2.6 presents a complete list of these orthographic forms of each Arabic grapheme. It is important here to highlight that Arabic does not allow the graphemes ى and ة to occur initially or medially in a word, they just take their place in the final position of a word (Alghamdi, 2015).

Table 2.6: *Arabic Graphemic Forms (Alghamdi, 2015)*

Base form	IPA symbol	Joined-to-the-following grapheme	Between-two-graphemes	Joined-to-the-preceding grapheme
ا	/aː/	ا	ـا	ـا
آ	/aː/	آ	ـآ	ـآ
أ	/ʔ/	أ	ـأ	ـأ
إ	/ʔ/	إ	ـإ	ـإ
ئ	/ʔ/	ئـ	ـئـ	ـئ
ؤ	/ʔ/	ؤ	ـؤ	ـؤ
ء	/ʔ/	ء	ء	ء
ى	/aː/			ـى
ة	/t/			ـة
ب	/b/	بـ	ـبـ	ـب
ت	/t/	تـ	ـتـ	ـت
ث	/θ/	ثـ	ـثـ	ـث
ج	/dʒ/	جـ	ـجـ	ـج
ح	/ħ/	حـ	ـحـ	ـح
خ	/χ/	خـ	ـخـ	ـخ
د	/d/	د	ـد	ـد
ر	/r/	ر	ـر	ـر
ز	/z/	ز	ـز	ـز
س	/s/	سـ	ـسـ	ـس
ش	/ʃ/	شـ	ـشـ	ـش
ص	/sˤ/	صـ	ـصـ	ـص

Table 2.6 cont.: *Arabic Graphemic Forms (Alghamdi, 2015)*

Base form	Base form	Base form	Base form	Base form
ض	/dˤ/	ضـ	ـضـ	ـض
ط	/tˤ/	طـ	ـطـ	ـط
ظ	/ðˤ/	ظـ	ـظـ	ـظ
ع	/ʕ/	عـ	ـعـ	ـع
غ	/ʁ/	غـ	ـغـ	ـغ
ف	/f/	فـ	ـفـ	ـف
ق	/q/	قـ	ـقـ	ـق
ك	/k/	كـ	ـكـ	ـك
ل	/l/	لـ	ـلـ	ـل
م	/m/	مـ	ـمـ	ـم
ن	/n/	نـ	ـنـ	ـن
ه	/h/	هـ	ـهـ	ـه
و	/w/	و	ـو	ـو
ي	/j/	يـ	ـيـ	ـي

2.5.3.1 The Role of Diacritics in Arabic

Despite the fact that Arabic orthography is transparent, which means that there is high one-to-one grapheme-phoneme correspondences, there are a number of Arabic sounds that do not have graphemic representations (Saiegh-Haddad & Henkin-Roitfarb, 2014). By way of illustration, Arabic orthography does not represent gemination and short vowels as letters, but as diacritics instead (Saiegh-Haddad & Henkin-Roitfarb, 2014). However, users and writers of Arabic do not always write diacritics because their pronunciation can be predicted from the context. Usually, Arabic context includes diacritics in the case where it is targeting beginners or young learners, or in some literary and religious forms (Saiegh-Haddad & Henkin-Roitfarb, 2014). This sometimes leads to homographic words. For example, the word , كتب which can be read when written without diacritics as /kataba/ 'he wrote', /kutub/ 'books', and /kutiba/ 'it was written', while it is written with diacritics as follows, respectively: ,كَتَبَ , كُتُب and . كُتِب (Saiegh-Haddad & Henkin-Roitfarb, 2014; Simon et al., 2006; Tibi & Kirby, 2018). Because Arabic morphology depends on root-pattern interaction, which is the case in a number of Sematic languages (Holes, 2004), the function of diacritics is not restricted to semantic functions but also has other functions such as using the silence diacritic (sukun) [ْ] to indicate the ending case which occurs usually at the end of an utterance (Mubarak, Abdelali, Darwish, Eldesouki, Samih, & Sajjad, 2019; Simon et al., 2006). Saiegh-Haddad and Henkin-Roitfarb (2014) state that diacritics in Arabic script are divided into two sets. The first set is graphemic and represented by dots above or below the letter, it is used to distinguish between consonants, such as ب /b/, ت /t/, and ث /θ/ which are distinguished by the number and position of the dots (one for /b/ below the letter, two for /t/ above the letter, and three for /θ/ above the letter). According to Tibi et al. (2022), due to this characteristic, Arabic is considered as having high degree of visual similarity between its graphemic symbols, which is found to cause difficulties for L1 Arabic readers (Tibi et al., 2021). This is because 22 letters can be categorised into visual similar graphemes that are distinguished by having either one, two, or three dots (e.g., ب، ت، ث). The second set of diacritics, according to Saiegh-Haddad and Henkin-Roitfarb (2014), is phonemic which represents some vocalization features along with short vowels, it is used to enhance semantic functions depending on the position of a word in a sentence. This feature of enhancing semantic functions indicates that there are up to four allographs for every single letter in Arabic. These allographs differ depending on the position of a letter in a word

(Simon et al., 2006). Alhumaid (2019, p.31) provides an example of how these diacritics have a great role in changing the meaning of words:

> “For instance, the word درس may have more than one meaning as it is written without diacritics, whereas when the diacritics are added, it means as follows: دَرْس /dars/ ‘a lesson’, دَرَسَ /darasa/ ‘he learned’, دُرِسَ /durisa/ ‘it was learned’, دَرّسَ /dar rasa/ ‘he taught’ and دُرِّسَ /dur risa/ ‘it was taught’. In the previous example, the root word is دَرَسَ /darasa/, and all the other forms are different derivations of this root”

2.5.3.2 Opaqueness of Arabic Orthography

Despite the fact that it has been observed that Arabic orthography is transparent, there are a number of features that reflect Arabic orthographic opaqueness. Saiegh-Haddad & Henkin-Roitfarb (2014) provide several examples to elaborate on this. To start with, there are some different graphemes that may represent the same sound. For instance, if the vowel /a:/ occurs in final position in an Arabic word, it might be represented by either the grapheme ‘ا’ or the grapheme ‘ى’. This largely depends on the derivational etymology of the word. Saiegh-Haddad & Henkin-Roitfarb (2014) also elaborate on how the Hamza (glottal stop) is represented in Arabic. The position of this phoneme and the surrounding vowels in a word determine how this phoneme is represented. If this phoneme is followed by the short vowel /i/, it is represented by the grapheme ئ, if it is followed by the vowel /u/, it is represented by the grapheme ؤ, and if it is followed by the vowel /a/, it is represented by the grapheme أ.

The opaqueness of Arabic orthography can also be reflected by the fact that there are some graphemes that are pronounced differently. By way of illustration, the two graphemes ‘و’ and ‘ي’

might be pronounced as consonants /w/ and /j/ or as long vowels /u:/ and /i:/, respectively. This can be demonstrated by the fact that the approximate consonants /w/ and /j/, sometimes referred to as "semi-vowels" (Saiegh-Haddad & Henkin-Roitfarb, 2014), are consonantal when they occur in the initial position of a word or syllable such as و /w/ in ورد /ward/ 'flower' and ي /j/ in يد /jəd/ 'hand', but represent vowels when the preceding consonant is marked by a diacritic that represents a similar shorter vowel such as و /u:/ in قُلوب /qulu:b/ 'hearts' and ي / i:/ in بدِين /bədi:n/ 'fat' (Saiegh-Haddad & Henkin-Roitfarb, 2014). Thus, these grapheme-phoneme correspondences are inconsistent, but predictable, according to the position in the word that a grapheme occupies.

Furthermore, having different pronunciations for the same grapheme might be caused by cases such as pause and juncture. The suffix 'ة', for example, which marks the feminine gender in Arabic, is sometimes pronounced as /h/ and at other times as /t/. The direct and clear rule for these two pronunciations is: it is pronounced as /h/ when the suffix has a pre-pausal position, and it is pronounced as /t/ when it is in a juncture context. Alhumaid (2019, p.32) provides examples of these two cases:

> " The word وردة 'flower' is pronounced as /wardah/ when it occurs in a pre-pausal position, such as سقيت وردة/saqeitu wardah/ 'I watered a flower', whereas it is pronounced as /wardat/ when it precedes another word, such as سقيت وردة الجوري /saqeitu wardat əldʒu:ri:/ 'I watered the rose flower'. In this context, it precedes the wordالجوري /əldʒu:ri:/ 'rose' which affects the pronunciation of the final suffix < ة > and changes it to /t/ instead of /h/ "

2.5.3.3 Arabic Orthographic Representations of the Phonemes Investigated in this Study

Table 2.7 presents how the phonemes investigated in this study are represented orthographically in two forms: base form and initial-position form.

Table 2.7: Orthographic Representations of the Phonemes Investigated in this Study

Pair	Phoneme 1	Base form	Initial-position form	Phoneme 2	Base form	Initial-position form
Identical phonemes	/ð/	ذ	ذ	/θ/	ث	ثـ
Similar phonemes	/sˤ/	ص	صـ	/s/	س	سـ
New phonemes	/ʁ/	غ	غـ	/χ/	خ	خـ

By comparing the graphemic symbols of the phonemes in Table 2.7, it can be seen that, while some of the symbols look similar, there are differences in each case.

2.5.4 Summary of the Section

This section presented a detailed description of the Arabic language, particularly MSA. It elaborated on how MSA phonology is structured, including a deeper discussion of how some phonemes are recognised in relation to each other. In addition, it briefly touched upon a comparison between the Arabic and English phonological systems. This section also demonstrated the basic rules of MSA orthography, touching upon the opaqueness of the orthographic system of Arabic.

2.6 Chapter Summary and Research Questions

2.6.1 Summary of the Chapter

As reviewed in previous sections, over the last two decades, the link between orthographic input and L2 phonological development has been at the centre of much attention. Studies have involved different methodologies including various types of orthographic input and/or different levels of orthographic depth and many other different factors, such as the perceived difficulty of target phonemes. Therefore, studies have concluded with contradictory results (as mentioned in Section 2.4). However, there are several other factors that may modulate the role of orthography in L2 phonological development which remain largely unexamined. These include learners' internal factors such as individual differences, and/or external factors such as L1/L2 phoneme similarity. Too little attention, to the best of the researcher's knowledge, has been paid to link individual differences, and/or L1/L2 phoneme similarity in order to investigate the role of orthography in L2 phonological development. Thus, the aim of this investigation was to explore the relationship between the effects of exposure to orthographic input (in terms of orthography availability and/or script familiarity) and L1/L2 phoneme similarity, as well as learners' phonemic coding ability, which is a component of language aptitude, in learners' ability to discriminate the contrasts between these L2 phonemes. This emphasises the gap that the current study aimed to bridge. This is because a systematic understanding of how phonemic coding ability and L1/L2 phoneme similarity contribute to the effect of orthographic input in L2 phonological development is still lacking. Therefore, the study sought to answer the following five specific research questions:

RQ1: Does the presence of orthography help participants differentiate between two phonemes in the case of:

- An unfamiliar script such as Arabic script; or
- A Combination of Familiar Script such as Romanised-Arabic Script and *Unfamiliar Script such as Arabic Script?*

RQ2: Does L1/L2 phoneme similarity affect the difficulty of participants' ability to differentiate between two phonemes?

RQ3: Is the effect of orthographic availability influenced by L1/L2 phoneme similarity?

RQ4: Does the participants' ability to discriminate phonemes when embedded within words (word-level discrimination) correlate with discriminating phonemes when uttered in isolation (phoneme-level discrimination)?

RQ5: Is the effect of orthographic availability influenced by the combination of phonemic coding ability and L1/L2 phoneme similarity?

2.6.2 Explanation of Research Questions

***RQ1:* Does the Presence of Orthography Help Participants Differentiate between two Phonemes in the Case of:**

- **An Unfamiliar Script such as Arabic Script; or**
- **A Combination of Familiar Script such as Romanised-Arabic Script and Unfamiliar Script such as Arabic Script?**

The role of orthographic input during L2 phonological development has been investigated in a number of recent studies (e.g., Bassetti, 2006; Bassetti, 2017; Bassetti et al., 2018; Bassetti et al., 2020; Cerni et al., 2019; Erdener & Burnham, 2005; Escudero, 2015; Escudero et al., 2008; Escudero et al., 2014; Hayes-Harb et al., 2010; Mathieu, 2016; Showalter & Hayes-Harb, 2013, 2015). The findings of these studies, however, are at odds with each other; some found that being exposed to orthography helps in L2 phonological development (Escudero et al., 2008; Showalter & Hayes-Harb, 2013), some found it hinders it (Bassetti, 2006; Hayes-Harb et al., 2010), and some found it has no effect (Mathieu, 2016; Showalter & Hayes-Harb, 2015) or has a two-sided effect, meaning that it helps in some cases and hinders or has no effect in others (Bassetti, 2008; Bassetti & Atkinson, 2015; Erdener & Burnham, 2005; Escudero, 2015; Nimz & Khattab, 2019; Simon et al. 2006). It is important to clarify that not all of these studies targeted L2 learners; i.e., some were investigating L2 learners (e.g., Bassetti, 2006; Bassetti & Atkinson, 2015; Nimz & Khattab, 2019) and some recruited naïve listeners (e.g., Erdener & Burnham, 2005; Escudero, 2015; Mathieu, 2016; Showalter & Hayes-Harb, 2013;2015). Bassetti (2008) argues that a number of factors may play a role in modulating the effect of orthographic input on L2 phonological development. One of these factors might be the characteristics of the writing systems of L1 and L2, including the type of writing system. Therefore, *RQ*1 consists of two main parts that include two types of orthographic input: familiar and unfamiliar. This is because Arabic (the target language in this study) has a totally different script from English (the first language of the participants in this study) in terms of directionality (Arabic is written from right to left) and graphemic shapes (as discussed in detail in Section 2.5).

*RQ*1 aimed to explore the role of written forms in terms of orthographic availability and/or script familiarity in discriminating specific phonemes and differentiating between minimal pairs contrasting initially in one of the target phonemes. In an attempt to answer the core concept of *RQ*1, which is the effect of the availability of orthographic input during L2 phonological acquisition, the input modality had to be different among three different groups: a group with no orthographic input at all, and two groups with orthographic input: one with unfamiliar script (Arabic) and the other with combined familiar and unfamiliar scripts: Romanised and Arabic. In order to answer the first part of *RQ*1, one group of three was provided with Arabic script, which was totally unfamiliar to the participants. However, according to Mathieu (2014, 2016) and Showalter and Hayes-Harb (2015), the difficulty and novelty of L2 script (Arabic in all three studies) resulted in a lack of orthographic effect on L2 phonological development. This led Showalter and Hayes-Harb (2015) to suggest that it might be more helpful to provide participants with a combination of both familiar (Romanised) and novel (Arabic) scripts. Thus, the third group was provided with two types of script: Romanised accompanied by Arabic script, which was done in order to answer the second part of *RQ*1.

***RQ2:* Does L1/L2 phoneme similarity affect the difficulty of participants' ability to differentiate between two phonemes?**

Flege (1995) suggests that the difficulty in acquiring similar phonemes which contrast in L2 but not in L1 occurs due to learners' classification of these phonemes as equivalent to a single L1 phoneme category. This is not the case with new phonemes that do not occur in L1 at all. These seem to be easier because of learners' ability to easily notice the differences between L1 and L2. Flege (1986) refers to a phoneme similarity process, what he calls "equivalence classification". According to Flege's (1995) fifth hypothesis, equivalence classification might lead the learner to block the category formation of an L2 sound. This results in having one single phonemic category for both linked L1 and L2 phonemes. To test Flege's fifth hypothesis, the audio stimulus has to include phonemes that have different levels of similarity. Thus, based on Flege's equivalence classification, three pairs of L2 phonemes are included in this study. The first pair is an identical pair which consists of /θ/ and /ð/ that correspond to two phonetic categories in English as well as in Arabic. The second pair is a similar pair which consists of the Arabic phonemes /s/ and /sˤ/ that correspond to only one phonetic category in English, /s/. The third pair of Arabic sounds is a new pair which consists of /χ/ and /ʁ/ that do not have any English phonetic category correspondences. This categorization is based on the acoustic features of these phonemes in Arabic and English (as discussed in Section 2.5).

*RQ*2 aimed to examine the effect of English and Arabic phoneme similarity on the participants' ability to discriminate between two L2 phonemes. For this purpose, the design of this study is guided by this objective and the experiment is divided into three conditions, with each condition being dedicated to the acquisition of one pair: the first condition is for /θ/ and /ð/, second condition is for /s/ and /sˤ/, and the last condition is for /χ/ and /ʁ/. Every condition is tested in a different session. In order to answer *RQ*2, participants were asked to discriminate between the two phonemes in each pair in a number of perception tasks (this is discussed in Subsection 3.3.5).

***RQ3:* Is the Effect of Orthographic Availability Influenced by L1/L2 Phoneme Similarity?**

According to Bassetti (2008), the interpretation of orthographic input is sometimes done depending on L1 orthography-phonology correspondences, which in turn interact with acoustic input to form the phonological representation of learners. This interaction may lead to affecting phonological awareness tasks and perception. This is supported by Bassetti's (2017) finding that L2 learners may establish two phonological categories in which both are different from their L1 category depending on the orthographic representation of these phonemes. Bassetti (2017) suggests that the current dominant models of L2 phonological development, such as the SLM (Flege, 1995) and PAM-L2 (Best & Taylor, 2007), can be associated with orthographic input during L2 phonological development as it is found that L2 speakers produce phonological contrasts depending on the orthographic representation.

*RQ*3 aimed to examine the interaction between orthographic input effect and L1/L2 phoneme similarity in the participants' ability to discriminate between L2 phonemes. To answer *RQ*3, all participants' performance on perception tasks in three orthographic-input groups was compared in three different conditions of L1/L2 phoneme similarity. These conditions include an identical pair in the first condition, a similar pair in the second condition, and a new pair in the third condition.

***RQ4:* Does the Participants' Ability to Discriminate Phonemes when Embedded Within Words (Word-Level Discrimination) Correlate with Discriminating Phonemes when Uttered in Isolation (Phoneme-Level Discrimination)?**

According to Escudero (2015), L2 learners' ability to discriminate phonemes might not be the same when occurring within words or uttered in isolation. She claims that it might happen that L2 learners are able to discriminate lexical representations that include perceptually difficult L2 contrasts, and at the same time they might not be able to distinguish these phonemes when they are uttered in isolation. According to Escudero (2015), this inability to distinguish contrasted phonemes leads to having a discontinuity between perceptual and lexical performance. This refers to the case where learners can distinguish words in the lexicon without the ability to notice the phonological contrasts in these words.

*RQ*4 aimed to test Escudero's (2015) claim by testing the participants' ability to distinguish target phonemes at both discrimination levels: word-level discrimination and phoneme-level discrimination. To answer *RQ*4, data were collected from two different sources in the testing phase. The first phase is the Phonological Test Phase, which asked participants to decide which image represents the word they heard in which this image either represents the word they heard or another word that was contrasted only in the other phoneme in the target pair. The second phase is the

Post-Sound Recognition Phase which targeted phoneme-level discrimination and included target phonemes uttered in isolation, participants were asked to decide whether these phonemes sound different or the same.

***RQ5:* Is the Effect of Orthographic Availability Influenced by the Combination of Phonemic Coding Ability and L1/L2 Phoneme Similarity?**

According to Meara (2005) and Carroll (1993), high phonemic coding ability provides learners with the ability to structure words into smaller phonetic units which gives them the opportunity to analyse words and divide them into smaller forms, including pronunciation and morphological rules and meaning aspects. Therefore, it is argued that the role of phonemic coding ability in L2 learners can be seen in the fact that it helps L2 learners to develop input processing strategies and consequentially be able to recognise and integrate new linguistic units (Reynolds, 2002). According to Carroll (1962), learners who have lower phonemic coding ability, on the other hand, are found to have difficulty in remembering phonetic form as well as mimicking speech sounds (Bassetti et al., 2020). Bassetti (2008) argues that there are some factors that modulate the effect of orthographic input on L2 phonological development which may include learners' internal factors. Bassetti (2008) suggests investigating the role of the phonemic coding ability of learners in their reliance on orthographic input during L2 phonological development.

*RQ*5 aimed to link the participants' cognitive abilities, specifically phonemic coding ability, to their performance in perception activities. It attempted to investigate learners' performance on a phonemic coding ability test (Llama-E) and its role in predicting their performance on discriminating contrasts in minimal pairs. It aimed to explore the interaction between phonemic coding ability, L1/L2 phoneme similarity, and orthographic input type taken together on participants' ability to discriminate between two L2 contrasts.

In order to answer *RQ*5, participants were allocated into three orthography groups: Arabic-script, Romanised-script, and No-script groups, in which each orthography group had a range of different scores on the Llama-E test. The performance of the participants in their phonemic coding ability test (Llama-E test) was analysed as a predictor to measure the relationship between their performance in this test with their performance in the three phoneme similarity conditions and in all three orthographic input-type groups.

3 Methodology

3.1 Introduction

This chapter presents a detailed description of an exploratory study which aimed to explore the extent to which some factors might affect L1 English participants' ability to benefit from L2 orthographic input during their Arabic L2 phonological learning. These factors included script availability and familiarity in which some participants had no script at all, some had a novel script (Arabic) along with a familiar script (Romanised), and some only had a novel script (Arabic). The second factor was English and Arabic phoneme similarity in which the target phonemes consisted of three pairs: a pair consisted of the two Arabic interdental phonemes, /θ/ and /ð/, which are identical to English phonemes and therefore map to two English phonemes, a pair consisted of the two Arabic alveodental phonemes, /s/ and /sˤ/, which are similar to each other but not a separate category in English and therefore have only one counterpart in English, and a pair consisted of the two Arabic uvular phonemes, /ʁ/ and /χ/, which are entirely new phonemes that do not exist in English. The last factor was the level of the participants' phonemic coding ability in which participants had scores varying between 0 to 20 on a phonemic coding ability test. These three factors were linked in this study design in order to investigate their role in affecting the success of the discrimination between some Arabic phonemes that share some of their acoustic and articulatory features with English.

The study was built on the basis of providing participants with different activities designed to test their perception ability of some Arabic phonological pairs and their ability to discriminate the contrasts between these pairs. These activities were sound recognition activities, Arabic and Arabic-like minimal pairs discrimination, Arabic non-word learning in which participants had to learn to identify novel objects, along with learning novel labels (non-words) for the objects. All the words and non-words included target sounds placed in word initial position. According to Chan (2012), L2 perception is affected by the position of consonants in a word. It is found that initial consonants are more perceptible than final consonants because the former are more perceptually salient than the latter (Redford & Diehl, 1996 as cited in Chan, 2012). The study was divided into different stages and phases including a Pre-Sound Recognition Phase, Word Learning Phase, Criterion Test Phase, Phonological Test Phase, and Post-Sound Recognition Phase which are all discussed in more detail in Subsection 3.3.4.

The details of how this study was designed in terms of the research design, participant recruitment, types of stimuli, ethical considerations, instruments, procedures, and methods of data collection and analysis are presented in this chapter. It also includes a number of justifications for selecting some particular methods used in this study.

3.2 Research Questions

As previously mentioned, the following research questions were investigated:

RQ1: Does the presence of orthography help participants differentiate between two phonemes in the cases of:

- An unfamiliar script such as Arabic script; or
- A Combination of Familiar Script such as Romanised-Arabic Script and Unfamiliar Script such as Arabic Script?

RQ2: Does L1/L2 phoneme similarity affect the difficulty of participants' ability to differentiate between two phonemes?

RQ3: Is the effect of orthographic availability influenced by L1/L2 phoneme similarity?

RQ4: Does the participants' ability to discriminate phonemes when embedded within words (word-level discrimination) correlate with discriminating phonemes when uttered in isolation (phoneme-level discrimination)?

RQ5: Is the effect of orthographic availability influenced by the combination of phonemic coding ability and L1/L2 phoneme similarity?

3.3 Study Design

The study followed an experimental design. It was conducted to measure the effect of multiple independent variables on a single dependent variable. The primary dependent variable in all the analyses was phoneme discrimination which was assessed using two phoneme discrimination levels: word level and phoneme level, and two measures: accuracy and reaction time. This was done by comparing the performance of different groups with each other where these groups shared the same aspects except that the treatment being administered was manipulated (Dörnyei, 2007). There were three independent variables: the level of similarity between L1 and L2 phonemes, the type of orthographic input, and phonemic coding ability scores. The effect of these variables on phonological acquisition was investigated using different phonological discrimination tasks. According to Johnson and Christensen (2004 as cited in Dörnyei, 2007), having a controlled environment and having the ability to manipulate the experimental process by manipulating an independent variable among the groups while keeping all the other variables identical is an essential feature of the experimental design. Dörnyei (2007) supports this argument highlighting that this feature of manipulating the independent variable is the main reason for allocating participants into different groups because it helps researchers to isolate the target variable. Dörnyei also adds that the experimental design method is the best one to build a cause-effect relationship, which is a very common aim in applied linguistics research.

This study was designed to collect data using a quantitative method approach. According to Dörnyei (2007), a quantitative method approach starts by identifying problems which are analysed by proposing hypotheses and then tested by analysing the data collected using standardised procedures. This being the case in this study, a quantitative method approach was selected. Another reason for selecting a quantitative method is because the aim of this study was to collect data from large groups rather than looking at individuals, which was done by manipulating the variables among the groups. Linking different variables together, according to Hummersley and Atkinson (1995), is an essential feature of a quantitative method approach. In addition, Dörnyei (2007) highlights some strengths of quantitative research. By way of illustration, quantitative-method research tends to have an objective nature because the researcher's subjective perspective is kept away from the research data. However, it is possible to devise quantitative research to prove an outcome the researcher is biased towards, but it is easier for the researcher in quantitative research to maintain an objective nature than other types of research methods such as qualitative research. In addition, the generalizability of quantitative data usually goes beyond the context of the research. Given these features, a quantitative method approach was selected for use in this study.

This study had two main stages: an aptitude test stage which sought to examine the participants' phonemic coding ability; and a phoneme discrimination stage with three different conditions testing the effect of orthographic availability and script familiarity, as well as the effect of English and Arabic phoneme similarity on the acquisition of Arabic phonemes. These tests took place in three different sessions, as is described in full below.

3.3.1 Participants

Recruitment of the sample involved the following exclusion criteria: no participant had prior knowledge of the Arabic language; no participant was an advanced learner, that is, higher than GSCE level, or was a simultaneous bilingual of German or French. Speakers of German and French were excluded because German and French contain phonemes that are novel to the English language and were included in this study as novel phonemes for participants, (/ʁ/ and /χ/). However, bilingualism and prior instructional training are found to influence Llama-B and -F test-takers but not Llama-E (Rogers et al., 2017), which is the test used in this study, meaning that bilingualism in a language other than German or French was not taken into account. This exclusion criterion was enacted by asking participants to fill out a demographic questionnaire before conducting the study.

75 adults between 18–65 years old (M = 28.48, SD = 12.82), who were native speakers of English and only used the English language at home, took part in the study (58 females and 17 males). All the participants had no previous experience of Arabic and knowledge no higher than GCSE level of either French or German. However, some participants reported that they had learned a second language at school or in another naturalistic environment but pointed out that their level was no higher than GCSE. These languages included French, Spanish, German, Russian, Romanian, Greek, Welsh, Indonesian, Italian, Latin, Japanese, and Afrikaans.

Participants were allocated into three groups, with 25 participants in each one. Every group had participants with a mixture of different Llama-E scores (different phonemic coding ability levels) fluctuating from 0–20 (as shown in Table 3.3). This was done in an attempt to balance any confounding variables, i.e., to ensure that the participants in groups did not differ in terms of average language aptitude. (Llama-E test is discussed in more detail in Subsection 3.3.4.2).

The decision about the sample size was based on the sample size of previous studies (e.g., Bassetti, 2006; Mathieu, 2016; Showalter & Hayes-Harb, 2013) who found significant effects of orthographic input in word recognition. Bassetti (2006), who had 18 participants, found a large effect size of the orthographic effect (p <.001, $\eta p^2 = .87$). Mathieu (2016), who had 21 participants in each group, found a medium effect size of the orthographic input (p =.01, $\eta p^2 = .06$). Showalter and Hayes-Harb (2013) also found a large effect size of orthographic input (p =.03, $\eta p^2 = .18$) with 13 participants in each group.

3.3.2 Stimuli

The study included six Arabic fricatives that can occur initially in Arabic words: the two interdentals /ð/ and /θ/, the two alveolars /s/ and /sˁ/, and the two uvulars /ʁ/ and /χ/. The selection of these phonemes followed Flege's (1986) "equivalence classification" which is based on a "similarity to English phonemes" criterion using the variety of Standard British English as a reference. Two phonemes were selected because they are identical to two English phonemes and, therefore, have two counterparts in English (/ð/ and /θ/); two phonemes are similar to an English phoneme but have only one counterpart in English (/s/ and /sˁ/); and two phonemes are novel phonemes that have no counterparts in English (/ʁ/ and /χ/). These three types of phonemes were included to test the effect of L1 on the establishment of L2 phonemic categories. This is because if two L2 phonemes are similar where only one of them exists in the learner's L1, the learner tends to assimilate the two phonemes into one phonemic category (Best, 1995) and consequently one phonological category is established instead of two. Therefore, the current study consisted of new phonemes which are not mapped to any L1 phoneme to minimise the possibility of assimilation into one phonemic category (Mathieu, 2016). All the selected consonants were Arabic fricatives selected in particular due to their continuity feature (+Continuent). Binasfour (2018) argues that this feature gives the speaker the ability to produce these phonemes continuously as long as there is air in their lungs. This feature was important in recording the stimuli – especially in the Pre-Sound and Post-Sound Recognition Phases, where sounds were uttered out of context in one of the tasks, i.e., without any following or preceding phonemes. This was done to avoid the effect of any surrounding phonemes or the effect of coarticulation which might lead to sounds being perceived differently. Coarticulation refers to the case where there is an overlap in the articulatory movements from one sound to other surrounding sounds resulting in having articulatory and acoustic signals that provide information about more than one sound at the same time (Khalil & Mooshammer, 2021). Another feature that can affect the perception of a phoneme is what is known as emphasis spread (Holes, 2004). As mentioned in Chapter 2, it is found that emphatic sounds (e.g., /sˤ/) affect the surrounding vowels, changing them to pharyngealised vowels and resulting in causing these emphatic phonemes to be perceived differently (Alhumaid, 2019; Binasfour, 2018; Holes, 2004; Shar & Ingram, 2010).

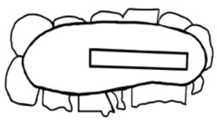
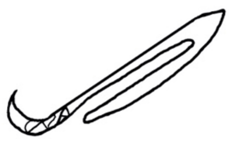
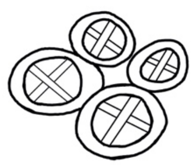
These six phonemes were embedded in 18 Arabic non-words, with the target sounds occurring word-initially. The aim of including non-words in the study was to give the researcher the chance to control some confounding variables that might affect the perception of these phonemes (Bassetti & Atkinson, 2015). These confounding variables included the surrounding vowel, discussed in the previous paragraph, and the syllabic structure which is discussed below.

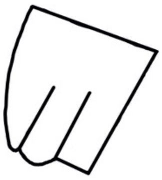
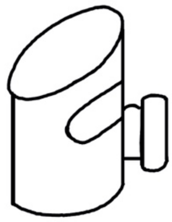
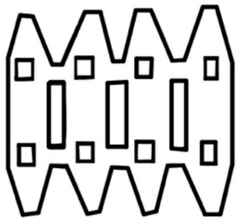
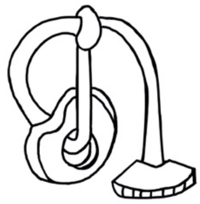
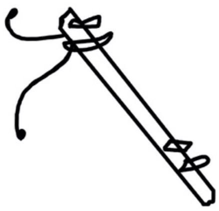
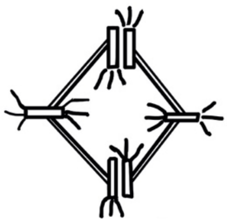
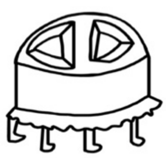
The syllabic structure of all the study lexical items was $C_1V_1C_2V_2$. As discussed in Chapter 2, it is found that F2 lowering as a result of surrounding emphatic consonants is not the same for all vowels (Binasfour, 2018; Card, 1983; Hayes-Harb & Durham, 2016). The literature has shown that the vowel /aː/ receives the greatest effect whereas the vowel /iː/ receives the lowest effect from the preceding emphatic phoneme, which results in making the surrounding emphatic consonant of the vowel /iː/ the most difficult to discriminate (Binasfour, 2018; Card, 1983; Hayes-Harb & Durham, 2016). Hayes-Harb and Durham (2016) argue that the reason for this might be because the vowel /aː/ has two corresponding English vowels, /æ/ when following non-emphatic phonemes, and /aː/ when following emphatic phonemes, whereas /uː/ and /iː/ do not have different English patterns. This led Binasfour (2018) to assume that the production or perception of words that have /aː/ or /uː/ vowels surrounding emphatic consonants is more likely to be easier for L2 Arabic learners than words that have the vowel /iː/. Moreover, it was evidenced in my previous study (Alhumaid, 2019) that even L1 Arabic speakers largely depend on surrounding vowels to determine the contrast between an emphatic and non-emphatic pair. It was found that L1 Arabic participants could not differentiate between /s/ and /sˤ/ when the vowel was manipulated and switched between the two phonemes. Therefore, it was expected that L2 Arabic learners would find it more difficult to discriminate, which is in line with Binasfour's (2018) prediction. For this reason, the vowel least affected by the preceding consonant is /iː/, and consequently the most difficult vowel to discriminate, which is why it was avoided in this study and the vowel that followed the target sounds in all three pairs (identical, similar, or new) was either /aː/ or /uː/. This was done in an attempt to help participants differentiate between emphatic and non-emphatic phonemes.

All the lexical items in this study were paired with 18 corresponding images for novel objects. and presented to the participants. These 18 non-real-object images were adapted from Escudero et al.'s (2008) study. Non-real-object images were used to avoid participants mapping any existing lexical representations that might affect their responses. In addition, as this study included non-words, it sought to avoid using real object images to avoid teaching participants any false information and wrong meanings. Table 3.1 presents these nonwords with their corresponding images.

All the lexical stimuli were phonotactically examined to measure their phonotactic probability in Arabic using the Arabic Phonotactic Probability Calculator (APPC) (Aljasser, & Vitevitch, 2018). Phonotactic probability refers to how likely a certain sound is to occur in a sound sequence in a given language (Storkel, Armbrüster & Hogan, 2006). The APPC is based on a written corpus of Arabic that consists of 5.8 billion words named Ar ten ten (Arts, Belinkov, Habash, Kilgarriff, & Suchomel, 2014, as cited in Aljasser & Vitevitch, 2018). The results of these phonotactic probability tests are presented in Table 3.2.

Table 3.1: *List of Visual Stimuli Used in the Study*

Pairs	**Arabic non-words**	**Visual representation**	**Arabic non-words**	**Visual representation**
Alveolar pairs	/sˤu:la/		/su:la/	
	/sˤa:bi/		/sa:bi/	

	/sˤaːtu/		/saːtu/	
	/ðaːmi/		/θaːmi/	
Interdental pairs	/ðaːnu/		/θaːnu/	
	/ðuːri/		/θuːri/	
Uvular pairs	/ʁaːdʒi/		/χaːdʒi/	

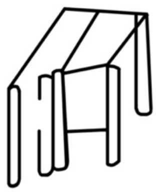
/ʁaːʃu/		/χaːʃu/	
/ʁaːri/		/χaːri/	

Table 3.2: *Phonotactic Probability of the Non-words Used in the Study*

Arabic non-word	**Position-specific**	**Position-specific biphone**	**All phonemes**
/sˤuːla/	.0212, .1831, .0579, .2247	.0018, .0096, .0116	.4869, .0230
/suːla/	.0334, .1831, .0579, .2247	.0053, .0096, .0116	.4991, .0265
/sˤaːbi/	.0212, .0903, .0458, .1173	.0031, .0053, .0046	.2746, .0130
/saːbi/	.0334, .5123, .0458, .1173	.0163, .0259, .0046	.7088, .0468
/sˤaːtu/	.0212, .0903, .0875, .0062	.0031, .0024, .0002	.2052, .0057
/saːtu/	.0334, .5123, .0875, .0062	.0163, .0088, .0002	.6394, .0253
/ðaːmi/	.0056 .0903 .0492 .0437	.0014 .0058 .0025	.1888 .0097
/θaːmi/	.0068 .0903 .0492 .0437	.0010 .0058 .0025	.1900 .0093
/ðuːri/	.0056, .1831, .0824, .1173	.0007, .0145, .0071	.3884, .0223
/θuːri/	.0068 .0052 .0824 .0437	.0001 .0009 .0063	.1381 .0073
/ðaːnu/	.0056 .0903 .0525 .0319	.0014 .0031 .0019	.1803 .0064
/θaːnu/	.0068 .0903 .0525 .0319	.0010 .0031 .0019	.1815 .0060
/ʁaːdʒi/	.0143 .0903 .0329 .0437	.0022 .0031 .0012	.1812 .0065
/χaːdʒi/	.0230 .0903 .0329 .0437	.0038 .0031 .0012	.1899 .0081
/ʁaːʃu/	.0143 .0903 .0259 .0319	.0022 .0020 .0004	.1624 .0046
/χaːʃu/	.0230 .0903 .0259 .0319	.0038 .0020 .0004	.1711 .0062
/ʁaːri/	.0143 .0903 .0824 .0437	.0022 .0083 .0063	.2307 .0168
/χaːri/	.0230 .0903 .0824 .0437	.0038 .0083 .0063	.2394 .0184

Table 3.2 presents the data in three different columns for each lexical item. In the first column, the position-specific probability for each phoneme in the item is presented. By way of illustration, the phoneme /sˤ/, in the lexical item /sˤuːla/, can occur initially with a probability of .0212, the phoneme /uː/, on the other hand, has .1831 probability in the second position; similarly, /l/ can occur in the third position in Arabic with a probability of .0579 and finally /a/ has a probability of .2247 in final position. The data in the second column show the position-specific biphone probability for each combination of two phones. For instance, in the same item /sˤuːla/, there are three possible combinations: /sˤ/ preceding /uː/ with a probability of .0018, /uː/ preceding /l/ with a probability of .0096, and finally, /l/ preceding /a/ with a probability of .0116. The sum of these phoneme and biphone probabilities for each item is presented in the third column. For instance, the total of .0212, .1831, .0579 and .2247 (the position-specific probability for each phoneme in the non-word /sˤuːla/) is 1.4869, whereas the total of .0018, .0096 and .0116 (the position-specific biphone probability for each combination in the non-word /sˤuːla/) is 1.0230. If any of these position-specific probabilities or position-specific biphone probabilities has a value of zero, that means that this phoneme or biphone combination is illegal in the language (Aljasser & Vitevitch, 2018). Looking at the table, it is found that none of the position-specific probabilities or position-specific biphone probabilities has a value of zero, meaning that all the positions of these phonemes and all these phoneme combinations are legal in MSA.

The stimuli were designed and presented to participants using PsychoPy 3 Software (Peirce et al., 2019) which enabled the researcher to measure the reaction time of responses for each trial presented to the groups. This is important when measuring the participants’ ability to discern the differences between target sounds within a given time. All the auditory items were recorded by the same native Arabic speaker using a lapel omni-directional microphone which has a frequency range of 65Hz – 18KHz and a 74dB SPL signal-to-noise ratio. The recordings were recorded by only one native Arabic speaker following Showalter and Hayes-Harb's (2015) finding that, after conducting four different experiments, the performance of their participants in discriminating phonological items was more accurate when the number of speakers was decreased from two to one. This is also consistent with a pilot study conducted by the same researchers (Showalter & Hayes-Harb, 2013). For this reason, this study involved only one native Arabic speaker.

3.3.3 Ethical Procedures

The researcher followed the ethical procedures specified by the Department of English Language and Applied Linguistics Ethics Committee of the University of Reading. The procedures included submission of the project (see Appendix E), a project description explaining the purpose of the project, and details about how participants would be selected and what procedures participants would be asked to perform. In addition, the researcher outlined information about where the data would be stored and how the confidentiality of data would be maintained (see Appendix F). After obtaining approval from the Department of English Language and Applied Linguistics Ethics Committee, participants were given an information sheet (see Appendix G) and provided with a consent form that had to be signed before they enrolled in the study (see Appendix H). For confidentiality purposes, information was saved to the researcher's personal computer protected by a password to which no one else had access.

3.3.4 Pilot Study

Before starting the data collection process, both instruments testing phoneme discrimination and phonemic coding ability were piloted with six participants (2 participants in each group: Arabic-script, Romanised-script, and no-script). The pilot study included all parts of the main study. This was done to ensure the methods were working properly in all stages and phases. Participants were guided by written instructions presented in the same main instrument of the actual study. However, because the pilot study went successfully and no changes were needed, these six participants were counted among the participants of the main study.

3.3.5 Procedure and Data Collection

3.3.5.1 **Demographic Questionnaire.**

All participants were required to complete a questionnaire online using Google forms (see Appendix I) that asked them about their linguistic background. This questionnaire was to satisfy the recruitment criteria to exclude any participant who had learned Arabic, or had higher than GSCE level in German or French. The questionnaire included questions about the participants' L1, the languages used at home, any other languages known by the participants, how these languages were learned, at what age they learned these languages, and what level they had in these languages including writing, reading, listening, and speaking. In addition, they were asked about Arabic in particular, and how they rated themselves in Arabic language proficiency, if they believed they had any.

3.3.5.2 First Instrument: Measuring Phonemic Coding Ability

Phonemic coding ability was investigated in this study to measure its role in phoneme discrimination, and whether it interacted with the effect of orthography. In order to measure phonemic coding ability, participants were asked to take a phonemic aptitude test (Llama-E) as a first stage of the study. Later, they were allocated into one of three groups for the next stage. In other words, this stage served as preparation for the following stage, because it should be taken into account to ensure that every orthography group has similarly mixed levels of phonemic coding ability fluctuating from 0 to 20.

The purpose of the test was to examine the participants' ability to associate sounds with their corresponding symbols by presenting unfamiliar symbols simultaneously with some recorded strings of sounds and, later, participants were asked to guess the correct spelling of some unfamiliar words consisting of these strings of sounds (Saito et al., 2018).

Llama-E consisted of two phases: a learning phase and a testing phase. The learning phase lasted for two minutes and provided participants with 24 recorded syllables along with their unfamiliar written representations. Participants could click on any syllable to hear its auditory form. The testing phase consisted of 20 trials and started by presenting a bisyllabic word orally accompanied by 20 written words to choose from and one option that consisted of question marks to be selected if the participant did not know the answer. For example, participants heard the nonword '3ë3u' and had the choices of '3ë3ë, 3ë3u, 9ë3u, 9ë3ü, 0u0u, 0ë3ë' and many others to choose from. These options did not change when the trial changed. In other words, the multiple options were constant in all 20 trials of the testing phase (see Appendix C). The participants' task was to guess the spelling of the word they heard depending on how every syllable in this word was pronounced. The score was calculated as the number of correct answers out of 20. In my previous study (Alhumaid, 2019), the methodology was subject to considerable criticism in a CoR[2] report in that

---

[2] CoR stands for Confirmation of Registration which is a transfer stage at the University of Reading for doctoral research candidates to PhD status. It takes place at the end of the first year or the beginning of the second year.

participants were only divided into high and low phonemic coding ability groups; the difference between the highest score in the low group and the lowest score in the high group was very close, which consequently may have affected the validity of the results. In an attempt to overcome this issue, phonemic coding ability scores were treated as a continuous variable (out of 20) instead of dividing the participants into phonemic coding ability groups. In addition, Alhumaid (2019) used a version of Llama-E test which provided only two options for the participants in each trial in the testing phase. This gave the participants a 50% possibility of answering correctly by chance, which might have affected its reliability, or scoring validity as Weir (2005) calls it. This is because it largely depends on multiple-choice questions. According to Ali, Carr, and Ruit (2016), multiple-choice questions provide test-takers with options, thereby giving them the opportunity to guess the correct answer, which consequently affects the test scoring validity. The version of Llama-E test used in Alhumaid's (2019) study provided test-takers with only two options, which increased the chance of guessing the correct answer with 50% accuracy. This issue was avoided in the current study by using a new version of the Llama-E test which presents more than two multiple options at the same time, which decreases the possibility for participants to answer by chance. The aptitude test stage of the study aimed to answer the fifth research question which links phonemic coding ability as a predictor of their performance in the phoneme discrimination experiment. Table 3.3 shows the scores in the first stage (aptitude test stage).

Table 3.3: *Study Procedure: Aptitude Test Stage*

Llama test score	Numbers of participants obtaining these scores			Total
	Arabic-script group	Romanised-script group	No-script group	
0	1	2	1	4
1	2	3		5
2	1	4	1	6
3	2		1	3
4			1	1
5			2	2
6		2	2	4
7	3	1		4
8			1	1
9	2	1		3
10			2	2
11		1	2	3
12		1	1	2
13	1	1	1	3
14	1	1		2
15		2	2	4
16	2	1	2	5
17	4	1	2	7
18	1	3	3	7
19	4	1		5
20	1		1	2
Total	25	25	25	75

3.3.5.3 Second Instrument: Measuring the Effect of Orthographic Availability and L1/L2 Phoneme Similarity on Phonological Acquisition

After taking the Llama-E test, participants had to report their scores to the researcher. They were then allocated one of three orthography groups, Arabic-script, Romanised-script along with Arabic script, and no-script. An attempt was made to balance the groups in terms of their phonemic coding ability scores. However, there were some difficulties in keeping the balance because it was not possible to monitor the participants while taking the test as it was done online due to Covid 19 pandemic. This means that there was a possibility that they did not give accurate information about their scores, but the researcher had to take them on trust and accept them. Later, when the researcher got all the data of the Llama-E test from the test designer (Prof. Paul Meara), it was surprising that a few participants had been dishonest about their scores, and other took the test more than once and reported the best score. Therefore, the researcher had to retrospectively refer to the actual score of each participant, leading to a lack of balance in the three orthographic groups in their phonemic coding ability scores. Fortunately, there was no significant difference between the three groups in terms of their phonemic coding ability scores.

In Alhumaid (2019), only two groups were investigated: an Arabic-orthography group and a no-orthography group. Despite the fact that participants in Alhumaid (2019) were given a short introduction to the Arabic language, including basic rules of how Arabic script is written, the no-orthography group performed slightly better than the other group. Therefore, even though the difference was not significant, it showed that orthography did not help the participants discriminate between phonemes as in Showalter and Hayes-Harb's (2015) and Mathieu's (2014, 2016) studies, who attributed the findings to the difficulty and novelty of the Arabic script. In an attempt to test this, Showalter and Hayes-Harb (2015) provided their participants with Romanised-script, which was also found not to be helpful, leading the researchers to suggest providing participants with a combination of both familiar (Romanised) and novel (Arabic) scripts. This is the rationale for the presentation of stimuli in the three groups in the current study. The Arabic-script group had only Arabic orthographic input and were alerted by an arrow to the direction of Arabic script (from right to left). The Romanised-script group had both Arabic and Roman orthographic input in the full script condition, i.e., they had the stimuli presented in Roman script where all characters were in Roman, and they had it also in Arabic script where all characters were in Arabic, leading to two different scripts at the same time. They were alerted to the direction of each script. However, they were not explicitly told that the Arabic starts from right to left. Despite Showalter and Hayes-Harb's (2015) finding that instructions on how Arabic script is written was not helpful and did not affect participants' performance, Alhumaid (2019) did give explicit instructions and comprehensive introduction about Arabic orthography; this was also not helpful, as the performance of participants who had orthographic input was not significantly different from those who did not have orthographic input. However, in the current study, the procedure was done online due to the Covid 19 pandemic and, therefore, the introduction about how Arabic is written was removed, given that the instruction did not have an effect in the previously mentioned studies (Alhumaid, 2019; Showalter & Hayes-Harb, 2015). In the no-script group, participants had the symbol (,(طططط which corresponds to the Roman symbol (xxxx), instead of orthographic input. The details of these three groups are discussed in more detail in the following subsections.

This stage (the phoneme discrimination stage) aimed to answer the first, second, third and fourth research questions by examining the effect of orthographic input in terms of script familiarity and the similarity of L2 and L1 phonemes. This was done by giving participants some activities that required them to differentiate between two Arabic-like minimal pairs that only contrasted in the initial position. The procedure in this stage was adapted from Showalter and Hayes-Harb (2015) with the addition of some activities which will be demonstrated in detail later in this subsection.

The phoneme discrimination stage was divided into three conditions with five phases in each condition, as mentioned in Subsection 3.3.2. These conditions followed a phoneme-similarity criterion:

- An identical-phonemes condition (/ð/ and /θ/)

- A similar-phonemes condition (/s/ and /sˤ/)
- A new-phonemes condition (/ʁ/ and /χ/)

Some of the studies conducted to test the effect of L1/L2 phoneme similarity showed that L2 phonemes that are mapped to a single L1 phoneme were easier to discriminate than those which are not mapped to any L1 phonemes (Han & Oh, 2018). This was not in line with SLM-r (Flege & Bohn, 2021) and PAM (Best, 1995). Due to this discrepancy in the effect of L1/L2 phoneme similarity, the three conditions in this study were presented to each participant in the same order. Having the same order in the three phoneme conditions (identical, followed by similar, followed by new phoneme condition) aimed to control the difficulty level equally for all participants, starting with what was thought to be the easiest as identical phonemes were mapped to two L1 phonemes, then moving to those which were mapped to only one L1 phoneme (similar phonemes), and finally those which were not mapped to any L1 phonemes (new phonemes).

These three Arabic pairs went through the same five phases: a Pre-Sound Recognition Phase, Word Learning Phase, Criterion Test Phase, Phonological Test Phase, and Post-Sound Recognition Phase. These phases are explained in detail below. Using PsychoPy 3 Software (Peirce et al., 2019), the stimuli in these phases were conducted online and presented to participants containing different auditory, visual, and written stimuli. Instructions were given to the participants in English and included in writing prior to each phase, after the end of the first task, and before the beginning of the second task in each phase that included two tasks. Questions within the trials were also presented in writing at the top of each trial.

3.3.5.3.1 Pre-Sound Recognition Phase

The aim of this phase was to measure the participants' ability to discriminate between two target phonemes before getting involved in the Word Learning Phase. This phase included two similar tasks and two different types of auditory input: isolated phonemes, i.e., phonemes were separated from any other preceding or following phonemes (Phoneme Discrimination Task), and Arabic minimal pairs including these phonemes in initial position (Word Discrimination Task). The design of the tasks in this phase was Same – Different where participants were introduced to two stimuli in each trial, separated by one second of silence. These two stimuli are either the same or different (McGuire, 2010).

In the Phoneme Discrimination Task, participants were presented with 12 pairs of stimuli. In each of the 12 stimuli, two isolated Arabic phonemes were uttered sequentially by the same native Arabic speaker. The participants' task was to determine whether these two phonemes sounded the same or different by choosing the "1" key corresponding to "Yes" or the "0" key corresponding to "No". Participants had only five seconds to answer. The next trial was played automatically after five seconds and if the participant did not respond in those five seconds, his/her missing answer was recorded as a wrong answer. This was applied for all stimuli in this task. This task included the two target phonemes in addition to a distractor phoneme. The phonemes presented in this phase were /ð/, /θ/ and /ʃ/ in the identical-phonemes condition, /s/, /sˤ/ and /z/ in the similar-phonemes condition, and /ʁ/, /χ/, and /ħ/ in the new-phonemes condition. Each phoneme was uttered twice with the same phoneme (the answer is Yes) and twice with the other two different phonemes (the answer is No), resulting in six matching answers and six not matching answers. Table 3.4 presents all 12 pairings in all three conditions.

Table 3.4: *List of Pairs in the Pre-Sound Recognition Phase: Phoneme Discrimination Task*

Pairs	Trial	First phoneme	Second phoneme
Identical-phonemes condition (Interdental pairs)	1	/θ/	/ð/
	2	/θ/	/θ/
	3	/θ/	/ʃ/
	4	/θ/	/θ/
	5	/ð/	/θ/
	6	/ð/	/ð/
	7	/ð/	/ʃ/
	8	/ð/	/ð/
	9	/ʃ/	/θ/
	10	/ʃ/	/ð/
	11	/ʃ/	/ʃ/
	12	/ʃ/	/ʃ/
Similar-phonemes condition (Alveolar pairs)	1	/s/	/s/
	2	/sˤ/	/sˤ/
	3	/s/	/z/
	4	/z/	/z/
	5	/sˤ/	/s/
	6	/z/	/s/
	7	/s/	/s/
	8	/s/	/sˤ/

Table 3.4 cont.: *List of Pairs in the Pre-Sound Recognition Phase: Phoneme Discrimination Task*

Pairs	Trial	First phoneme	Second phoneme
	9	/z/	/sˤ/
	10	/sˤ/	/z/
	11	/sˤ/	/sˤ/
	12	/z/	/z/
New-phonemes condition (Uvular pairs)	1	/ʁ/	/χ/
	2	/ħ/	/ʁ/
	3	/ʁ/	/ʁ/
	4	/χ/	/χ/
	5	/ʁ/	/ħ/
	6	/ħ/	/ħ/
	7	/ʁ/	/ʁ/
	8	/χ/	/ħ/
	9	/χ/	/χ/
	10	/ħ/	/χ/
	11	/ħ/	/ħ/
	12	/χ/	/ʁ/

The Word Discrimination Task was very similar to the first one. The difference was that words consisting of real Arabic minimal pairs, i.e., words in which the first consonant contrasted initially, were uttered instead of isolated phonemes. The same phonemes used in the first task were used as consonants, including a distracting one. The participants’ task was to determine whether the two words sounded the same or different. Table 3.5 presents the Arabic minimal pairs used in the second task. The procedure was identical to the first task, including the number of trials, the number of matching answers, and the number of mismatching answers (see Appendix J).

Table 3.5: *List of the Arabic Minimal Pairs in the Pre-Sound Recognition Phase: Word Discrimination Task*

Pairs	Arabic word	Arabic script	Arabic word	Arabic script
Identical-phonemes condition (Dental pairs)	θamma	ثمّ	ðamma	ذمّ
	θa:b	ثاب	ʃa:b	شاب
	ðara:	ذرا	θara:	ثرى
	ðabaħ	ذبح	ʃabaħ	شبح
	ʃaraf	شرف	ðaraf	ذرف
	ʃara:	شرى	θara:	ثرى
Similar-phonemes condition (Alveolar pairs)	saif	سيف	sˤaif	صيف
	sahar	سهر	zahar	زهر
	sˤarħ	صرح	sarħ	سرح
	sˤa:d	صاد	za:d	زاد
	zir	زر	sir	سر
	zaraʕ	زرع	sˤaraʕ	صرع
New-phonemes condition (Uvular pairs)	ʁaimah	غيمة	χaimah	خيمة
	ʁa:li:	غالي	ħa:li:	حالي
	χair	خير	ʁair	غير
	χad	خد	ħad	حد
	ħa:ʔim	حائم	ʁa:ʔim	غائم
	ħal	حل	χal	خل

3.3.5.3.2 Word Learning Phase

This was the only phase where the input provided to the participants in the three orthographic groups was different. The Arabic-script group had three different types of input, both auditory and visual, including pictures and written (Arabic) stimuli. The Romanised-script group also had three different types of input. The difference between the two groups is that the Romanised-script group had both Arabic and Roman written input and they were guided by a red arrow to the direction of the script, as Arabic script reads from right to left and the Roman script reads in the opposite direction. The No-script group had the same auditory and visual input as the other two groups except that the written input was linguistically "empty" as it was substituted by the symbol (ططط) which corresponds to the (xxxx) symbol (see Appendix K). This symbol was used to ensure that all groups were balanced and had the same amount of visual input (Showalter & Hayes-Harb, 2015). According to Winke, Sydorenko, and Gass (2013), participants might be negatively affected by the number of different inputs, as reported in Chai and Erlam (2008, as cited in Winke et al., 2013), who found that some participants reported that they did not have enough time to move from written to visual input or vice versa, others reported that they were not able to listen and read at the same time.

In the Word Learning Phase, the participants' task was only to learn the lexical representation of each lexical item, including its written (if available), auditory, and visual representations. They were not required to react or undertake another activity. Table 3.6 presents the non-words introduced to participants in the three conditions.

All groups had lexical input presented as minimal pairs in the same trial with five seconds between the appearance of the first and second non-words. Each condition included three Arabic-like minimal pairs in three trials (six Arabic non-words) with ten repeated blocks, totalling 60 new word exposures in every condition. It is argued that the number of repetitions plays an important role in determining the success of vocabulary learning (Nation, 2014). It is not agreed, however, how many repetitions should take place in order for learning to occur. Webb (2007, as cited in Nation, 2014) claims that ten repetitions are sufficient for learning to occur, whereas Waring and Takaki (2003, as cited in Nation, 2014) claim that eight repetitions is sufficient. Therefore, this study provided ten repetitions to ensure that participants had enough time to learn vocabulary items.

In the Romanised-script group, as seen in Table 3.6, the researcher attempted to avoid any ambiguity between the pronunciation of the digraph <th>, which has two different pronunciations in English language /ð/ and /θ/, by underlining the digraph that represents the phoneme /θ/ in an attempt to attract participants' attention to the two different pronunciations. The two similar phonemes /sˤ/ and /s/ have only one corresponding phoneme in English /s/, which means there is only one corresponding grapheme. The researcher inserted the symbol (') to indicate the difference in pronunciation of the two phonemes in which <s> corresponds to the phoneme /s/ and <s'> corresponds to the phoneme /sˤ/. The last pair of phonemes was novel to the participants. That is why the researcher used <gh> corresponding to /ʁ/ and <kh> corresponding to /χ/ as graphemic representations of these two Arabic phonemes.

Table 3.6: *List of the Non-words Used in the Study*

Pairs	Arabic non-words	Arabic script	Romanised script	Arabic non-words	Arabic script	Romanised script
Identical-phonemes condition (Dental pairs)	/ða:mi:/	ذامي	Thami	/θa:mi:/	ثامي	Thami
	/ða:nu:/	ذانو	Thano	/θa:nu:/	ثانو	Thano
	/ðu:ri:/	ذوري	Thuri	/θu:ri:/	ثوري	Thuri
Similar-phonemes condition (Alveolar pairs)	/sˤu:la:/	صولا	S'ula	/su:la:/	سولا	Sula
	/sˤa:bi:/	صابي	S'ato	/sa:bi:/	سابي	Sato
	/sˤa:tu:/	صاتو	S'abi	/sa:tu:/	ساتو	Sabi
New-phonemes condition (Uvular pairs)	/ʁa:dʒi:/	غاجي	Ghaji	/χa:dʒi:/	خاجي	Khaji
	/ʁa:ʃu:/	غاشو	Ghasho	/χa:ʃu:/	خاشو	Khasho
	/ʁa:ri:/	غاري	Ghari	/χa:ri:/	خاري	Khari

3.3.5.3.3 Criterion Test Phase

This phase was in completion of the Word Learning Phase. It aimed to enhance the participants' learning process by examining their acquisition of words' meanings rather than phonological discrimination. In other words, this phase aimed to ensure that participants had learnt the association between a lexical item and its corresponding picture. This means that the purpose of this phase was not to measure their ability to distinguish the two phonemes in minimal pairs, as this discrimination was not required to perform this task. Therefore, in the case of mismatching pairs, there was no similarity between the auditory form heard and the picture presented, i.e., it was not a minimal pair (e.g., the picture presented was /ða:mi/ and the auditory form heard was /θu:ri/). This phase included two different tasks. The design of the tasks in this phase was

Identification tasks which have many types, among them being the ones used in this phase: Yes-No Identification Task and Forced Choice Identification Task. These two types of tasks differ from each other in that, in the former, one stimulus is presented per trial and compared to another one (e.g., Was the stimulus x or y?), whereas the latter offers one or multiple stimuli and participants are required to label these (McGuire, 2010).

In the first task (Yes-No Identification Task), participants heard an auditory form and saw a picture using PsychoPy. Their task was to determine whether the picture matched the auditory form they had just heard or not by choosing "1" for "Yes" if the picture matched the auditory form and "0" for "No" if the picture did not match the auditory form. This phase did not include any orthographic information for any of the three orthography groups. Participants had five seconds to answer in 12 trials that included six matched items (e.g., the picture presented was /ðaːmi/ and the auditory form heard was /ðaːmi/) and another six mismatched items (e.g., the picture presented was /ðaːmi/ and the auditory form heard was /θuːri/) (see Appendix L).

Participants also did another task in this phase (Forced Choice Identification Task). In this task, two pictures were presented at the same time and numbered "1" and "0", while the participants heard one auditory form. In this task also, no groups had any orthographic input. The participants' task was to choose the matching picture by pressing either "1" or "0" within five seconds. There was no phonological relationship between the two pictures (e.g., the first picture represented /ðaːmi/ and the second picture represented /θuːri/ while the auditory form was /θuːri/). This task also comprised 12 trials. Every auditory form was presented in two trials in which the distracting picture was not the same both times (e.g., the first time the auditory form /θuːri/ was presented with a picture that represented /θuːri/ as well as a picture that represented /ðaːmi/, and the second time with a picture that represented /ðaːno/) (see Appendix L).

3.3.5.3.4 Phonological Test Phase

The Phonological Test Phase was identical to the Yes-No Identification Task in the Criterion Test Phase, but the difference was that the purpose of this phase was to measure the participants' ability to discriminate phonological contrasts. The stimulus in the task involved minimal pairs. For instance, the picture presented in the task in the mismatched trial was /ðaːmi/ and the auditory form heard was /θaːmi/. So, the participants' task was to differentiate between the two phonemes /θ/ and /ð/. Every trial was repeated randomly twice, so 12 trials in total. Everything else was identical to the Yes-No Identification Task in the Criterion Test Phase (see Appendix M).

3.3.5.3.5 Post-Sound Recognition Phase

This phase was exactly the same as Task 1 in the Pre-Sound Recognition Phase, except that the distracting phoneme was removed, meaning that this phase only included the target phonemes. The purpose of this phase was to measure the effect of the orthographic input available in the Word Learning Phase on the participants' ability to discriminate between isolated phonemes. Their performance was compared with their performance in the Phonological Test Phase to see if there was a difference in the participants' performance in the case of phonemes uttered in isolation or phonemes embedded in words.

It is important to acknowledge that the Word Learning Phase and the Yes-No Identification Tasks in the Criterion Test and Phonological Test Phases were adapted from Showalter and Hayes-Harb (2015). However, the Pre- and Post-Sound Recognition Phases as well as the Forced Choice Identification Task of the Criterion Test were new tasks in this study. Table 3.7 summarises the tasks performed in all five phases of the phoneme discrimination stage.

Table 3.7: *Study Procedure: Phoneme Discrimination Stage*

Conditions			Procedure
Identical phonemes (/ð/ & /θ/)	Similar phonemes (/s/ & /sˤ/)	New phonemes (/ʁ/ & /χ/)	
Pre-Sound Recognition Phase			Phoneme Discrimination Task: participants heard 2 phonemes and decided whether they were different or the same Word Discrimination Task: participants heard 2 Arabic minimal pairs contrasted initially and decided whether they were different or the same
Word Learning Phase			Participants were exposed to 3 minimal pairs contrasting in word initial position accompanied by non-real-object pictures. Some participants had orthographic input depending on the group they belonged to
Criterion Test Phase (non-minimal pairs)			Yes-No Identification Task: participants heard a word and saw a picture and decided whether the picture matched the word Forced Choice Identification Task: participants heard a word and saw 2 pictures and decided which picture matched the word
Phonological Test Phase (minimal pairs)			Yes-No Identification Task: participants heard a word and saw a picture and decided whether the picture matched the word
Post-Sound Recognition Phase			Phoneme Discrimination Task: Participants heard 2 phonemes and decided whether they were different or the same

3.4 Data Analysis

The main analysis in this study involved an examination of the performance in the Phonological Test Phase and the Post-Sound Recognition Phase. This analysis had one dependent variable in two levels (word level and phoneme level) with two measures: accuracy and reaction time in correct trials. Because there was more than one predictor variable (Field, 2018) in the key analysis for this study, the study followed a factorial design and was analysed using two-way [L1/L2 phoneme similarity (identical phonemes, similar phonemes, new phonemes) by Orthographic input type (Arabic-script, Romanised-script, no-script)] analysis of variance (ANOVA) (Loewen & Plonsky, 2015) with each dependent variable considered separately. Phonemic coding ability was analysed in this study as a predictor vatiable. Therefore, this study included also a multiple linear regression. The factorial design was selected because one of the main aims of the study was to measure the effect of many variables (L1/L2 phoneme similarity, orthographic input type, and phonemic coding ability), which had a number of levels, as mentioned above, on another variable (phonological discrimination). In addition, it aimed to measure the interaction between these factors in their effect on the dependent variable (Loewen & Plonsky, 2015). Data were collected from different groups of orthographic type (between-subjects), same group in different levels of L1/L2 phoneme similarity (within-subjects), and phonemic coding ability scores as a continuous variable, following a mixed-design method that had independent groups variable (orthographic input type) and a repeated measures variable (phoneme similarity level). Phonemic coding ability, on the other hand, was a continuous variable and included in analyses as a predictor.

In order to answer the five research questions, the study followed three methods of analysis. The first method was a 2-way ANOVA which was carried out to answer *RQs*1, 2, and 3. The second method was a correlational analysis which was carried out to answer *RQ*4. The last method was multiple linear regression which was carried out to answer *RQ*5.

2-way ANOVA was selected as a method for analysing the data in this study because all ANOVAs allow the measurement of one dependent variable on a continuous scale, whereas independent variables have a categorical nature. In other words, ANOVA is a flexible technique that gives the researcher the opportunity to test the effect of any given number of categorical predictive variables or factors at the same time on any continuous response measure. In addition, all the different types of ANOVA involve the researcher in determining whether the data are collected from different groups (between-subjects) or from the same group (within-subjects); both are used in this study. ANOVA also tests the effect of each factor at the same time that it controls the variance caused by other factors, which means that it allows testing the statistical interaction between factors and can tell whether the effect of one variable is consistent across the different levels of another variable.

3.5 Summary of the Chapter

This chapter presented a detailed description of the experimental design of the study. It elaborated on how the stimuli were presented to participants, how ethical procedures were followed, how the study was piloted, how participants were recruited, how instruments and procedures were used to collect data, and what methods were used to analyse those data. This chapter also provided a justification for using and selecting certain procedures and methods, and some types of stimuli.

4 Results

4.1 Introduction

This chapter presents the results of the current study which aimed to explore the effect of certain factors on L1 English participants' ability to benefit from L2 orthographic input during their Arabic L2 phonological learning. These factors include orthographic input type (novel script, familiar and novel script together, and no script at all), L1/L2 phoneme similarity (identical phonemes, similar phonemes, and new phonemes), and the level of the participants' phonemic coding ability, where participants had different scores varying from 0–20. The data set for the study in all three orthographic input-type groups is presented in Appendix N.

To start, data screening was conducted on all data at all levels to check that the assumptions underlying the use of parametric tests were met. According to Field (2018), there are four main common statistical procedures used in the social sciences to test that data set meet the assumptions underlying the use of parametric statistical tests. These are assumptions of linearity, normality, equality of variance, and independence. Assumption of linearity refers to, according to Field (2018), the condition where the scores for the outcome variable (phonological discrimination in the current study) are connected to any predictor in a linear relationship, meaning that if there is more than one predictor (as in the current study) then adding the effects of these predictors together will result in having a combined effect. According to Field (2018), the meaning of the normality distribution assumption is not restricted to the case where data are normally distributed, but also applies to the effects of this distribution on fitting models and assessing them. In non-normally distributed models, parameter estimates, such as mean and median, are affected by extreme scores. However, the median is less biased than the mean in the case of a skewed distribution. Moreover, because confidence intervals are computed from values which come from a standard normal distribution, if parameter estimates come from a non-normal distribution, the values used to compute confidence intervals do not make sense and are inaccurate. The equality of variance assumption assumes that all groups involved in the study "come from populations with the same variance" (Field, 2018, p.237). According to Hayes and Cai (2007, as cited in Field, 2018), violating this assumption results in having bias and leads to the standard error associated with parameter estimates being inconsistent. The assumption of independence means that there is no relation between errors in the model, i.e., the error in predicting one participant's response is not influenced by the error in predicting another participant's response (Field, 2018). This is important in estimating the standard error which will not be valid unless the study has independent observations (Field, 2018). How these assumptions were tested and the results of these tests on participants' Phonemic Coding Ability scores, Pre-Sound Recognition Phase scores, Criterion Test Phase scores, Phonological Test Phase scores, and Post-Sound Recognition Phase scores in all the orthographic input-type groups are presented in Appendix O. As shown in Appendix O, the assumptions of linearity and of independence were met, whereas the assumptions of normality and of equality of variance were not satisfied in all cases.

It is argued that ANOVA is relatively robust to violations of statistical assumptions, such as normality distribution (Field, 2018) or equivalence of variance. According to Lunney (1970, as cited in Field, 2018). *F*-statistics are likely to be robust to violations of normality if the sizes of the groups are equal, which is the case in the current study. However, Field (2018) disagrees to some extent with that as he claims that even in the case of having equal group sizes, *F*-statistic is usually affected by the absence of normality and/or violations of equality of variance. Therefore, an ANOVA was selected to analyse this stage of the current study. However, as the assumptions underlying ANOVA were not fully met, as shown in Appendix O, the researcher took the approach of carrying out non-parametric tests as well in order to validate the findings of the parametric tests. Because the results of these non-parametric tests did not show any large discrepancies from the ANOVA test results – though slight differences did occur in some places – and because the study sought to undergo a factorial design, it was decided to report the ANOVA results here and present the results of the non-parametric tests in the Appendices (P and Q). The slight differences between the parametric and non-parametric tests will be pointed out as appropriate throughout this chapter. Given the current awareness of the issues regarding replicability crisis, the researcher decided to only further investigate the significant results with follow up non-parametric tests rather than verifying non-significant results to be as conservative as possible about claiming significant effect.

4.1.1 Chapter Outline

This chapter starts with an ANOVA which was conducted on phonemic coding ability and Pre-Sound Recognition Phase scores. The collection of these two types of data preceded the treatment in the orthographic input-type groups and, therefore, are seen as baseline phases. Thus, Section 4.3 of this chapter provides the results of a one-way ANOVA of phonemic coding ability scores and a two-way [Orthographic input type (Arabic script, Romanised with Arabic scripts, No script) by L1/L2 phoneme similarity (identical phonemes, similar phonemes, new phonemes)] ANOVA of Pre-Sound Recognition Phase scores. Section 4.4 shows the results of conducting one sample *t*-test aiming to test whether the participants were answering better than by chance or not in the Criterion Test Phase, the Phonological Test Phase, and the Post-Sound Recognition Phase. Section 4.5 presents the main analysis of the study. The first subsection of Section 4.5 presents a two-way [Orthographic input type (Arabic script, Romanised with Arabic scripts, No script) by L1/L2 phoneme similarity (identical phonemes, similar phonemes, new phonemes)] ANOVA for the Phonological Test Phase and the Post-Sound Recognition Phase (Subsection 4.5.1). This was done in order to provide answers to *RQ*1, *RQ*2, and *RQ*3. The second subsection reports a correlational analysis between responses and reaction time in the Phonological Test Phase with those in the Post-Sound Recognition Phase in the three conditions of L1/L2 phoneme similarity, which provides an answer to *RQ*4 (Subsection 4.5.2). Finally, the third subsection presents a multiple linear regression with three predictors (Orthographic input type, L1/L2 phoneme similarity, and phonemic coding ability), which provides an answer to the final research question (Subsection 4.5.3). Section 4.6 presents a summary of the chapter.

As previously mentioned, data were analysed to address the following research questions:

RQ1: Does the presence of orthography help participants differentiate between two phonemes in the case of:

- An unfamiliar script such as Arabic script; or
- A Combination of Familiar Script such as Romanised-Arabic Script and Unfamiliar Script such as Arabic Script?

RQ2: Does L1/L2 phoneme similarity affect the difficulty of participants' ability to differentiate between two phonemes?

RQ3: Is the effect of orthographic availability influenced by L1/L2 phoneme similarity?

RQ4: Does the participants' ability to discriminate phonemes when embedded within words (word-level discrimination) correlate with discriminating phonemes when uttered in isolation (phoneme-level discrimination)?

RQ5: Is the effect of orthographic availability influenced by the combination of phonemic coding ability and L1/L2 phoneme similarity?

The data from three orthographic groups were analysed as a between-subjects variable (Arabic script, Romanised with Arabic script, and no script), from the same group in different conditions as a within-subjects variable (identical phonemes, similar phonemes, and new phonemes), and individually as a continuous variable (phonemic coding ability scores).

4.2 Analysis of the Baseline Phases

After providing data screening (Appendix O) and before moving to the main analysis of the study, it is important to analyse the scores for phonemic coding ability and the Pre-Sound Recognition Phase scores. It is worth noting that the phonemic coding ability test and the Pre-Sound Recognition Phase preceded orthographic input in the Word Learning Phase, and as all participants in the three groups had identical input for phonemic coding ability and in the Pre-Sound Recognition Phase, regardless of which orthographic input type they were exposed to, phonemic coding ability and the Pre-Sound Recognition Phase are considered baseline phases.

4.2.1 Analysis of Phonemic Coding Ability

Table 4.1 shows descriptive statistics for phonemic coding ability scores in all the orthographic input-type groups, including mean and standard deviation (*SD*).

Table 4.1: *Descriptive Statistics for Phonemic Coding Ability Scores*

Orthographic input type	*M* (out of 20)	*SD*
Arabic-script	11.64	7.01
Romanised-script	9.04	6.88
No-script	11.12	5.98

A one-way ANOVA was conducted to ensure there was no prior differences between the three orthographic input-type groups in terms of their phonemic coding ability scores. Results of a one-way ANOVA indicate that there were no significant differences between the three groups in their phonemic coding ability scores ($F[2, 72] = 1.07$, $p = .34$, $\eta p^2 = .02$) with a low observed power (.23).

4.2.2 Analysis of Pre-Sound Recognition Phase

This phase consists of two tasks in each condition of L1/L2 phoneme similarity (for details about these two tasks, see Subsection 3.3.5.3.1). The purpose of this phase, as mentioned in Chapter 3, was to test the participants' discrimination and prior knowledge of these Arabic phonemes. Table 4.2 reports the descriptive statistics for this phase (mean and *SD*) including the three conditions of L1/L2 phoneme similarity in the three orthographic groups in both tasks.

A separate two-way: [Orthographic input type (Arabic script, Romanised with Arabic scripts, No script) by L1/L2 phoneme similarity (identical phonemes, similar phonemes, new phonemes)] ANOVA was conducted on each of the two tasks in the Pre-Sound Recognition Phase. A two-way ANOVA was carried out to ensure that there were no prior differences between the three orthographic input-type groups in terms of their performance in the Pre-Sound Recognition Phase. According to Field (2018), if a study design consists of repeated measures, the sphericity assumption should be met. A Mauchly's test indicated that the sphericity assumption was not met for the Pre-Sound recognition scores, therefore a Greenhouse-Geisser correction was applied (Field, 2018).

Table 4.2: *Descriptive Statistics for the Pre-Sound Recognition Phase (Tasks 1 and 2)*

Orthographic input type	*N*		Identical sounds condition (Out of 12)		Similar sounds condition (Out of 12)		New sounds condition (Out of 12)	
			Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Arabic-script	25	*M*	11.56	11.28	9.96	10.24	11.80	10.84
		SD	0.58	1.2	1.45	1.3	0.64	.98
Romanised-script	25	*M*	11.40	11.4	10.12	10.04	11.16	10.68
		SD	1.22	.81	1.09	1.45	1.51	1.34
No-script	25	*M*	11.48	10.8	9.84	10.08	11.52	9.96
		SD	1.63	1.41	1.99	1.07	2.00	2.49
Total	75	*M*	11.48	11.16	9.97	10.12	11.49	10.49
		SD	1.21	1.18	1.54	1.27	1.5	1.75

With regard to the Phoneme Discrimination Task (Task 1 in the table), as shown in Table 4.3, the two-way ANOVA indicated that there was a main effect of the level of L1/L2 phoneme similarity, suggesting that the accuracy across the groups differed significantly depending on whether phonemes were identical, similar, or novel. However, the analysis showed that there were no significant differences between the three orthographic input-type groups, suggesting that the accuracy of the participants did not differ depending on which orthographic input type they were allocated to. In addition, the interaction between the level of L1/L2 phoneme similarity and orthographic input was not significant.

Table 4.3: *Analysis of the Main Effects of L1/L2 Phoneme Similarity, Orthographic Input Type and the Interaction Between them in the Pre-Sound Recognition Phase*

Task	Variable	*F*	*p*	Eta Partial Squared	Observed Power
Phoneme Discrimination	L1/L2 phoneme similarity	31.68	<.001	.305	1.000
	Orthographic input type	0.36	.69	.010	.105
	Interaction between L1/L2 phoneme similarity and orthographic input type	0.63	.61	.017	.19
Word Discrimination	L1/L2 phoneme similarity	12.9	<.001	.152	.997
	Orthographic input type	1.98	.145	.052	.398
	Interaction between L1/L2 phoneme similarity and orthographic input type	0.84	.49	.023	.265

To explore which conditions of L1/L2 phoneme similarity were significantly different, contrast tests with a Bonferroni correction were carried out for both tasks. Phoneme Discrimination Task results showed that there was no significant difference between the identical-phoneme and new-phoneme conditions, (p =1). However, this was not the case with the similar-phoneme condition which was significantly different from the identical-phoneme condition and new-phoneme condition and had lower accuracy than that in the identical-phoneme and new-phoneme conditions (p <.001) (see Fig. 4.1 for the Means of the three conditions of L1/L2 phoneme similarity).

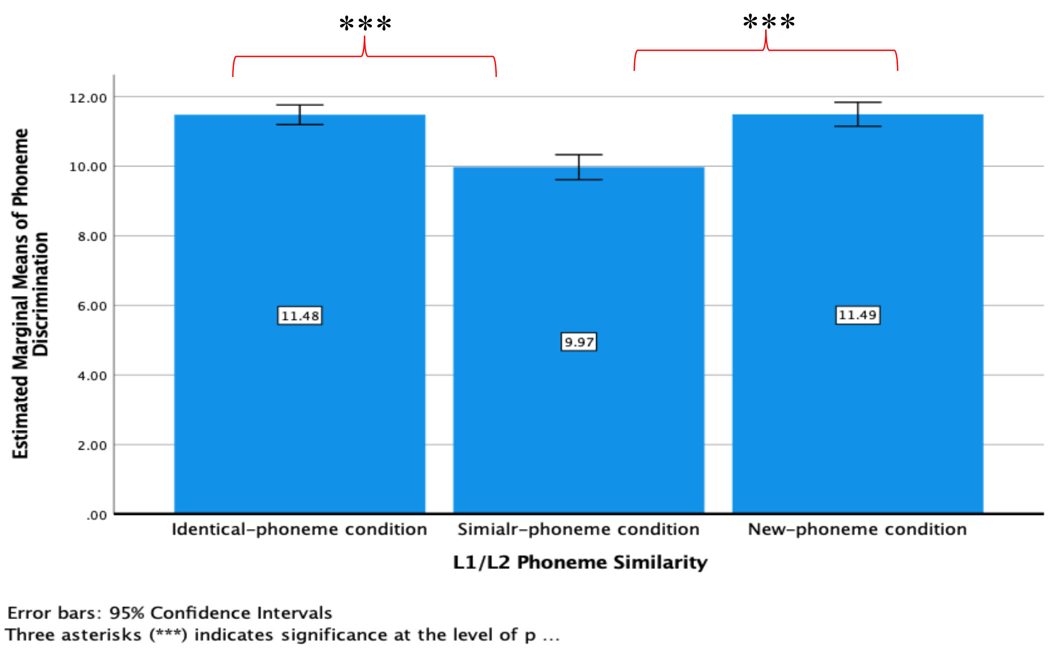


Figure 4.1 Accuracy in the three conditions of L1/L2 phoneme similarity (Pre-Sound Recognition Phase / Phoneme Discrimination Task)

Performance on Word Discrimination Task was very similar to that on Task 1, and data analysis produced very similar results. There was a main effect of L1/L2 phoneme similarity on the participants' accuracy. The analysis of orthographic input type effects showed that there were no significant differences in the accuracy of the participants in the three orthographic input-type groups. The interaction between them was not significant (as shown in Table 4.3).

Results of contrast tests with a Bonferroni correction for Word DiscriminationTask showed that participants' accuracy in the identical-phoneme condition was significantly different and better than their accuracy in the similar-phoneme condition (p <.001), and the new-phoneme condition (p = .014), By comparing the other two L1/L2 phoneme similarity conditions, it was found that the accuracy of the participants in the similar-phoneme condition was not significantly different from that in the new-phoneme condition (p = .23) (see Fig. 4.2 for the Means of the three conditions of L1/L2 phoneme similarity).

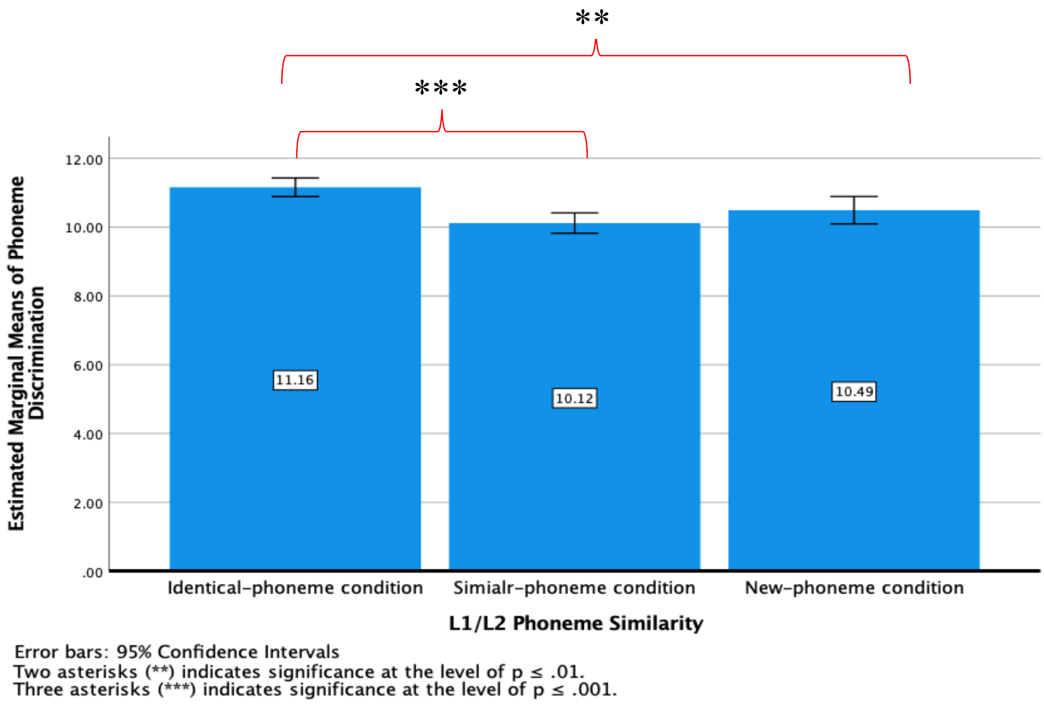


Figure 4.2 Accuracy in the three conditions of L1/L2 phoneme similarity (Pre-Sound Recognition Phase / Word Discrimination Task)

4.3 Testing the Possibility of Answering by Chance

Since all tasks in the second stage of the experiment (phoneme discrimination stage) were multiple-choice questions with two options per task, there was a 50% possibility of the participants choosing the correct answer simply by guessing. As the Criterion Test Phase was part of the Learning phase and its purpose was to enhance the participants' learning before moving on to the main analysis of the current study (the Phonological Test Phase and the Post-Sound Recognition Phase), a one-sample t-test was performed to measure participants' answers in the two tasks of the

Criterion Test Phase, the Phonological Test Phase, and the Post-Sound Recognition Phase. This was done to test whether their responses were significantly above chance (better than 50% correct).

After splitting the data and testing each L1/L2 phoneme similarity level separately, the results showed that, with regard to the Criterion Test Phase, for all the groups in the three conditions of L1/L2 phoneme similarity performance was significantly better than chance in each task except for one case, indicating that some learning had occurred in most cases. The only case where result was not significant was the Yes-No Identification Task in the identical-phoneme condition for the Romanised-script group *p* = .33. However, the responses for the Forced Choice Identification Task of the same group in the same condition of L1/L2 phoneme similarity were significant, indicating that their responses were better than by chance. Table 4.4 presents the results of the one-sample *t*-test of Criterion Test Phase in addition to the mean and *SD* of both the Yes-No Identification and Forced Choice Identification tasks.

Table 4.4: *Results of the One Sample T-test in the Criterion Test Phase (Two Tasks)*

Orthographic input type		Identical-phoneme condition (Out of 12)		Similar-phoneme condition (Out of 12)		New-phoneme condition (Out of 12)	
		Yes-No	Forced Choice	Yes-No	Forced Choice	Yes-No	Forced Choice
Arabic-script	*M*	8.64	9.84	10.2	11.04	9.92	10.44
	SD	1.86	2.13	1.55	1.56	2.73	2.39
	p	<.001	<.001	<.001	<.001	<.001	<.001
Romanised-script	*M*	6.56	7.92	9.6	10	9.16	9
	SD	2.84	3.13	2.67	2.36	2.88	2.27
	p	.33	.005	<.001	<.001	<.001	<.001
No-script	*M*	9.32	9.96	10.4	11.36	11.28	11.12
	SD	2.71	1.88	1.58	0.86	1.3	1.66
	p	<.001	<.001	<.001	<.001	<.001	<.001
Total	*M*	8.17	9.24	10.06	10.8	10.12	10.18
	SD	2.74	2.58	2	1.78	2.54	2.28

Moving on to the Phonological Test Phase, the results of the one-sample *t*-test were significant in all cases except for one case, the same one as in the Criterion Test Phase, i.e., the identical-phoneme condition for the Romanised-script group, $p = .39$, as shown in Table 4.5.

Table 4.5: *Results of the One Sample T-test in the Phonological Test Phase*

Orthographic input type		Identical-phoneme condition (Out of 12)	Similar-phoneme condition (Out of 12)	New-phoneme condition (Out of 12)
Arabic-script	*M*	8.36	7.68	10.16
	SD	2.11	2.42	2.26
	p	<.001	.002	<.001
Romanised-script	*M*	6.44	6.8	8.32
	SD	2.51	1.63	2.64
	p	.39	.02	<.001
No-script	*M*	8.76	8.08	10.52
	SD	2.33	2.21	1.93
	p	<.001	<.001	<.001
Total	*M*	7.85	7.52	9.66
	SD	2.51	2.15	2.46

The results of the one-sample *t*-test in the Post-Sound Recognition Phase were significant in all cases. This indicated that their responses were better than by chance, as shown in Table 4.6.

Table 4.6: *Results of the One Sample T-test in the Post-Sound Recognition Phase*

Orthographic input type		Identical-phoneme condition (Out of 12)	Similar-phoneme condition (Out of 12)	New-phoneme condition (Out of 12)
Arabic-script	*M*	11.84	9.84	11.92
	SD	0.55	1.81	0.40
	p	<.001	<.001	<.001
Romanised-script	*M*	11.60	9.52	10.04
	SD	0.81	1.85	1.74
	p	<.001	<.001	<.001
No-script	*M*	11.44	10.24	11.68
	SD	1.08	1.66	0.74
	p	<.001	<.001	<.001
Total	*M*	11.62	9.86	11.54
	SD	0.85	1.78	1.16

4.4 Main Analyses of the Study

4.4.1 Analysis of the Effect of L1/L2 Phoneme Similarity and Orthographic Availability on Phoneme Discrimination

As mentioned in the methodology chapter, two types of data were collected in this phase to test the effect of L1/L2 phoneme similarity and orthographic input type on L2 phonological development: accuracy and reaction time in correct trials. A two-way [Orthographic input type (Arabic script, Romanised with Arabic scripts, No script) by L1/L2 phoneme similarity (identical phonemes, similar phonemes, new phonemes)] ANOVA was carried out on the Phonological Test Phase as well as the Post-Sound Recognition Phase for both accuracy and reaction time. This was done to answer *RQs*1, 2, and 3, which are:

RQ1: Does the presence of orthography help participants differentiate between two phonemes in the case of:

- An unfamiliar script such as Arabic script; or
- A Combination of Familiar Script such as Romanised-Arabic Script and Unfamiliar Script such as Arabic Script?

RQ2: Does L1/L2 phoneme similarity affect the difficulty of participants' ability to differentiate between two phonemes?

RQ3: Is the effect of orthographic availability influenced by L1/L2 phoneme similarity?

This stage of analysis included the participants responses in only the Phonological Test Phase and the Post-Sound Recognition Phase as the main analysis of the study. As mentioned above, because this stage of analysis consists of repeated measures, the sphericity assumption should be met. As shown in Table 4.7, the assumption of sphericity was met for the Phonological Test Phase in both measures, but it was not met in the Post-Sound Recognition Phase in both measures and so a Greenhouse Geisser correction was therefore applied.

Table 4.7: *Mauchly's Test Results for Sphericity in a Phonological Test and Post-Sound Recognition*

Variable	*p*	Decision
Phonological Test: Accuracy	.98	Sphericity was perfectly met
Phonological Test: Reaction-time	.17	Sphericity was met
Post-Sound Recognition: Accuracy	<.001	Sphericity was not met
Post-Sound Recognition: Reaction-time	<.001	Sphericity was not met

4.4.1.1 Phonological Test Phase

4.4.1.1.1 Analysis of Accuracy Data

See Table 4.5 for descriptive statistics of the accuracy data of participants' responses in the Phonological Test Phase.

As shown in Table 4.8, a two-way ANOVA revealed that there was a main effect of L1/L2 phoneme similarity on the participants' accuracy. There was also a main effect of orthographic input type on the participants' accuracy. The interaction between L1/L2 phoneme similarity and orthographic input type was not significant. However, in this analysis, when looking at the distribution of normality and equality of variance as shown in Table O4 and O6 respectively in Appendix O, it was found that there was no normal distribution and equality of variance in some cases. Therefore, results of the non-parametric tests are presented here. The results of the non-parametric tests (carried out to validate results because of violations of assumptions for ANOVA) showed that L1/L2 phoneme similarity was significant with the Arabic-script and No-script groups, but not the Romanised-script group. Orthographic input type, on the other hand, was significant in the identical-phoneme and new-phoneme conditions, but not the similar-phoneme condition[3], as shown in Appendices P and Q.

Table 4.8: *Analysis of the Main Effects of L1/L2 Phoneme Similarity, Orthographic Input Type and the Interaction Between them in the Phonological Test Phase*

Variable	*F*	*p*	Eta Partial Squared	Observed Power
L1/L2 phoneme similarity	26.04	<.001	.266	1.000
Orthographic input type	10.44	<.001	.225	.985
Interaction between L1/L2 phoneme similarity and orthographic input type	0.71	.583	.019	.227

The contrast tests with Bonferroni correction results for L1/L2 phoneme similarity as a main effect showed that accuracy in the identical-phoneme condition was not significantly different from that in the similar-phoneme condition ($p = .91$). However, the accuracy in the new-phoneme condition was significantly better than in the identical-phoneme condition and the similar-phoneme condition ($p < .001$). As for the orthographic input type, the results of contrast tests as a main effect of orthography input type indicated that participants in the Romanised-script group had lower accuracy than the Arabic-script group ($p = .003$), and the No-script group ($p < .001$).

[3] As parametric tests showed that there was no significant interaction, and non-parametric tests suggested that there were interesting differences between the orthographic input type groups and L1/L2 phoneme similarity levels, this may be worn for other investigations.

However, by comparing the Arabic-script group's accuracy with that of the No-script group, it is shown that the No-script group had no significant difference at all in their accuracy from those in the Arabic-script group (p = 1) (see Figs 4.3, 4.4, and 4.5 for Means).

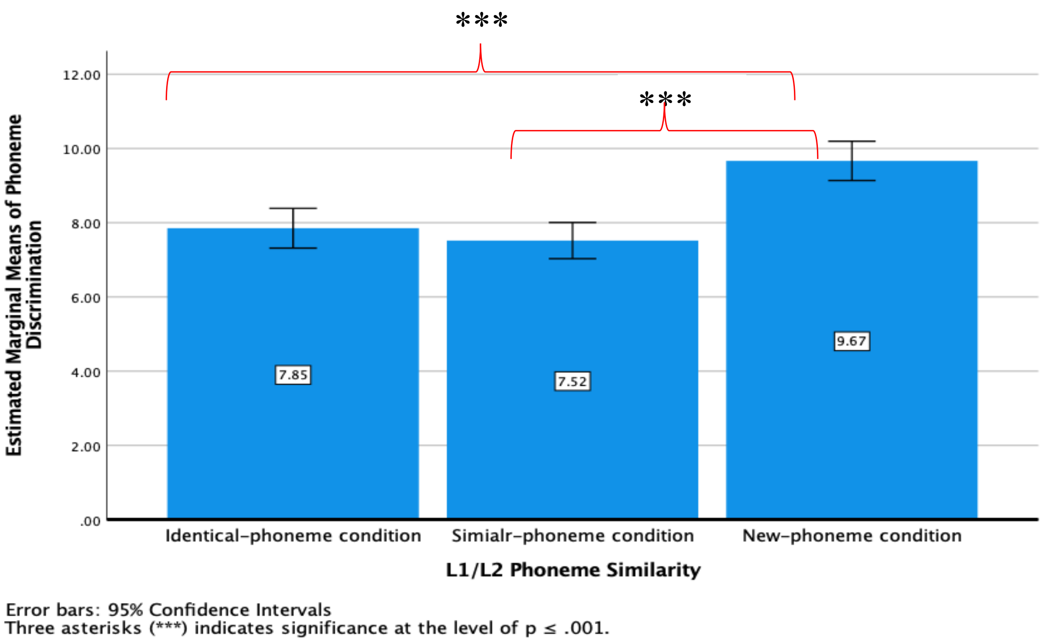


Figure 4.3 Accuracy in the three conditions of L1/L2 phoneme similarity (Phonological Test Phase)

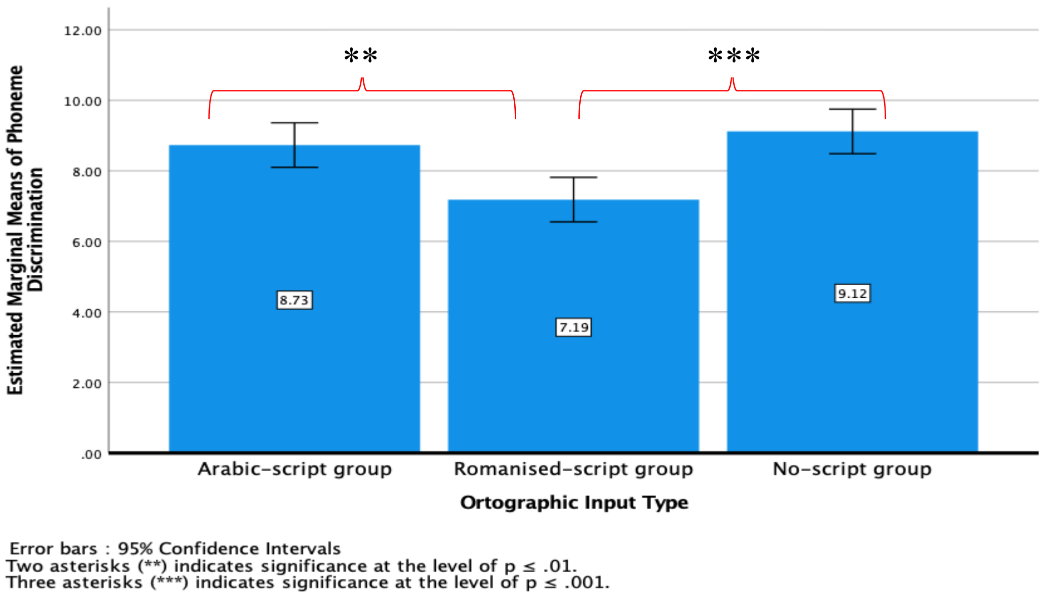


Figure 4.4 Accuracy in the three different orthographic input types (Phonological Test Phase)

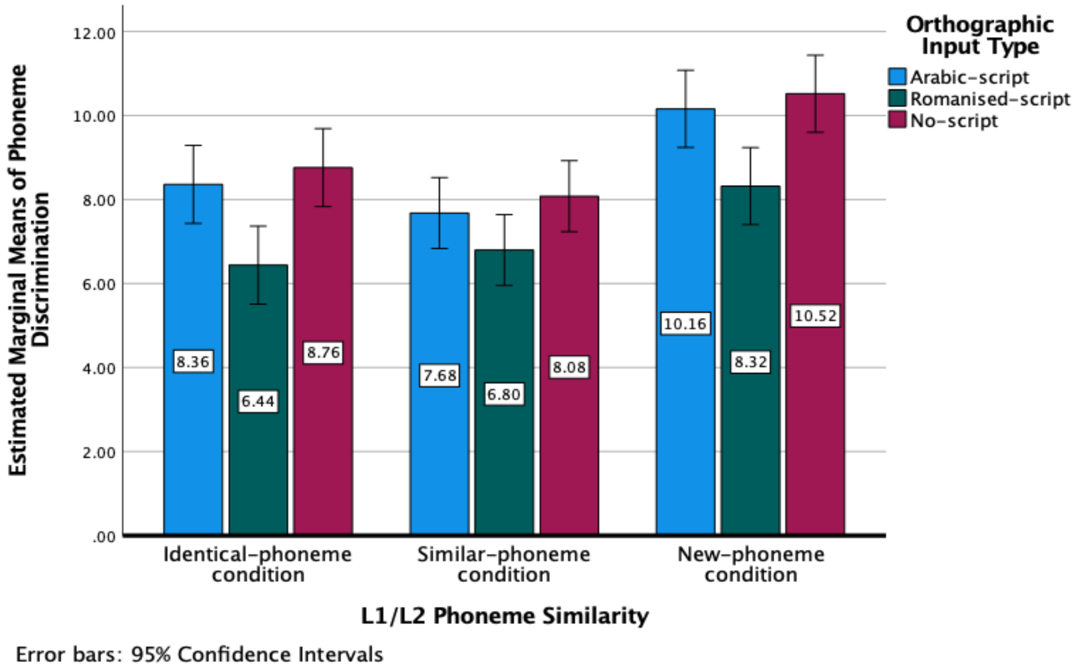


Figure 4.5 Accuracy for the three different orthographic input types in the three conditions of L1/L2 phoneme similarity (Phonological Test Phase)

4.4.1.1.2 Analysis of Reaction Time Data

Table 4.9 presents descriptive statistics for the participants' reaction time (on correct responses only) in the Phonological Test Phase.

Table 4.9: *Descriptive Statistics for Reaction Time in the Phonological Test Phase*

Orthographic input type	*N*		Identical-phoneme condition (Out of 12)	Similar-phoneme condition (Out of 12)	New-phoneme condition (Out of 12)
Arabic-script	25	*M*	2.25	2.03	2.13
		SD	0.87	0.88	1.16
Romanised-script	25	*M*	2.39	2.1	2.23
		SD	1.1	1.2	1.01
No-script	25	*M*	2.35	1.95	2.2
		SD	0.85	0.92	0.95
Total	75	*M*	2.33	2.03	2.19
		SD	0.93	1.00	1.03

As shown in Table 4.10, the results showed that there was a main effect of L1/L2 phoneme similarity on the time participants took to respond. Analysis of orthographic input type showed that there was no significant difference in the reaction times of the three orthographic input-type groups. The interaction was also not significant.

Inconsistently with accuracy measure, results of contrast tests with a Bonferroni correction showed that the reaction time in the similar-phoneme condition was significantly faster than that in the identical-phoneme condition (p = .002). However, the reaction time in the new-phoneme condition was not significantly different from that in the identical-phoneme condition (p =.23) and similar phoneme condition (p = .31) (see Figs 4.6, 4.7, and 4.8 for Means).

Table 4.10: *Analysis of the Main Effects of L1/L2 Phoneme Similarity, Orthographic Input Type and Interaction Between them on Reaction Time in the Phonological Test Phase*

Variable	F	p	Eta Partial Squared	Observed Power
L1/L2 phoneme similarity	5.94	.003	.076	.873
Orthographic input type	0.08	.92	.002	.062
Interaction between L1/L2 phoneme similarity and orthographic input type	.22	.92	.006	.097

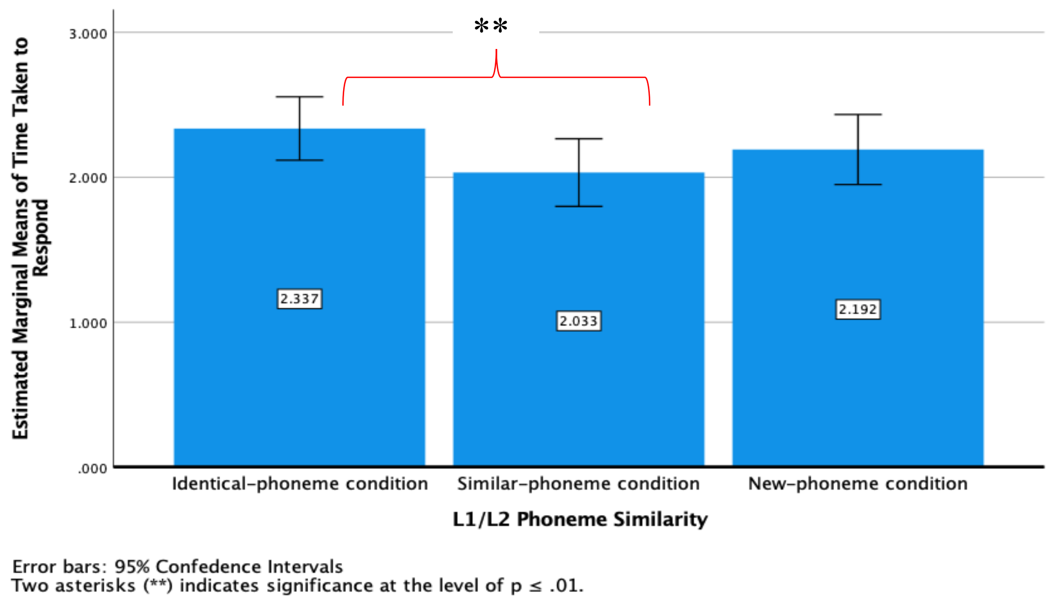


Figure 4.6 Reaction time in the three conditions of L1/L2 phoneme similarity (Phonological Test Phase)

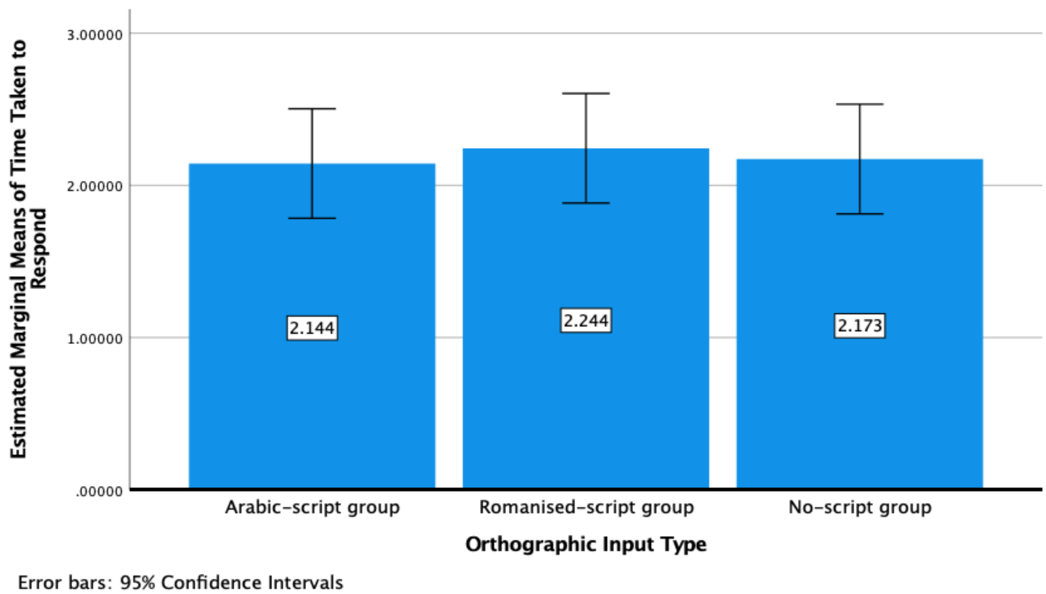


Figure 4.7 Reaction time for the three different orthographic input types (Phonological Test Phase)

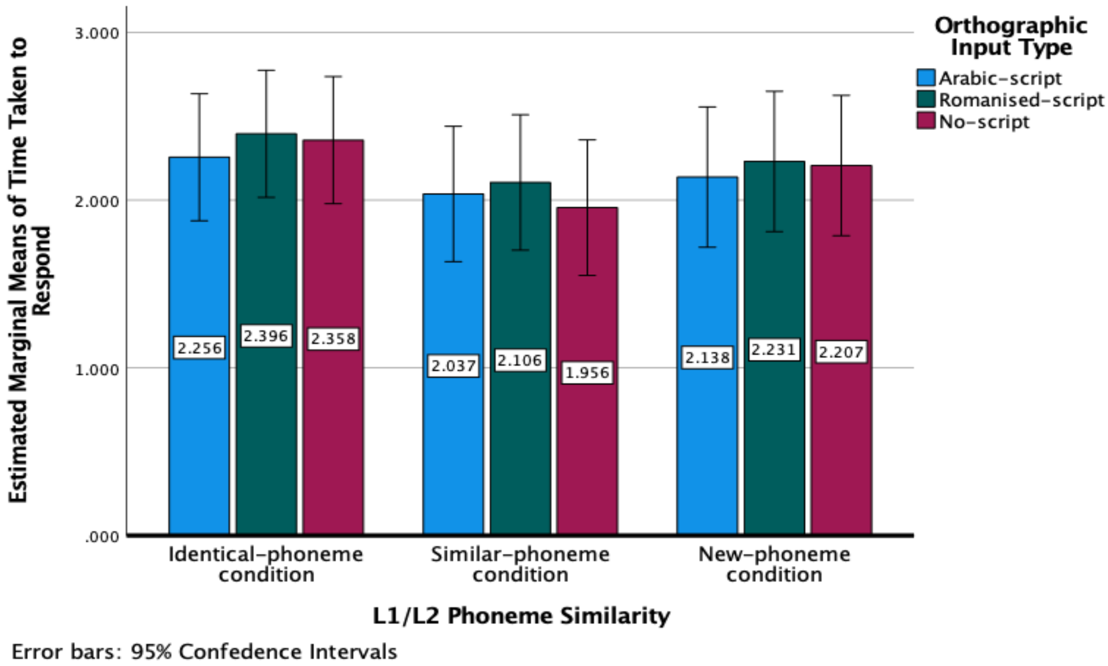


Figure 4.8 Reaction time of the three different orthographic input types in the three conditions of L1/L2 phoneme similarity (Phonological Test Phase)

4.4.1.2 Post-Sound Recognition Phase.

4.4.1.2.1 Analysis of the Accuracy Data.

Refer to Table 4.6 for descriptive statistics for the accuracy data of participants' responses in the Post-Sound Recognition Phase.

As shown in Table 4.11, two-way ANOVA reveals that there was a main effect of L1/L2 phoneme similarity on participants' accuracy. There was no main effect of orthographic input type on participants' accuracy. The interaction between L1/L2 phoneme similarity and orthographic input type was not significant.

Results of contrast tests with a Bonferroni correction showed that accuracy in the similar-phoneme condition was significantly lower than that in the identical-phoneme and new-phoneme conditions (p <.001). However, their accuracy in in the identical-phoneme condition did not significantly differ from that in the new-phoneme condition (p =1) (see Figs 4.9, 4.10, and 4.11 for Means).

Table 4.11: *Analysis of the Main Effects of L1/L2 Phoneme Similarity, Orthographic Input Type and the Interaction Between them in the Post-Sound Recognition Phase*

Variable	F	p	Eta Partial Squared	Observed Power
L1/L2 phoneme similarity	44.40	<.001	.381	1.000
Orthographic input type	2.75	.07	.071	.527
Interaction between L1/L2 phoneme similarity and orthographic input type	1.34	.26	.036	.281

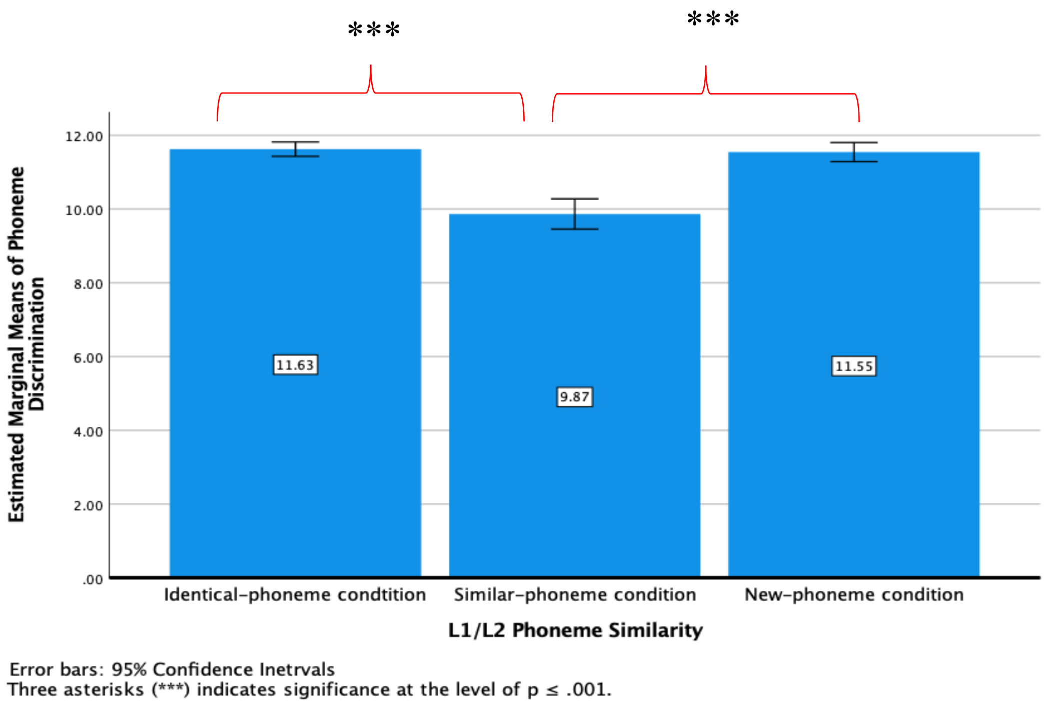


Figure 4.9 Accuracy in the three conditions of L1/L2 phoneme similarity (Post-Sound Recognition Phase)

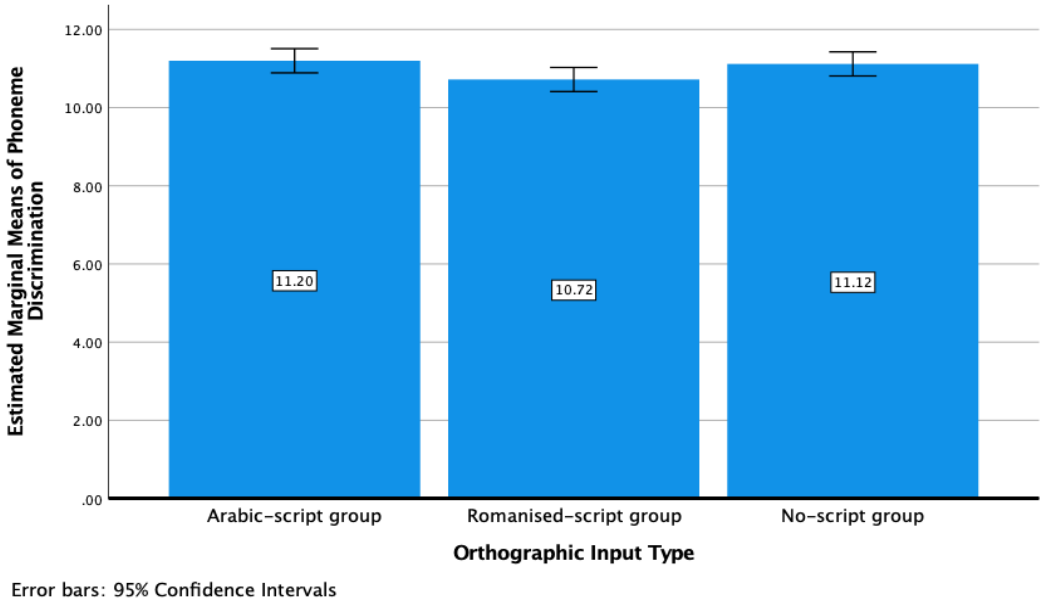


Figure 4.10 Accuracy for the three orthographic input types (Post-Sound Recognition Phase)

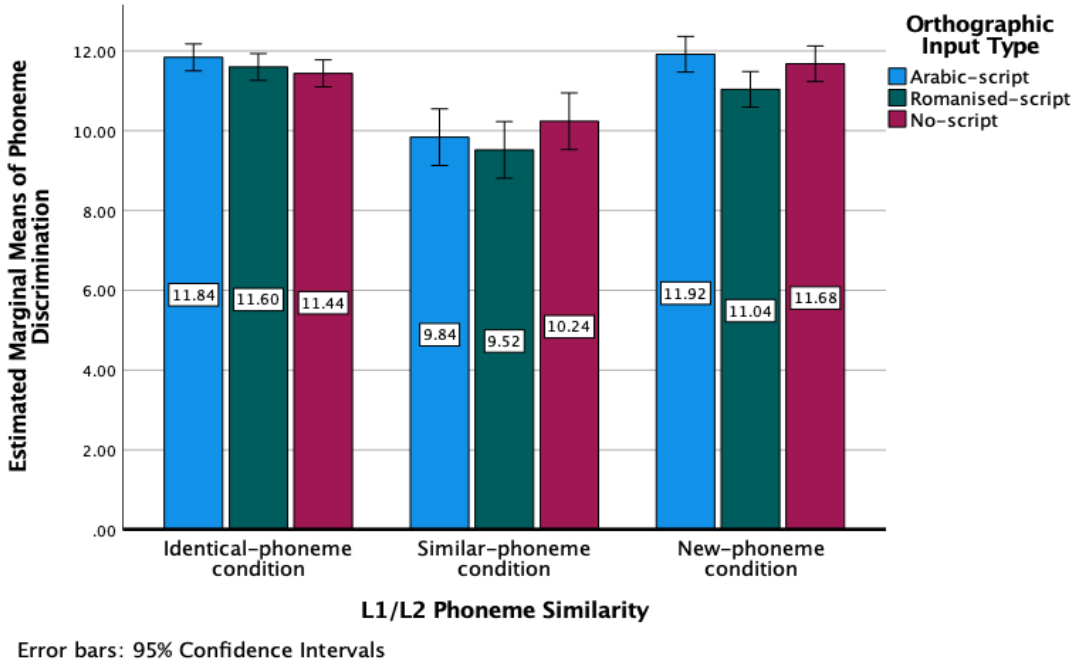


Figure 4.11 Accuracy for the three different orthographic input types in the three conditions of L1/L2 phoneme similarity (Post-Sound Recognition Phase)

4.4.1.2.2 Analysis of the Reaction Time Data.

Table 4.12 presents descriptive statistics for the participants' reaction time in their correct responses in the Post-Sound Recognition Phase in the three conditions of L1/L2 phoneme similarity and three orthographic input-type groups.

Table 4.12: *Descriptive Statistics for Reaction Time in the Post-Sound Recognition Phase*

Orthographic input type	*N*		Identical-phoneme condition (Out of 12)	Similar-phoneme condition (Out of 12)	New-phoneme condition (Out of 12)
Arabic-script	25	*M*	1.21	0.90	1.03
		SD	0.48	0.50	0.43
Romanised-script	25	*M*	1.27	0.83	0.96
		SD	0.53	0.38	0.45
No-script	25	*M*	1.28	1.00	1.19
		SD	0.49	0.41	0.45
Total	75	*M*	1.26	0.91	1.06
		SD	0.49	0.43	0.45

As shown in Table 4.13, two-way ANOVA reveals that there was a main effect of L1/L2 phoneme similarity on reaction time. Orthographic input type effect was not significant. Two-way ANOVA showed that these two effects (L1/L2 phoneme similarity and orthographic input type) had no significant interaction between them.

Table 4.13: *Analysis of the Main Effects of L1/L2 Phoneme Similarity, Orthographic Input Type and the Interaction Between them on Reaction Time in the Post-Sound Recognition Phase*

Variable	*F*	*p*	Eta Partial Squared	Observed Power
L1/L2 phoneme similarity	19.38	<.001	.212	.999
Orthographic input type	0.92	.40	.025	.203
Interaction between L1/L2 phoneme similarity and orthographic input type	0.87	.50	.021	.213

Contrast tests with a Bonferroni correction results showed that the reaction time in all three conditions of L1/L2 phoneme similarity differed significantly. Inconsistently with accuracy findings, the fastest reaction time was recorded in the similar-phoneme condition, which was significantly faster than those in the identical-phoneme condition ($p < .001$) and new-phoneme condition ($p = .04$). The reaction time in the new-phoneme condition was significantly faster than that in the identical-phoneme condition ($p < .001$) (see Figs 4.12, 4.13, and 4.14 for Means).

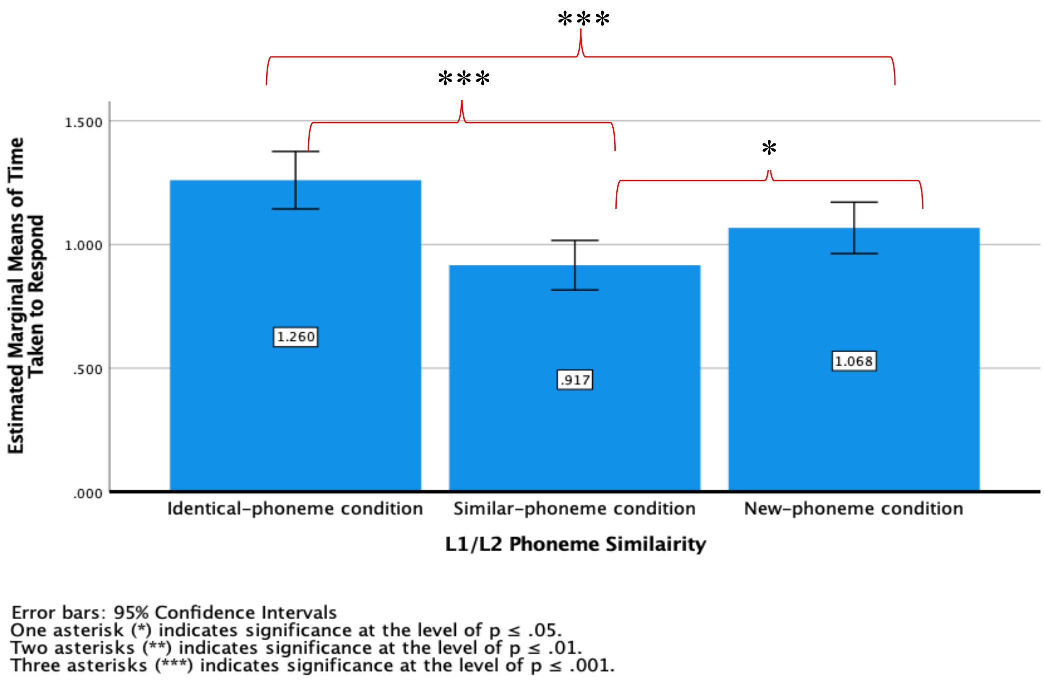


Figure 4.12 Reaction time in the three conditions of L1/L2 phoneme similarity (Post-Sound Recognition Phase)

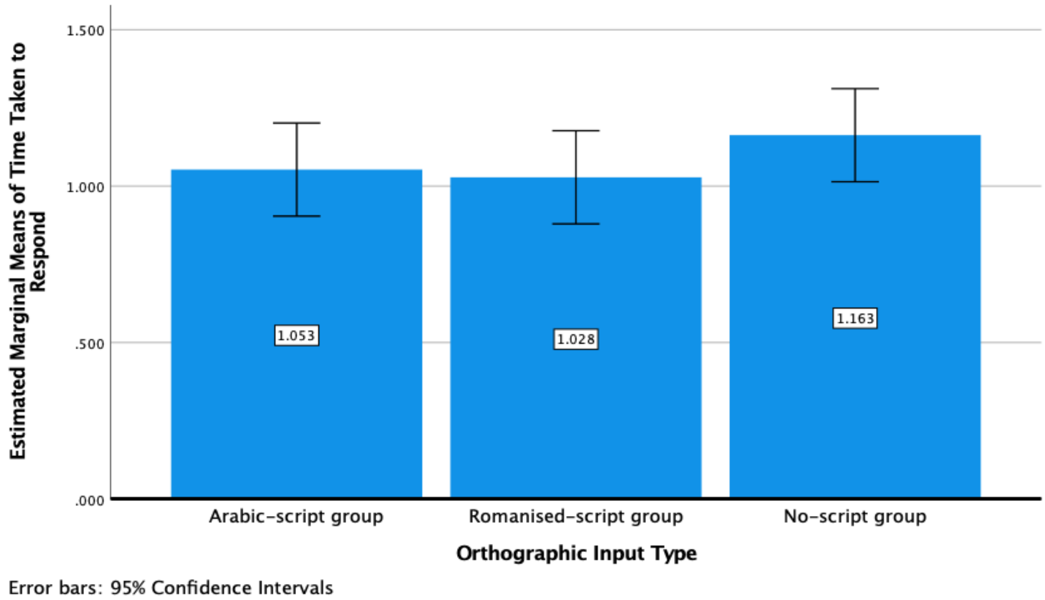


Figure 4.13 Reaction time for the three different orthographic input types (Post-Sound Recognition Phase)

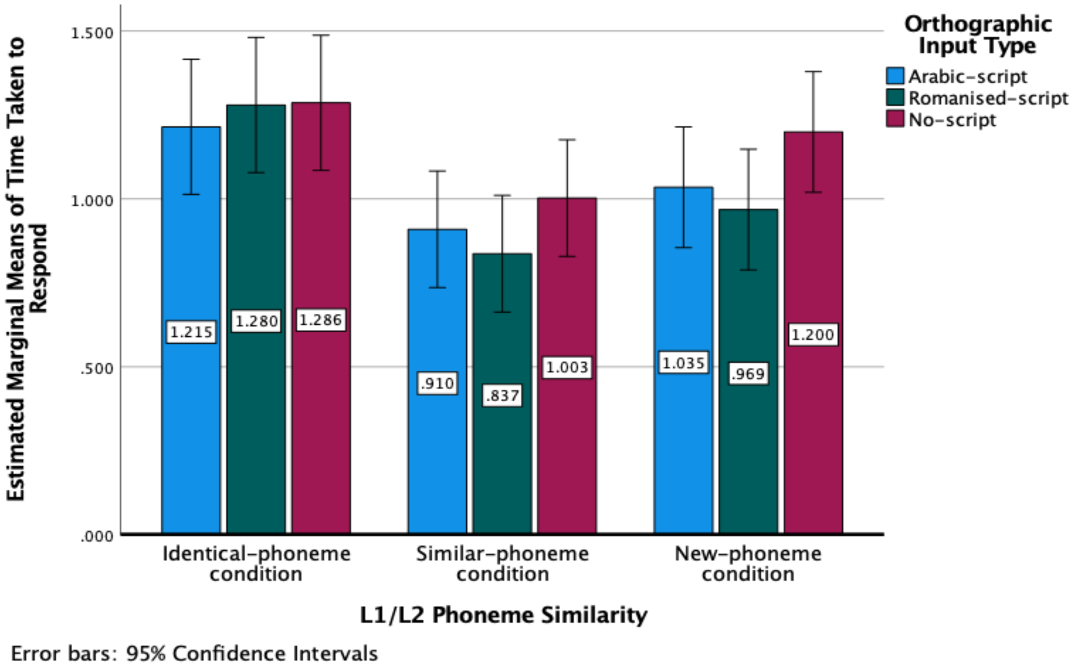


Figure 4.14 Reaction time for the three different orthographic input types in the three conditions of L1/L2 phoneme similarity (Post-Sound Recognition Phase)

4.4.2 Correlational Analysis between Participants' Performance in the Phonological Test Phase Assessing Word-Level Phoneme Discrimination and Post-Sound Recognition Phase Assessing Phoneme-Level Phoneme Discrimination

The data analysed in this study was obtained from two types of phonemes discrimination: the Phonological Test Phase which involved word-level discrimination, and the Post-Sound Recognition Phase which involved phoneme-level discrimination.

To answer the fourth research question: Does the participants' ability to discriminate phonemes when embedded within words (word-level discrimination) correlate with their ability to discriminate phonemes when uttered in isolation (phoneme-level discrimination)?, a correlational analysis between participants' responses in the Phonological Test Phase assessing word-level phoneme discrimination and Post-Sound Recognition Phase assessing phoneme-level phoneme discrimination was conducted on the two measures of the study, accuracy and reaction time, separately. As the assumptions underlying parametric tests were not fully met (as described in Appendix O), the researcher conducted the Spearman's correlation.

A Spearman correlation analysis of the participants' performance in terms of both accuracy and reaction time showed that, as displayed in Table 4.14, none of the conditions in the Phonological Test Phase correlated with performance in the Post-Sound Recognition Phase indicating that there is no relationship between the participants' performance in the Post-Sound Recognition Phase and the Phonological Test Phase.

Table 4.14: *Correlational Analysis of Participants' Performance and Reaction Time Between the Phonological Test Phase and the Post-Sound Recognition Phase*

Measure	Post-Sound Recognition Phase	Phonological Test Phase	r_s	*p* (2-tailed)	95% CI (2-tailed) Lower	Upper
Accuracy	Identical	Identical	.144	.21	-.093	.365
	Similar	Similar	.078	.50	-.158	.306
	New	New	.040	.73	-.195	.271
Reaction Time	Identical	Identical	.140	.23	-.096	.362
	Similar	Similar	.114	.33	-.123	.338
	New	New	.207	.07	-.028	.420

4.4.3 Analysis of Phonemic Coding Ability as a Predictor Variable

Before answering *RQ*5 (Is the effect of orthographic availability influenced by the combination of phonemic coding ability and L1/L2 phoneme similarity?), a correlational analysis between all independent variables (phonemic coding ability, orthographic input type, and L1/L2 phoneme similarity) and the dependent variable (participants' performance) in the Phonological Test Phase and Post-Sound Recognition Phase was conducted separately on the two measures of the study, accuracy and reaction time (Table 4.15).

A shown in Table 4.15, the performance of the participants in the Phonological Test Phase in terms of accuracy correlated with the participants' phonemic coding ability scores and L1/L2 phoneme similarity, $r_s = .14$, $p = .02$ and $r_s = .29$, $p < .001$ respectively. However, there was no correlation between phonemic coding ability and the participants' accuracy in the Post-sound Recognition Phase. The reaction time of the participants in the Phonological Test Phase also did not show any significant correlation with the participants' phonemic coding ability scores in both Phonological Test Phase and Post-sound Recognition Phase.

Table 4.15: *Correlational Analysis of Participants' Performance and Reaction Time in the Phonological Test Phase and Post-Sound Recognition Phase with all independent variables*

Phase	Measure	Variable	r_s	*p* (2-tailed)	95% CI (2-tailed) Lower	Upper
Phonological Test	Accuracy	Phonemic coding ability	.14	.02	.013	.277
		Phoneme similarity	.29	<.001	.166	.412
		Orthographic input type	.06	.32	-.069	.199
	Reaction time	Phonemic coding ability	.08	.2	-.050	.217
		Phoneme similarity	-.05	.38	-.191	.077
		Orthographic input type	.01	.86	-.123	.146
Post-Sound Recognition:	Accuracy	Phonemic coding ability	-.001	.98	-.136	.133
		Phoneme similarity	-.006	.92	-.141	.128
		Orthographic input type	-.04	.51	-.177	.091
	Reaction time	Phonemic coding ability	.01	.87	-.124	.145
		Phoneme similarity	-.14	.02	-.277	-.014
		Orthographic input type	.09	.13	-.036	.231

Given that phonemic coding ability correlated only with accuracy data in the Phonological Test Phase, multiple linear regression was carried out to test if there was a significant effect of phonemic coding ability as a predictor only on the accuracy data for the Phonological Test Phase. Regression analysis were not carried on on data from the Post-Sound Recognition Phase as well as reaction time in the Phonological Test Phase.

To answer the final research question, to determine how changes in the independent variables (phonemic coding ability score, orthographic input type, and L1/L2 phoneme similarity) are associated with changes in the dependent variables (accuracy in the Phonological Test Phase), the multiple linear regression analysis was used, where all variables are included simultaneously in the model. The categorical predictor variables were coded using the dummy coding system. To investigate the effect of orthography input type, two dummy coded variables were created, one encoding the comparison between the Arabic-script group and the No-script group and the other encoding the comparison between the Romanised-script group and the No-script group. In addition, to investigate the effect of the L1/L2 phoneme similarity a further two dummy variables were created, each with the new-phoneme condition as the reference category. The first allowed us to test for the difference between the identical-phoneme condition and new-phoneme condition and the second allowed us to test the difference between the similar-phoneme condition and the new-phoneme condition.

Multiple linear regression was run to test if phonemic coding ability score, orthographic input type, and L1/L2 phoneme similarity significantly predicted the phoneme discrimination score during the Phonological test phase. The fitted regression model was:

Phoneme discrimination score = 10.077 + 0.033*(Phonemic coding ability score) - 0.404*(Arabic-script group) - 1.865*(Romanised-script group) - 1.813*(Identical-phoneme condition) - 2.147*(Similar-phoneme condition)

The overall regression was statistically significant (R^2 = .25, *F* (5, 219) = 14.68, $p < .001$). In these results, the model explains approximately 25% of the variation in the accuracy data during the Phonological test phase.

As shown in Table 4.16, it was found that phonemic coding ability score did not significantly predict the phoneme discrimination score during the Phonological test phase ($\beta = 0.03$, $p = .15$).

Table 4.16: *Multiple Linear Regression of L1/L2 Phoneme Similarity, Orthographic Input Type, Phonemic Coding Ability on Participants' Responses in the Phonological Test Phase*

Variable	Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for *B*	
	B	Std. Error	Beta	*t*	*p*	Lower Bound	Upper Bound
PCA_Score	.03	.023	.085	1.42	.15	-.013	.078
Arabic-script vs. No-script	-.40	.365	-.075	-1.10	.27	-1.124	.317
Romanised-script vs. No-script	-1.86	.368	-.345	-5.06	<.001	-2.591	-1.140
Identical-phoneme vs. New-phoneme	-1.81	.365	-.335	-4.96	<.001	-2.533	-1.093
Similar-phoneme vs. New-phoneme	-2.14	.365	-.397	-5.87	<.001	-2.867	-1.427

It was found that the dummy variable encoding the Arabic-script group relative to the No-script group did not significantly predict the phoneme discrimination score ($\beta = -0.4$, $p = .27$). Therefore, there was no difference between these two groups. The dummy variable encoding the Romanised-script group relative to the No-script group was a significant predictor ($\beta = -1.86$, $p < .001$). The negative beta, however, demonstrates that the phoneme discrimination score during the Phonological Test Phase for the Romanised-script group was 1.865 less than the control group (No-script group), and this was significant.

It was found that the dummy variable encoding the identical-phoneme condition relative to the new-phoneme condition significantly predicted the phoneme discrimination score ($\beta = -1.81$, $p < .001$). The negative beta demonstrates that the phoneme discrimination score in the identical-phoneme condition is 1.81 less than in the new-phoneme condition. Similarly, the dummy variable encoding the similar-phoneme condition relative to the new-phoneme condition significantly predicted the phoneme discrimination score (β = -2.14, $p < .001$). The negative beta demonstrates that the phoneme discrimination score in the identical-phoneme condition is 2.14 less than in the new-phoneme condition.

4.5 Summary of the Chapter

This chapter conducted a detailed statistical analysis of the effect of orthographic input in relation to L1/L2 phoneme similarity and/or phonemic coding ability on participants' discrimination of L2 sounds. Two measures were looked at in this study: accuracy and reaction time.

Beginning with the first measure, the statistical analysis of the participants' responses in the Phonological Test Phase and Post-Recognition Phase concluded with findings that can be summarised as follows:

- Accuracy of the participants did not significantly differ when asked to discriminate between two identical or two similar phonemes in the Phonological Test Phase.
- Accuracy of the participants did not significantly differ when asked to discriminate between two identical or two new phonemes in the Post-Sound Recognition Phase.
- Accuracy of the participants was significantly higher when they were asked to discriminate between two new phonemes in the Phonological Test Phase than when they were asked to discriminate between two identical or two similar phonemes.
- Accuracy of the participants was significantly lower when they were asked to discriminate between two similar phonemes in the Post-Sound Recognition Phase than when they were asked to discriminate between two identical or two new phonemes.

- Accuracy of the participants in the Arabic-script group was not significantly different from that in the No-script group in the Phonological Test Phase.
- Accuracy of the participants in the Romanised-script group was significantly the lowest among the three orthographic input-type groups in the Phonological Test Phase.
- Accuracy of the participants was not affected significantly by orthographic input type in the Post-Sound Recognition Phase.
- L1/L2 phoneme similarity (identical, similar, new) did not show any significant interaction with orthographic input type (Arabic, Romanised along with Arabic, no script) in both the Phonological Test Phase and the Post-Sound Recognition Phase. Because of some slight differences which were shown by non-parametric results, these findings may be worn for further investigations.
- The responses in the Phonological Test Phase did not correlate with those in the Post-Sound Recognition Phase.
- Scores in the phonemic coding ability test did not significantly predict the phoneme discrimination score during the Phonological test phase.

As for the reaction time, statistical analysis of the participants' reaction time in the Phonological Test Phase and Post-Recognition Phase concluded with findings that can be summarised as follows:

- Reaction time did not significantly differ when asked to discriminate between two identical or between two new phonemes in the Phonological Test Phase.
- Reaction time was significantly faster when participants were asked to discriminate between two similar phonemes than when asked to discriminate between two identical phonemes in the Phonological Test Phase.

- Reaction time did not significantly differ when asked to discriminate between two similar or between two new phonemes the Phonological Test Phase.
- Reaction time was the fastest when the participants were asked to discriminate between two similar phonemes in the Post-Sound Recognition Phase.
- Orthographic input type had no significant effect on discriminating phonemes in the Phonological Test Phase as well as the Post-Sound Recognition Phase, in either the identical-phoneme, similar-phoneme, or new-phoneme conditions.
- L1/L2 phoneme similarity (identical, similar, new) did not interact significantly with orthographic input type (Arabic, Romanised along with Arabic, no script) in the Phonological Test Phase as well as the Post-Sound Recognition Phase.
- Reaction time in the Phonological Test Phase did not correlate with those in the Post-Sound Recognition Phase.
- Because phonemic coding ability did not correlate with reaction time in any condition, regression was not carried out on this measure.

5 Discussion

5.1 Introduction

This chapter presents a detailed explanation of the findings discussed in Chapter 4 and provides answers to the five research questions. As previously mentioned, this study aimed to explore the effect of orthographic input on Arabic phonological development by L1 English speaker with regard to the participants' phonemic coding ability level in different situations. These situations included having an unfamiliar script, unfamiliar and familiar scripts together, or no script at all in the acquisition of identical to English, similar to English, and novel phonemes. Participants were allocated into three orthographic groups after taking a phonemic coding ability test (LLAMA-E). The first group was the Arabic-script group where the participants were exposed to written, visual and auditory input. Written input was presented to the participants in Arabic script along with non-object corresponding images while hearing their auditory forms. The second group was the Romanised-script group where the participants were also exposed to written, visual and auditory input. The difference was that written input was presented to the participants in both Arabic and Romanised scripts along with non-object corresponding images while hearing their auditory forms. The third group was the No-script group where the participants had the same input as the other groups except for the written input being replaced by a string of letters (طططط) which corresponds to (xxxx) in English. Phonemic coding ability was included in this analysis as a predictor variable to measure its role in the effect of orthographic input type and/or L1 and L2 phoneme similarity on the participants' ability to acquire three different pairs of Arabic phonemes. The study sought to investigate the following research questions:

RQ1: Does the presence of orthography help participants differentiate between two phonemes in the case of:

- An unfamiliar script such as Arabic script; or
- A Combination of Familiar Script such as Romanised-Arabic Script and Unfamiliar Script such as Arabic Script?

RQ2: Does L1/L2 phoneme similarity affect the difficulty of participants' ability to differentiate between two phonemes?

RQ3: Is the effect of orthographic availability influenced by L1/L2 phoneme similarity?

RQ4: Does the participants' ability to discriminate phonemes when embedded within words (word-level discrimination) correlate with discriminating phonemes when uttered in isolation (phoneme-level discrimination)?

RQ5: Is the effect of orthographic availability influenced by the combination of phonemic coding ability and L1/L2 phoneme similarity?

5.1.1 Chapter Outline

This chapter consists of two main sections. The next section (Section 5.2) presents a general summary of the findings. The following section (Section 5.3) presents a general discussion of the findings of the 2-way ANOVA that measured the effect of orthographic input type on the acquisition of three pairs of the Arabic phonemes to answer *RQ*1, the effect of L1/L2 phoneme similarity on the acquisition of three pairs of Arabic phonemes to answer *RQ*2, the findings of the interaction between the two effects (orthographic input type and L1/L2 phoneme similarity) to answer *RQ*3, the findings of the correlational analysis between participants' responses in a word-level discrimination task and a phoneme-level discrimination task which was carried out to answer

*RQ*4, and finally the findings of the multiple linear regression which was carried out to measure the prediction of phonemic coding ability on the performance of the participants in three orthographic input-type groups in the three conditions of L1/L2 phoneme similarity, to answer *RQ*5.

5.2 Summary of Findings

The performance of the participants in these three groups was analysed in terms of two measures: accuracy and reaction time in two different levels of phoneme discrimination; word-level (phonemes embedded in words assessed in the Phonological Test Phase) and phoneme-level (isolated phonemes assessed in the Post-Sound Recognition Phase). In word-level discrimination, the participants were tested after being exposed to a Learning Phase and a Criterion Test Phase, which was part of their learning. Following these phases, a Phonological Test Phase took place. In this phase, the participants were asked to discriminate between two minimal pairs that were contrasted in the initial position with the two target phonemes by matching them to a representative picture. Following this phase, a Post-Sound Recognition Phase took place. In this phase, which was assessing phoneme-level discrimination, the participants were exposed to 12 trials where each trial included two phonemes uttered in isolation (without any other following or preceding phoneme). The participants' task was to decide whether the two recordings represented identical or different sounds.

Regarding the accuracy of the participants' responses, it was found that L1/L2 phoneme similarity played a significant role in the participants' performance in both phoneme discrimination levels. In terms of word-level discrimination, new phonemes were easier than both similar and identical pairs, but there was no difference between participants' discrimination of similar and identical phonemes. However, this was not the case when participants were asked to discriminate between isolated phonemes. Their discrimination of a similar phonemes pair was the poorest whereas discrimination of a new phonemes pair did not significantly differ from an identical phonemes pair.

Contrary to expectations, orthographic input (as described below) was found to affect learning negatively, but only in the case of word-level discrimination; it did not play any significant role in phoneme-level discrimination. The negative effect of orthography in word-level discrimination was restricted to the case where Romanised script was provided to the participants along with Arabic script. Having more than one script at the same time was found to impair learning as the Romanised-script group had the lowest performance among the three groups. Novel script alone, on the other hand, did not have any effect on the participants' performance, as the Arabic-script group did not significantly differ from the group that did not have any orthographic input. In addition, results showed that there was no interaction between orthographic input type and L1/L2 phoneme similarity. This finding, however, needs further investigation as non-parametric tests showed that, in the word-level discrimination, the effect of orthographic input on the participants' ability to discriminate between phonemes was not the same in the three levels of L1/L2 phoneme similarity.

Moreover, correlational analysis showed that participants' ability to discriminate phonemes in word-level discrimination did not correlate with their ability to discriminate phonemes in phoneme-level discrimination in all conditions. Finally, phonemic coding ability as a predictor variable was found to have no significant effect on participants' performance.

As for reaction time, it was expected that it would be faster in the case of new phonemes than in the case of similar ones, as it is easier for the participants to grasp the difference between two phonemes. In addition, it was expected that reaction time would also be faster in the case of identical phonemes than in the case of similar ones, as they are already acquired in the participants' L1, which makes them easier to discriminate. On the other hand, it was expected that reaction time would be the slowest in the case of similar phonemes, because it is difficult to separate two such phonemes. However, the findings were not in line with these expectations. It is somewhat surprising that the fastest reaction time was recorded in the case of discriminating between two similar phonemes, whereas discriminating between two identical phonemes was generally the slowest in most cases in both discrimination levels, which was contrary to expectations and not consistent with the accuracy data. Inconsistently with the accuracy data, in word-level discrimination, discriminating between two similar phonemes was significantly faster than discriminating between two identical phonemes, whereas discriminating between two identical or two new phonemes did not significantly differ. In phoneme-level discrimination, discriminating between two new phonemes was faster than discriminating between two identical ones, which is in line with the accuracy data. Given that non-significant result of overall ANOVA, it must be treated with caution. Notwithstanding that ANOVA showed that discriminating between two similar phonemes was significantly faster than discriminating between two identical or two new phonemes, non-parametric test showed that discriminating between two similar phonemes did not significantly differ from discriminating between two new phonemes, but both are significantly faster than discriminating between two identical phonemes. However, this significant difference in the parametric test results between similar phoneme and new phoneme conditions is marginal (p =.04), which might be the reason for the inconsistency between the parametric and non-parametric results.

On the other hand, unlike accuracy data, orthographic input had no significant effect on the participants' reaction time in both word-level and phoneme-level discrimination. Moreover, correlational analysis showed that reaction time in word-level discrimination did not correlate with the participants' reaction time in phoneme-level discrimination in all three levels of L1/L2 phoneme similarity. This finding was in line with those in the accuracy data.

Looking at the findings of the reaction time measure in the two levels of phoneme discrimination, it is shown that only L1/L2 phoneme similarity level had a significant effect on the time participants took to respond, whereas neither orthography nor phonemic coding ability significantly affected their reaction time. For this reason, the focus of this chapter is purely on accuracy data and there will be few further discussions of reaction time in the following subsections.

5.3 General Discussion

Before discussing the five research questions, it is important to mention that participants were exposed to pairs of target phonemes and were tested in their discrimination of these pairs of phonemes in the Pre-Sound Recognition Phase before they were exposed to orthographic input and discrimination tasks which were included in the main analysis. The Pre-Sound Recognition Phase aimed to measure the participants' a priori ability to discriminate between the two target phonemes in each pair before the learning phase. Their responses in this phase were not included in the main analysis. It included two different tasks. In the first task, they heard two isolated phonemes and their task was to determine whether these two phonemes sounded the same or different. In the second task, they heard two Arabic minimal pairs contrasted in the initial position and their task was to decide whether the two words were different or identical. Both tasks consisted of three phonemes shuffled and presented as pairs: two target phonemes and one more distracting phoneme. For example, the identical-phonemes condition consisted of the target phonemes /θ, ð/ and one distracting phoneme /ʃ/. These three phonemes were presented to the participants randomly in pairs. Findings suggest that the performance of the participants was affected by the L1/L2 phoneme similarity of the pair in both tasks, i.e., whether the two phonemes mapped to one, two, or no L1 phonemes. However, their performance in the first task (phoneme-level discrimination) was different and did not correlate with their performance in the second task (word-level discrimination). From their performance in the first task, it was found that phoneme pairs existing in the participants' L1 (identical phonemes) and novel phonemes (new phonemes) have the same level of perceptual difficulty when uttered in isolation, as can be seen from their similar responses to both pairs. This was not the case when perceiving two L2 phonemes that map to only one L1 category (similar phonemes); in this case, participants found it more difficult to discriminate these than identical or new phonemes. However, when these phonemes occurred in real-word contexts, like in the second task, the performance of the participants differed from their performance when discriminating between phonemes uttered in isolation. Their ability to discriminate contrasts was best when the two phonemes mapped to two L1 categories (identical phonemes), whereas their performance when the two phonemes mapped to only one L1 category (similar phonemes) was not significantly different from the case when the two phonemes did not map to any L1 category (new phonemes).

The difference between the performance in the two tasks with /s/ and /sˤ/ (similar phonemes) was not surprising. This can be attributed to the fact of having pharyngealisation as a phonological cue in the second task. As mentioned in Subsection 2.5, one essential characteristic of Arabic is that emphatic consonants affect the production of vowels preceding or following these consonants, resulting in having different allophones for these vowels from those following their non-emphatic counterparts (Shar & Ingram, 2010). These different allophones might have served as phonological cues in the second task, resulting in having more accurate performance than in the first task which lacked these cues. Walley and Carrell (1983, as cited in Hayes-Harb & Durham, 2016) found that, in order to identify consonants, listeners usually depend on the acoustic information of surrounding vowels. This is supported by Hayes-Harb and Durham (2016), who state that their participants were able to detect the Arabic allophone [ɑ] because it overlaps with the two English vowels /ɑ/ and /æ/. It is important to mention that two pairs out of three in the current study involved the vowel /a/ which has two Arabic allophones, [ɑ] and [æ]. Thus, it might be easier for L1 English participants to detect this due to the existence of the two English vowels /ɑ/ and /æ/. The third pair involved [ʊ]. The reliance on surrounding vowels as a phonological cue for discriminating the pharyngealisation feature of this consonant as used by L1 Arabic speakers was also reported by Jongman et al. (2011) and Alhumaid (2019). Both Jongman (2011) and Alhumaid (2019) found that L1 Arabic speakers rely on information provided by rime more than on information provided by the onset consonant itself.

Moreover, having different performance when discriminating phonemes uttered in isolation as opposed to real word contexts is supported by Escudero (2015), who argues that L2 learners might be able to discriminate lexical representations even if they include perceptually difficult L2 contrasts. They might, however, not be able to distinguish these phonemes when they are uttered in isolation. Escudero built her argument on the results of Weber and Cutler (2004), who found that L1 Dutch learners of L2 English were not able to distinguish between the two perceptually difficult vowels /ɛ/ and /æ/, as they referred to, despite being able to distinguish between words that included these two vowels such as “pencil” and “panda” (as discussed in Subsection 5.4.4).

According to Escudero (2015), the participants' inability to distinguish what she categorised as difficult vowels caused them to have a discontinuity between perceptual and lexical performance, which is the case where learners can distinguish words in the lexicon without the ability to notice the phonological contrasts in these words. For this particular reason, the main analysis of the current study included two levels of phoneme discrimination: word-level discrimination, provided by the Phonological Test Phase; and phoneme-level discrimination, provided by the Post-Sound Recognition Phase (for details about these two phases, see Subsections 3.3.5.3.4 and 3.3.5.3.5).

After looking at the participants' performance in the Pre-Sound Recognition Phase, which preceded the Word Learning Phase, the following subsections discuss the analyses to the Phonological Test Phase and the Post-Sound Recognition Phase which provide answers for the five research questions.

5.3.1 First Research Question

RQ1: Does the Presence of Orthography Help Participants Differentiate between Two Phonemes in the Case of:

- An Unfamiliar Script such as Arabic Script; or
- A Combination of Familiar Script such as Romanised-Arabic Script and Unfamiliar Script such as Arabic Script?

In order to answer both parts of this research question, three groups of participants were examined during their acquisition of six Arabic phonemes (/θ/, /ð/, /s/, /sˤ/, /χ/, and /ʁ/): the Arabic-script group, the Romanised-script group, and the No-script group. The analysis included two levels of phoneme discrimination: phoneme-level discrimination and word-level discrimination.

In phoneme-level discrimination (Post-Sound Recognition Phase), ANOVA showed that participants in all three orthographic input-type groups did not significantly differ in their performance and their reaction time, indicating that orthographic input, familiar and unfamiliar, did not affect their performance either positively or negatively in discriminating isolated phonemes. However, this was not the case when discriminating phonemes that were embedded within words where orthographic input type did have a significant effect in the accuracy measure but not in reaction time measure, as discussed below. Following discussion of *RQ*1 is restricted to only the accuracy measure, as reaction time was not significantly affected by orthographic input type.

As for the first part of *RQ*1 (having an unfamiliar script such as Arabic script), findings for word-level discrimination indicated that participants' ability to discriminate Arabic phonemes was not affected at all by Arabic orthographic input as their performance was no different from those who had no orthographic input. There are two possibilities to explain this null effect. The first possibility is that the Arabic script in the Word Learning Phase was totally ignored because it was totally novel to the participants in terms of its graphemic symbols as well as its directionality. Therefore, participants were not familiar enough with this script. This can be seen in the fact that, throughout the tasks, the Arabic-script group did not significantly differ from the No-script group. This indicates that presenting Arabic script to the participants was equivalent to presenting nothing to them as they were not getting any benefit from it. Notwithstanding that the two groups did not significantly differ, there was a slightly better performance by the No-script group. This leads to the second possibility. It is possible that participants in the Arabic-script group were trying to attend to orthography, but they failed. This might have happened for many reasons. First, it might be because of the reduction in the cognitive load with the No-script group by having fewer elements of information needing to be processed. According to Sweller (2011), having spoken rather than written input results in the visual channel having less work to do. This is because it does not need to process text or to convert it into auditory form. In this case only the auditory channel needs to process spoken text, which in turn decreases the cognitive load and eventually enhances learning. The second reason might be the fact that the Arabic writing system starts from the opposite direction to that of English, which might direct the participants' attention to the wrong direction, although they were guided by an arrow showing the correct direction. In other words, it is possible that participants in the Arabic-script group were trying to attend to orthography but they just could not cope with the directionality. Therefore, they received useless information that they tried to attend to, because they might have been looking at what they thought was the onset of a word, the left side of a word, but that was wrong as Arabic script runs from right to left. This is supported by the findings of Winke et al., (2013) who found that L1 English learners of Arabic spent more time on captions than did learners of Spanish and Russian. Winke et al. (2013) attribute this to the directionality of the script, among other reasons. This is because Arabic starts from the opposite direction to that of English, whereas Spanish and Russian do not. However, due to the short time of the trials, participants of the current study may have failed to find orthographic contrasts in the script because of its direction.

The findings of the current study are in line with the findings of Showalter and Hayes-Harb (2015) in the case of both Arabic script and Romanised script (as discussed below). Showalter and Hayes-Harb (2015) did not find Arabic script to play any role in the performance of their L1 English participants. They attributed their findings to two possible reasons, one of which is the fact that the two target phonemes /k/ and /q/ map to only one English phoneme, namely /k/. This, according to Showalter and Hayes-Harb, increased the perceptual difficulty of the two phonemes.

This attribution may be the reason behind the performance of the participants in the current study in the case of similar phonemes, as both /s/ and /sˤ/ map to only one English phoneme, which is /s/. The other possible reason suggested by Showalter and Hayes-Harb (2015) is the fact that Arabic script is too difficult for and unfamiliar to the participants. They concluded that when the written input is entirely novel for participants, it is difficult for them to benefit from it. Mathieu (2014) agrees with Showalter and Hayes-Harb (2015) as having unfamiliar script, which included Arabic amongst other languages in his study, negatively affected and hindered participants' ability to discriminate contrasting sounds. As mentioned above, Winke et al. (2013) support this, arguing that their findings imply that spending more time on captions, particularly by L2 Arabic learners, among other groups of learners, indicates that Arabic text is hard to process. Bassetti et al. (2022) provide some examples that instructions in novel languages with naïve learners might not help to reduce or increase the orthographic effect on L2 phonology. These include L1 English learners of L2 Arabic in Showalter and Hayes-Harb's (2015), L1 English learners of L2 German in Hayes-Harb et al's (2018 as cited in Bassetti et al., 2022), and L1 English learners of L2 Russian in Showalter's (2020, as cited in Bassetti, et al., 2022). However, it is important to acknowledge that this issue needs further investigation as these examples all included novel and difficult sounds which were represented by unfamiliar or difficult orthographic forms (Bassetti et al., 2022).

According to Bassetti et al. (2020), research has shown that there is a correlation between the amount of exposure to L2 orthographic input and the effect of orthography on L2 phonological development. It is important to mention that the participants in the current study were entirely unfamiliar with Arabic script and they were not learners of Arabic as a second language. This did indeed decrease the possibility of benefiting from the Arabic orthographic input, based on what Bassetti et al. (2020) state about the effect of the amount of exposure to L2 orthographic input. As discussed in Subsection 2.4.4.3, Mathieu (2016) argues that foreign scripts influence L2 phonological acquisition negatively. Mathieu built his argument based on his findings for the negative effect of orthographic input for L1 English learners of L2 Arabic. Based on the findings of previous studies (Showalter, 2012; Almahmoud, 2013), Mathieu attributed his findings for the Arabic script group to the combination of both the foreignness of the script and the perceptual difficulty of the target phonemes.

The findings of the second part of *RQ1* (having a combination of familiar script such as Romanised and unfamiliar script such as Arabic script) for word-level discrimination indicated that participants' ability in the Romanised-script group to discriminate Arabic phonemes was affected negatively by the existence of the two orthographic input types, as their performance was significantly the lowest among the other two groups. This can be attributed to many reasons. The first reason might be due to the unfamiliarity or partial unfamiliarity with the phonological system of Arabic. It is important to emphasise here that this unfamiliarity is not restricted to this group only. However, it might have interacted with other factors, which are discussed in detail below. To demonstrate, in the case of emphatic consonants, as mentioned above and in Subsection 2.5, vowels are affected by the surrounding consonants (Holes, 2004; Shar & Ingram, 2010). Despite the fact that stimulus words containing the vowel /i/ (the vowel least affected by pharyngealisation spread) were avoided in the current study to make it easier for participants to discriminate contrasts, pharyngealisation of /a/ or /u/ did not help all participants to discern the contrasts between the two consonants /s/ and /sˤ/, as these two consonants had the lowest performance among other consonants. Participants might not have been familiar with the fact that the difference which occurs when pronouncing the vowel due to pharyngealisation is not phonemic. This means that participants may have thought that the consonants were the same whereas the vowels are the ones that are different. The fact that some Arabic vowel allophones overlap with two English vowels such as /ɑ/ and /æ/ may be the reason behind this confusion, given that these allophones were presented in the Romanised script with the grapheme <a> which is also sometimes present in both vowels in English. In other words, being unfamiliar with this effect of pharyngealisation might have led participants to think that the consonants (they were asked about) were identical but a difference occurs in the following vowel, resulting in giving wrong answers.

The second reason might be because of having additional input, which was Romanised alongside Arabic scripts. This might have distracted the participants' attention from familiarizing themselves with phonological contrasts. This is evidenced by the finding when comparing the performance of this group with the other two groups in the Criterion Test Phase, which was a part of the learning phase and aimed to test the occurrence of learning regardless of phonological contrasts. The Romanised-script group had the lowest performance in the Criterion Test Phase in comparison with the other two groups with all three pairs (see Appendix R for descriptive statistics for all phases). This may imply that their learning of lexical representations was impaired by the number of inputs, or they were merely poorer learners. This especially can be considered when looking at their scores in the phonemic coding ability. The lowest mean for phonemic coding ability was recorded by the Romanised-script group (M = 9.04, SD = 6.88), compared to the Arabic-script group (M = 11.64, SD = 7.01) and the No-script group (M = 11.12, SD = 5.98). This will be discussed in detail below.

According to Winke et al. (2013), the process of balancing the simultaneous intake of different sources of input such as video, text, image and audio by learners is not clear. Learners themselves are sometimes unable to explain how they process different types of input at the same time. Winke et al. found support for their claim from the findings of a study conducted by Chai and Erlam

(2008, as cited in Winke et al., 2013), that learners are not able to concentrate on three types of input at the same time, which renders unhelpful the information provided to them. According to cognitive load theory (Sweller, 1988 as cited in Sweller, 2011), processing many elements of information at the same time might have led to an increase in extraneous cognitive load which in turn interfered with learning. Extraneous cognitive load is one of two essential sources of instructional cognitive load: intrinsic cognitive load referring to knowledge levels, and extraneous cognitive load referring to how information is presented. In the case of the Romanised-script group, participants were required to process many sources of information, such as audio and visual input including a picture, and two scripts at the same time, which might have restrained their learning. In other words, having additional unnecessary elements of information involves unnecessary working memory load, which is an extraneous cognitive load. Having two different scripts at the same time might act like redundant information which affected the participants' performance by the redundancy effect which results in having a high extraneous cognitive load. This is because the processing capacity of the visual channel is limited and having unnecessary elements of information may lead to overloading it (Sweller, 2011).

The third reason might be the fact that the English graphemic representation of the two sounds /θ/ and /ð/ is <th>, and of /s/ and /sˤ/ it is <s>. The first pair was distinguished in the main instrument of the current study by underlining one of the two, and the second pair was distinguished by adding (`) after the grapheme representing /sˤ/. Having digraphs representing the same target phonemes may have led participants not to learn the lexical representations of non-words. This can be evidenced by the results of the *t*-test conducted to test the possibility of answering by chance as seen in Section 4.4. The Yes-No Identification Tasks with the two phonemes /θ/ and /ð/ in the Criterion Test Phase and the Phonological Test Phase were not significantly different to chance, whereas the Post-Sound Recognition Phase was significantly different. This indicates that it is very possible that the participants did not acquire the meanings of the words. Their performance in the Forced Choice Identification Task in the Criterion Test Phase were significantly different to chance. Their performance in this task was clearly better than that in the Yes-No Identification Task. Despite that there are differences of 1.92 between the possibility of answering by chance and their means in the Forced Choice Identification Task of the Criterion Test Phase, there is still a possibility that they had not learned the lexical representations of non-words.

The Yes-No Identification Task and Forced Choice Identification Task in the similar-phoneme condition in the Criterion Test Phase were significantly better than chance as they had a difference of 3.6 and 4, respectively, between the possibility of answering by chance and their means. The Yes -No Identification Task with the two phonemes /s/ and /sˤ/ in the Phonological Test Phase were also significantly better than chance. This indicates that they learned the lexical representations of non-words as seen in their performance in the Criterion Test Phase, but they were not able to discriminate the contrasts between the two phonemes /s/ and /sˤ/ as seen in their performance in the Phonological Test Phase.

Looking at their performance in the Post-Sound Recognition Phase, which included isolated phonemes only, with no corresponding images, it was found that the participants' ability to discriminate between /θ/ and /ð/ was better than /s/ and /sˤ/, which supports the likelihood that they did not learn the lexical representations with /θ/ and /ð/ but they were able to discriminate the phonological representations of both phonemes. This was not the case with /s/ and /sˤ/ which had the lowest performance in the Post-Sound Recognition Phase and Phonological Test Phase, but not in the Criterion Test Phase, indicating that the participants learned the lexical representations but were not able to discriminate the phonological contrast.

Their learning of the pair /θ/ and /ð/ might have been impaired by having one additional input (Romanised script) or by the fact that they saw digraphs representing the target phonemes for the pair /θ/ and /ð/. However, having additional input affected their performance in the group in general, i.e., in the three levels of L1/L2 phoneme similarity, but not learning lexical representations seems only to have occurred with the phonemes /θ/ and /ð/, which strongly supports it having been due to seeing digraphs. Distinguishing these digraphs by underlining one pair might not have helped the participants to distinguish the two contrasts.

The negative effect of participants' L1 orthography on L2 phonological acquisition is supported by Bassetti (2008) and Bassetti et al. (2020), who claim that one of the factors that might moderate the effect of orthographic input on L2 phonological development is L1 grapheme-phoneme conversion rules. The effect of orthography that is caused by applying L1 rules to recode a writing unit into a sound unit of L2 is referred to as the inter-orthography effect (Bassetti et al., 2020). An example of this effect is seen in Bassetti (2008), who found that Italian learners of Chinese pronounced the spelling of <ui> as /ui/ although it represents /uei/ in Chinese. Bassetti suggests that this happened because this is how Italian speakers pronounce it in their L1. As mentioned above, this is also supported by the finding of Showalter and Hayes-Harb's (2015) study, who used the grapheme <k> to represent the phoneme /k/ and the grapheme <q> to represent the phoneme /q/ in their study. This might have led participants to think that both phonemes have the same pronunciation as the two graphemes <k> and <q> represent the phoneme /k/ in English, even though <q>, however, tends to be associated with /kw/. This resulted in finding that the Roman script affected the participants' performance negatively. This case may also be applicable to the current study for the grapheme <th> representing both /θ/ and /ð/, and <s> representing both /s/ and /sˤ/. Jackson (2016) also tested the orthographic effect on discrimination between the two Arabic phonemes /k/ and /q/ by L1 English speakers. Jackson found that the participants who were presented with novel grapheme to represent the phoneme /q/ performed more accurately than those who were having diacritic dot under the grapheme <k> representing the same phoneme (/q/). Jackson attributed this finding to the high similarity between the grapheme forms (with or without a diacritic dot), especially given the fact that the grapheme <k> (without a diacritic dot) has correspondence in the participants' L1. This is very comparable to the findings of this study in the case of <s> and <s'>.

Another reason that contributes to the possibility of not learning the meanings of lexical items with the pair /θ/ and /ð/ is that participants were exposed to the pair /θ/ and /ð/ in the first session, but the pair /s/, /sˤ/ in the second one and /χ/, /ʁ/ in the last session. Having the pair /θ/ and /ð/ in the first session might have affected the performance of the participants negatively as they were unfamiliar with what would happen next in the study, but in the following two sessions, they were sufficiently trained in the stages of the study, leading them to be more familiar and ready for the testing phase. This may also be one of the reasons why the session containing /χ/ and /ʁ/ prompted the best performance as it was the last session and the participants had enough training to do these types of tasks in that session. This is because the preceding tasks might have had an impact on the following tasks (McGuire, 2010), as they provided more training and exposed participants to the same strategies, leading to better performance in the subsequent sessions than preceding ones.

Apart from this, the lowest mean for phonemic coding ability was recorded by the Romanised-script group (M = 9.04, SD = 6.88), compared to the Arabic-script group (M = 11.64, SD = 7.01) and the No-script group (M = 11.12, SD = 5.98). Despite the fact that phonemic coding ability appears to have had no effect at all on the performance of the participants in all three conditions of L1/L2 phoneme similarity, and it was not significantly different among the three orthographic input-type groups, the lowest performance of the Romanised-script group might be because of their weakness in phonemic coding ability. As mentioned in Subsection 2.3.3, it is argued that learners who have high phonemic coding ability are more capable of easily structuring words into smaller phonetic units (Carroll, 1993; Meara, 2005). In addition, Reynolds (2002) argues that input processing strategies are better developed by L2 learners who have high phonemic coding ability. This increases learners' ability to recognise and integrate new linguistic units. This is important in highlighting the role played by phonemic coding ability in the success of language learning (Reynolds, 2002). Saito (2017) also claims that adult L2 learners who have a high level of phonemic coding ability are found to be more capable of remembering and analysing unfamiliar sounds. In contrast, having low phonemic coding ability leads to difficulties in remembering phonetic forms (Carroll, 1962). Sparks et al. (2011) claim that weak phonemic coding ability and/or phonological awareness leads to having inefficient and slow ability to decode sounds and words. Looking at all the evidence in the literature could lead one to suppose that one possible reason for having the lowest performance by the Romanised-script group is the fact that they were poorer in phonemic coding ability, but the test that was used in this study might not have been sensitive enough to thoroughly assess the participants' phonemic coding ability (this will be discussed in detail in Subsection 5.4.5). It is possible that if the study was carried out using a different phonemic coding ability test, it would have shown that the Romanised-script group actually had the lowest in phonemic coding ability and their performance in the experiment might account for this fact rather than the orthographic input.

Furthermore, as mentioned in subsection 2.4.4, it is found that the role of orthographic input during perception and production of native and non-native contrasts is largely affected by the learners' experience of phoneme-to-grapheme conversion rules (Bassetti et al. 2020; Nimz & Khattab, 2020). Erdener and Burnham (2005) support this, arguing that orthography has a more positive impact on L2 phonology only if learners have prior experience of a phonologically transparent L1 orthographic system, whereas it was found that orthography had a weaker impact with learners who had prior experience of phonologically opaque L1 orthographic system. Nimz and Khattab (2020) support this with the findings of Dornbusch (2012, as cited in Nimz & Khattab, 2020) as Dornbusch's German participants, who had a relatively transparent L1 orthographic background, were affected more by English orthographic input, leading them to make more mistakes, and had slower reaction times than Danish learners of L2 English, whose L1 had a more opaque orthographic system. Nimz and Khattab (2020) argue that the main source of the negative effects caused by L2 spelling is the characteristics of the L1 orthographic system. Looking at the current study, English has a number of phonemes that are represented differently in orthography leading to a number of words that have fewer consistent phoneme-to-grapheme correspondence rules than phonologically transparent languages (Holmes & Ng, 1993). This could be the reason behind the fact that orthographic input (either Arabic or Romanised) was found to be unhelpful in the current study in either phoneme-level or word-level discrimination as in the L1 English participants' experience, orthography was not always an accurate cue to phonology leading participants to have less reliance on orthography for phonology learning. This can be evidenced by revealing that the group which had no orthographic input had the best performance in phoneme discrimination among all three conditions of orthographic input type. However, this needs further investigation with cross-linguistic studies as this outperformance of the No-script group was not significantly different with that in the Arabic-script group. According to Cutler (2015), orthographic input is meant to assist in building lexical representations. However, in some cases, like the case in the current study, it is found that orthography fails to fulfil this purpose, resulting in having more difficulty in recognizing words or phonemes perceptually and, therefore, orthography has more disadvantages than benefits. Escudero (2015) also supports this claim; she argues that orthography may not help perception accuracy and it is better to avoid it in some cases because, according to Escudero, the positive role of orthography occurs only when there is a congruency between L1 and L2 grapheme-phoneme correspondences.

Apart from this, given that the effect of the orthographic input type in the accuracy data in the word-level discrimination was large (.22) with a very high power (.98), this means that there has been a sufficient sample size to detect that significant effect. However, the effect size of the orthographic input type in the phoneme-level discrimination was smaller (.07) with a medium power (.57). In this case, orthographic input type was found to have no effect, which might be caused by the fact that the sample size was not sufficient. Nevertheless, it is evidenced in the literature that the effect of orthographic input was found to have different roles in word recognition tasks from that in phoneme discrimination tasks (Escudero & Wanrooij, 2010; Escudero et al. 2014; Escudero, 2015; Han & Oh, 2018). This might account for the different effect size of orthographic input in the case of phoneme-level discrimination and word-level discrimination.

5.3.2 Second Research Question

RQ2: Does L1/L2 phoneme similarity affect the difficulty of participants' ability to differentiate between two phonemes?

In order to answer this research question, the experiment ran in three different L1/L2 phoneme similarity conditions carried out in three different sessions. The first session was dedicated to two

Arabic phonemes which are, based on their articulatory and acoustic features, identical to two English phonemes and, therefore, have two counterparts in English (/ð/ and /θ/). These two phonemes are referred to throughout the study as identical phonemes. The second session was dedicated to another two Arabic phonemes which are similar to one English phoneme and thus have only one counterpart in English (/s/ and /sˤ/). These two Arabic phonemes are referred to throughout the study as similar phonemes. The last session was dedicated to two Arabic phonemes which are novel phonemes that have no counterparts in English (/ʁ/ and /χ/). These two Arabic phonemes are referred to throughout the study as new phonemes.

Looking at the accuracy measure, findings showed that the performance of the participants was affected by whether the pair of phonemes consisted of identical (mapping to two L1 categories), similar (mapping to one L1 category), or new phonemes (not mapping to any L1 category) in both levels of phoneme discrimination. However, the extent of this effect was not the same in these levels of phoneme discrimination. In word-level discrimination, participants performed best when they were asked to discriminate between the two new phonemes, where the two L2 categories do not map to any L1 category. Despite the fact that the performance of the participants did not differ significantly when discriminating between two identical phonemes from when they were discriminating between two similar phonemes in the word-level discrimination, the means of both conditions showed that the similar pair was slightly more difficult to discriminate than the identical pair. This is also in line with the findings for the second discrimination level, phoneme-level discrimination. Their performance in discriminating between isolated phonemes was significantly the lowest when discriminating between two similar phonemes, whereas for identical and new phonemes it did not significantly differ.

This was predicted, as it is argued that the ease and difficulty in discrimination of non-native segmental contrasts might be directly influenced by listeners' L1 phonology (Best & Taylor, 2007). In other words, the existence or absence of sounds in L1 and L2 can predict L2 speech perception and production. A number of L2 speech perception models, such as the SLM-r (Flege & Bohn, 2021), PAM (Best, 1995), PAM-L2 (Best & Taylor, 2007), and many others, attribute the difficulties in sound perception to the similarity between L1 and L2 sounds rather than the differences between them. This means that similar sounds tend to be more difficult to discriminate than different ones, especially when two similar phonemes in L2 map to only one L1 phoneme (Oller & Ziahosseiny, 1970, as cited in Chan, 2012; Cutler, 2015). Schmidt (2007) also supports this, stating that the interpretation and perception of speech sounds are done depending on L1 phonemic categories and requirements, which is the reason why some items are perceived differently by listeners who come from different L1 backgrounds (Lisker & Abramson, 1964 as cited in Schmidt, 2007).

Regarding the performance on the identical pair (/ð/ and /θ/) in word-level discrimination, as discussed in the previous subsection, having the pair /θ/ and /ð/ in the first session might have had a negative effect on the performance of the participants due to their unfamiliarity with what would happen next in the experiment and being unprepared for the testing phase. Having lower performance in the Criterion Test Phase only indicates that the problem occurred in their word-picture matching, i.e., their learning the meanings of lexical items, but not their ability to discriminate two phonemes.

This was not the case with the following two conditions of L1/L2 phoneme similarity. Participants in the Romanised-script group performed no better than chance in the Yes-No Identification Task in the identical-phoneme condition in the Criterion Test Phase. This clearly affected the total achievement of the Romanised-script group participants in the identical-phoneme condition. However, by looking at the participants' performance in the Pre-Sound Recognition Phase in all three orthographic input-type groups, which preceded the Word Learning Phase, it is shown that their discrimination between two identical phonemes was the same as their discrimination between two new phonemes, which was better than discriminating between two similar phonemes. This was not surprising as it is hypothesised by PAM (Best, 1995) that the perception and discrimination of two L2 categories that are mapped to two L2 categories will be very good or excellent, because the two L2 sounds are assimilated to two different native categories and are perceived as two acceptable exemplars of two different native sounds (two-category 'TC') (Best 1995).

Depending on the means of the three conditions of L1/L2 phoneme similarity, the similar-phonemes condition had the lowest performance in the Pre-Sound Recognition Phase, the Phonological Test Phase, and the Post-Sound Recognition Phase, but not the Criterion Test Phase (see Appendix R for descriptive statistics). This indicates that participants had learned the lexical representations, as shown in their performance in the Criterion Test Phase, but their discrimination of phonological contrasts was the lowest, as shown in their performance in the Pre-Sound Recognition, the Phonological Test, and the Post-Sound Recognition Phases.

Having the lowest performance on the similar pair is in line with an argument presented by Escudero et al. (2008). They argue that the difficulties in distinguishing L2 similar-sounding words and the inability to build these words phonologically could be as a result of the lack of lexical representations of these words, which result in dealing with these words as homophones. It is argued that when there are two contrasting phonemes in L2 that map to only one L1 category phoneme in L1, it will be more difficult to distinguish these contrasts (Best, 1995; Best & Taylor, 2007; Cutler, 2015; Flege, 1995; Nimz & Khattab, 2020). Due to the absence of these contrasts in the learners' L1, learners are likely to be unable to discriminate these phonemic contrasts either perceptively or productively (Best, 1995; Best & Taylor, 2007; Escudero, 2015; Flege, 1995). This inability to discriminate contrasts leads to the misperception of minimal pairs that include these contrasts (Nimz & Khattab, 2020). This may be illustrated by the case in the current study with the phonemes /s/ and /sˤ/ in the Phonological Test Phase and in Showalter and Hayes-Harb's (2015) study, whose participants failed to discriminate the contrast between the two Arabic phonemes /k/ and /q/. Both phonemes in the two pairs (/s/ and /sˤ/) and (/k/ and /q/) map to only one English category, namely /s/ in the former and /k/ in the latter.

A number of speech perception models, such as the SLM-r (Flege & Bohn, 2021), PAM and PAM-L2 (Best 1995; Best & Taylor; 2007 respectively) and L2LP (van Leussen & Escudero, 2015), support these findings. The SLM-r (Flege & Bohn, 2021) states that if two L2 sounds map to one phonetic category in learners' L1 phonology, learners might not be able to discern the phonetic difference between L1 and L2 sounds or between two L2 sounds. According to the fifth hypothesis of the SLM, the category formation of L2 sounds might be blocked by the mechanism of equivalence classification. In other words, if the listeners fail to discern a mismatch in the phonetic feature that signals a contrast between L1 and L2 sounds, the formation of an L2 phonetic category might be blocked, leading to have difficulty in perceiving L2 sounds that have features which do not signal phonological contrasts in L1. This leads to having one single phonetic category for both linked L1 and L2 sounds, which are named diaphones. As mentioned in the previous subsection, having the lowest performance in the similar-phonemes condition might be attributed to this as participants were not able to discern the pharyngealisation feature, which does not exist in English, and this eventually led to the inability to discriminate the contrast between /s/ and /sˤ/ as both sounds map to only one English phonetic category, which is /s/.

PAM (Best, 1995) states there are three possibilities for the perception of non-native phones by L2 listeners: categorise it as either a good or a poor exemplar of a native language phonological segment; uncategorise it and see it as different from any single native phoneme, or; hear it as a non-linguistic nonspeech sound which means it is unassimilated. According to PAM (Best, 1995) and PAM-L2 (Best & Taylor, 2007), due to the articulatory similarities between L1 and L2 sounds, non-native listeners tend to perceptually assimilate an unfamiliar L2 sound to its closest similar sound in their L1.

The concept of the L2LP model (van Leussen & Escudero, 2015) agrees with SLM-r and PAM, in that it claims that the role of L1 perceptual behaviour in the initial stages of L2 acquisition is important. It states that, due to the acoustic similarities shared by L1 and L2 sounds, learners tend to rely on their L1 perceptual behaviour in that stage. This behaviour affects how non-native sounds are perceived, which also agrees with Flege's SLM and Best's PAM (Escudero, 2005). In other words, the acoustic similarities between L1 and L2 sounds, especially those that correspond to only one L1 category rather than those that correspond to two different categories, can predict the difficulty of L2 phonological development.

A number of studies provide examples of how this difficulty occurs when two L2 sounds are perceived as one single L1 sound. For instance, Escudero and Wanrooij (2010) and Escudero and Williams (2012) found that their L1 Spanish learners of L2 Dutch faced difficulties in distinguishing Dutch vowel contrasts that map to only one category in Spanish. In Escudero and Williams' (2012) study, it is shown that the Dutch contrasts /a/ vs /ɑ/ and /y/ vs /ʏ/ are categorised as one single Spanish vowel for each pair: /a/ for the former and /u/ for the latter. Escudero et al. (2013) found that in some cases L1 Dutch learners of L2 English can easily discriminate the contrasts in other minimal pairs, either familiar or novel ones, and this is because these phonemes contrasting minimal pairs are already distinguished in their L1. In the current study, participants' lowest performance was reported in the similar-phonemes condition. Despite having recorded the lowest scores in this condition, they were not unable to discern the contrast as their performance was better than chance. This is hypothesised in PAM's third pattern (the category goodness 'CG'). According to the category goodness pattern, two L2 sounds are assimilated to a single native category in which they are perceived differently in terms of goodness of fit to the native-language phoneme (one is a better exemplar than the other). The perception and discrimination of contrasts in this case is predicted to be intermediate (Best 1995).

As for the new-phonemes condition, it is suggested that the perception of new sounds tends to be easier because it is easier for the learner to notice the difference between L1 sounds and new L2 sounds (Flege, 1995).

Chan (2012) agrees with Flege's argument, stating that it is predictable that L2 learners will judge L2 phonemes depending on the realization of L1 categories. If learners can notice the difference between an L2 sound and its closest L1 sound, they will be able to build a new L2 phonetic category. However, learners will struggle to detect two L2 contrasts if they are distinguished in L2 but not in L1, as in the case of /s/ and /sˤ/. Flege (1995) believes that L2 learners recognise the L2 phonological system by either adding new phonetic categories or modifying existing phonetic categories. The findings of the current study support all these claims as it was found that the best performance was reported in the new-phonemes condition in word-level discrimination, and phoneme-level discrimination was significantly better than in the similar-phoneme condition, though it did not differ from the identical-phoneme condition.

According to the SLM-r (Flege & Bohn, 2021), the greater the distance between an L2 sound and its closest L1 sound is, the more possible it is that a separate phonetic category of L2 sound will be established. PAM (Best, 1995) suggests that if two L2 sounds do not match any native-language phoneme, as in the case of (/ʁ/ and /χ/), the perception of these two sounds might be one of two cases. The first case is when one L2 phoneme is assimilated to a native phoneme, whereas the other is perceived as an uncategorised speech sound. In this case, the phonological distinction is clearly reflected, which predicts having very good perception and discrimination of these contrasts. The second case is when both L2 sounds are perceived as uncategorised. The perception and discrimination of these contrasts in this case are predicted to fluctuate between poor and moderately well. PAM-L2 (Best & Taylor, 2007), however, extended the PAM framework by proposing four possible cases for contrasts in L2. The fourth case is in line with the findings of the current study in the new-phoneme condition. Best and Taylor (2007) argue that in the fourth case, L2 learners do not perceptually assimilate L2 sounds to any L1 sounds. Instead, they have a mixture of similarities to several L1 phonological categories for both L2 sounds. In this case, it is predicted that learners will be able to easily perceive or discriminate one or both sounds, which supports the concept of the SLM-r's new sounds. Major (2001) agrees with that, claiming that if learners categorise a particular L2 sound as more similar to an L1 sound, such as /sˤ/, than another L2 sound, such as /ʁ/ or /χ/, the acquisition of the first one will be slower than that of the second one.

It is worth highlighting here that the participants in the current study were not active learners of Arabic. Best and Taylor (2007) differentiate between L2 learners and naïve listeners, in that L2 learners are going through a process of learning and have goals to achieve that might be functional or communicative and are not restricted to educational requirements. Naïve listeners, on the other hand, are not actively learning the language; they are merely functional monolinguals who are not using L2 and are linguistically naïve to this language. According to Best and Taylor (2007), naïve listeners have difficulties in categorising and discriminating some phonetic contrasts which have no counterparts in their L1 due to their lack of non-L1 experience. This implies that the performance of the participants might have been affected by the fact that they were not learning Arabic and findings might be different if the current study was carried out with L2 Arabic learners.

Regarding the reaction time measure, as mentioned above, it was found that the fastest reaction time was recorded in the similar-phoneme condition. Having the fastest reaction time in similar phonemes might be attributed to the fact that there could be some sort of speed/ accuracy trade-off. Participants were not able to discriminate between contrasts in the similar-phoneme condition and did not realise that the two phonemes were different, leading them to respond faster than usual and give fewer correct answers, and therefore lower values for reaction time because only the reaction time of correct responses was included in the analysis. This can be seen in that the lowest performance in the accuracy measure among the three conditions of L1/L2 phoneme similarity was that in the similar-phoneme condition.

According to Pisoni and Tash (1974), responses take less time when discriminating between phonemes in the case of having matched phonemes, i.e., phonemes that belong to the same phonetic category (e.g., /ba/ and /ba/) than in the case of having mismatched phonemes, i.e., phonemes that belong to different phonetic categories (e.g., /ba/ and /pa/). Looking at the findings of the current study, the participants took less time to respond when the pair of phonemes consisted of two phonemes with one existing in the participants' L1 and another similar to it but not existing in their L1 (similar-phoneme condition), than when they were presented with two sounds that exist in their L1 (identical-phoneme condition). Despite the fact that analysis showed that the only significant difference was between the similar- and identical-phonemes conditions, and the differences between similar-phoneme condition and new-phoneme condition and also new-phoneme condition and identical-phoneme condition were not significant, the means suggested the similar-phonemes condition was the fastest followed by new-phonemes condition with the identical-phonemes condition being the slowest. Given that the two phonemes in the similar-phoneme condition map to only one phonetic category in the participants' L1, reaction time might have been affected because the participants simply could not tell the difference between them and were confident that the two sounds were the same, which resulted in a quicker reaction time. However, despite the fact that reaction time was reported only in the case of correct responses, it seems that the participants might have thought that all trials sounded the same, which is why they did not need time to think and took less time to respond. This eventually resulted in having faster reaction times in correct trials than for new and identical phonemes. This is specifically applicable in the case of phoneme-level discrimination because the two phonemes were uttered in isolation without any following or preceding vowels. As previously mentioned, studies have shown that listeners, either L2 Arabic learners (Binasfour, 2018; Hayes-Harb & Durham, 2016) or L1 Arabic speakers (Alhumaid, 2019; Jongman et al., 2011), tend to rely on surrounding vowels to discriminate Arabic emphatic consonants such as /sˤ/ from their non-emphatic counterparts such as /s/ because of the pharyngealisation spread from the emphatic consonant to the following vowel, providing an acoustic cue for perceivers. This means that hearing emphatic sounds in isolation makes them harder to discriminate from their non-emphatic counterparts. However, the unfamiliarity of the participants with the phonological system of Arabic might have resulted in this being applicable in word-level discrimination as well, because they did not know that the effect of pharyngealisation spread affects the consonant per se and might think that this effect occurs on the vowel instead. This is discussed in detail in subsection 5.4.1. However, because of some slight differences which were shown by non-parametric results, these findings may be worn for further investigations. As mentioned in the previous chapter, Reaction time data in the parametric tests were inconsistent with those in the non-parametric tests. Despite that these findings are slightly different and occurred only on one discrimination level (phoneme-level discrimination), looking at reaction time is more conservative to conclude that there were no significant reliable effects involving reaction time and it seemed to make no contribution to the general understanding of the participants' behaviours in the current analysis.

5.3.3 Third Research Question

RQ3: Is the effect of orthographic availability influenced by L1/L2 phoneme similarity?

Before discussing the findings of this research question, it is important to highilight that this section discusses the findings of the accuracy measure only as reaction time was found to not being affected by orthographic input type and there was no interaction between L1/L2 phoneme similarity level and orthographic input type. In order to answer this research question, the study was designed to include an independent groups factor (orthographic input type) which had three levels (Arabic script, Romanised with Arabic scripts, and no script) and a repeated measures factor (L1/L2 phoneme similarity) which also had three levels (identical phonemes, similar phonemes, and new phonemes). The performance of three orthographic input-type groups was analysed in all three conditions of L1/L2 phoneme similarity. As shown in the previous two subsections, L1/L2 phoneme similarity influenced the participants' ability to discriminate different types of phonemes in both word-level discrimination and phoneme-level discrimination. Orthographic input type, on the other hand, only influenced the participants' performance in word-level discrimination, not in phoneme-level discrimination. Notwithstanding that the results of ANOVA indicated that, in word-level discrimination, L1/L2 phoneme similarity and orthographic input type did not have a significant interaction, non-parametric tests suggested that there were interesting differences between the orthographic input type groups and L1/L2 phoneme similarity levels. Given that the assumptions of distribution of normality and equality of variance were not met, these findings were presented in the previous chapter and are discussed in this chapter. However, this may be worn for other investigations. The non-parametric tests showed that the effect of L1/L2 phoneme similarity was significantly different in some orthographic input-type groups, namely the Arabic-script group and the No-script group, and not significant in the other one, the Romanised-script group. On the other hand, orthographic input type was significantly different in the identical-phoneme and new-phoneme conditions, but not in the similar-phoneme condition. Therefore, this section discusses the findings for word-level discrimination only, as orthographic input type had no significant effect on phoneme-level discrimination. It is worth highlighting here that these findings clearly need to be verified with further research due to the inconsistency between parametric and non-parametric results. By way of demonstration, having bigger sample might solve the problem of violating the assumptions of parametric tests which makes analysis more robust.

Mack (1988 as cited in Flege, 1995) argues that learning L2 speech involves more analytic ability than L1 acquisition. These requirements, according to Mack (1988, as cited in Flege, 1995), increases in the case when exposing L2 learners to written input in early stages of learning. Wayland (2007) highlights the difference between adults and children, in that adults are language-specific perceivers. This means that they largely depend on their L1 phonological filtering when perceiving speech sounds, leading them to consider the perceptual relationships between L1 and

L2 speech sounds as the base for the perceptual discriminability of L2 speech sounds. Schmidt (2007) supports this by providing a definition for the study of L2 speech perception. The study of L2 speech perception, according to Schmidt, refers to examining the process of perceiving acoustic or gestural information, where this examination is based on a pre-existing sound's individual systemization, which is the listener's L1 phonology. This implies that this process, at least at the beginning, is derived from the perception of L1 speech. Speech perception models such as the SLM-r (Flege & Bohn 2021), PAM (Best, 1995) and PAM-L2 (Best & Taylor, 2007) lack a connection between L2 sounds perception and orthographic input (Bassetti, 2017; Nimz & Khattab, 2020). Bassetti's (2017) findings encouraged her to suggest including the effect of orthographic input during L2 phonological development in the current dominant speech perception models, such as the SLM (Flege, 1995) and PAM-L2 (Best & Taylor, 2007). However, PAM-L2 (Best & Taylor, 2007) implies that L1 English learners of L2 French tend to assimilate the voiceless uvular fricative /ʁ/ to the English liquid /r/. The French uvular fricative /ʁ/ and the English liquid /r/ are phonetically very distinct. Nimz and Khattab (2020) attribute the assimilation of these phonemes by L1 English learners of French to the graphemic symbol in the two orthographic systems of French and English, namely <r>. This is not possible if the model is only based on the perceptual similarity between the two phonemes because they are phonetically distinct. Flege (2016), on the other hand, claims that orthography raises problematic issues that hinder speech perception when these issues are not related to speech learning. Flege (2016) provides examples of the negative effect of orthographic input due to the incongruency between L1 and L2 orthographic systems, and/or orthographic depth, as explained in Subsection 2.4. Therefore, Flege (2016) believes that the phonetic input that is needed to establish a phonetic category cannot be accessed via orthographic input.

As shown in Subsection 2.4, the effect of orthography on L2 phonological development is not always the same and there are several factors that may moderate this effect. Escudero (2015) claims that the positive effect of orthography in the case of unperceivable contrasts is restricted to the case where there is a congruency between L1 and L2 grapheme-phoneme correspondences; this is because orthographic representation has already been learned in the learners' L1 and, therefore, orthography acts as redundant cue. On the other hand, orthography does not provide helpful information, even with easily perceived contrasts, if there was no congruency between L1 and L2 grapheme-phoneme correspondences. However, Escudero's (2015) findings revealed that orthographic input did not help the participants in her study to discriminate perceptually easy contrasts, which was attributed to the fact that orthography does not help when native graphemes represent different phonemes from those in their L1, i.e., they are incongruent. Although L1/L2 phoneme similarity and orthographic input type did not have a significant interaction in the current study, as shown by the results of ANOVA analysis, Escudero's attribution might be applicable to the findings of the current study in the case of the Romanised-script group when they were asked to discriminate between two similar phonemes. In this particular condition, L1/L2 phoneme similarity did not have a significant effect in the Romanised-script group, and orthographic input type did not have a significant effect in the similar-phoneme condition, resulting in this condition having the lowest performance in comparison with other conditions. As shown above, /s/ and /sˤ/ are the most difficult pair to distinguish, as predicted by the SLM-r (Flege & Bohn, 2021) and PAM (Best, 1995). As previously mentioned, in the current study, the phonemes /s/ and /sˤ/ were both represented by the grapheme <s> and distinguished with (') to represent /sˤ/. Having orthographic information that does not mark contrast in the participants' L1, in this case an apostrophe ('), may have resulted in participants not benefitting from orthographic input. In contrast, participants in this condition had the lowest performance among all the other conditions in this study. This can be attributed to many reasons as mentioned in the discussion of *RQ*1 and *RQ*2, including having one more input resource than the other two conditions of orthographic input type, being unfamiliar with the L2 phonological system, having perceptual difficulty with the two target phonemes, and/or lacking orthographic cues as both were represented with <s> and distinguished with (’), which does not mark contrast in their L1.

Because the interaction was not significant, as shown by ANOVA, and non-parametric tests showed that the effect of L1/L2 phoneme similarity was significantly different in some orthographic input-type groups but not in others, answering this research question is not straightforward. Therefore, any conclusions drawn should be treated with caution and would need to be verified by further research. The findings of non-parametric tests in the current study can be summarised as follows: when the participants were asked to discriminate phonemes in word-level discrimination, orthographic input type only affected the participants’ performance when they were asked to discriminate between two new phonemes or two identical phonemes, but not two similar phonemes. This effect of orthographic input type did not play any role when it was totally novel to the participants (Arabic), whereas it negatively affected their performance when it was provided in a novel script accompanied by a familiar one (Arabic with Roman script). In the case of having a novel script or having no script at all, their ability to discriminate new phonemes was better than their ability to discriminate identical or similar phonemes. However, this was not the case when participants had a familiar script along with a novel one, as their performance did not significantly differ in all three levels of L1/L2 phoneme similarity. In fact, they were the lowest among the other two orthographic input-type groups in discriminating between identical and new phonemes. The participants’ discrimination differed slightly when they were asked to discriminate between phonemes in phoneme-level discrimination. In this case, L1/L2 phoneme similarity had a significant effect in all three orthographic input-type groups. Furthermore, participants had identical performance when asked to discriminate between new phonemes or identical phonemes in all three orthographic input-type groups. Their discrimination of identical and new phonemes was better than that of similar phonemes. Orthographic input type, however, did not affect their performance, in neither the case of an unfamiliar script alone nor familiar and unfamiliar scripts together.

Mathieu (2016) tried to attribute his findings for the negative effect of novel orthographic input (Arabic script) on his L1 English participants' performance by looking at the effect of the perceptual difficulty of the phonemes and/or orthographic input. By linking his findings to those of Showalter (2012), who found that Arabic script did not play any significant role in participants' performance, Mathieu (2016) excluded novel orthographic input being the sole reason, given that he used a similar experimental design to Showalter's and a similar research context to that of Showalter. On the other hand, referring to Almahmoud (2013), who claims that the contrast between both pairs /χ – ħ/ and /k – q/ shares the same perceptual discriminability as it was found that there was no significant difference between them, Mathieu (2016) suggests that the negative effect of a novel script in his study cannot be attributed to the perceptual difficulty of contrast alone. Therefore, Mathieu believes that two factors (perceptual difficulty of phonemes and novel orthographic input) interacted together and led to a negative effect. However, it is worth noting here that Showalter (2012), who did not find any role for Arabic script, used two phonemes which map to only one phonemic category in the participants' L1, whereas Mathieu (2016), who found a negative effect for Arabic script, used two phonemes which did not map to any phonemic category in the participants' L1, i.e., new phonemes. In addition, the graphemic symbols of the two phonemes in Showalter's study were not similar in Arabic < قـ and كـ >, representing /q/ and /k/

respectively, and their participants were guided as to the direction of Arabic script (from right to left), whereas the graphemic symbols of the two phonemes in Mathieu’s study were very similar in Arabic and merely distinguished by a dot < ـحـ and ـخـ >, representing /ħ/ and /χ/ respectively, and their participants were not guided as to the direction of Arabic script. This might have affected the performance of Mathieu’s participants, resulting in having a negative effect of Arabic script. The finding of the current study, in the case of similar phonemes, can be compared to Showalter’s findings as both phonemes /sˤ - s/ map to only one phonetic category in English: /s/. Furthermore, the graphemic symbols of these phonemes in Arabic should be distinctive because their form is different, < ـصـ and ـسـ > representing /sˤ/ and /s/ respectively, and participants were guided by an arrow as to the direction of Arabic script. The findings of the current study also agree with Showalter’s (2012) findings, in that Arabic script had no significant effect at all. Therefore, the findings cannot be attributed to the interaction between novel orthographic input and perceptual difficulty (caused by L1/L2 phoneme similarity). Thus, apparently, in the case of novel script, only perceptual difficulty affected the participants’ performance.

5.3.4 Fourth Research Question

*RQ*4: Does the participants’ ability to discriminate phonemes when embedded within words (word-level discrimination) correlate with their ability to discriminate phonemes when uttered in isolation (phoneme-level discrimination)?

To answer this question, two levels of phoneme discrimination were included in the analysis of the current study, word-level discrimination (phonemes embedded in words) and phoneme-level discrimination (isolated phonemes). The findings of the current study revealed that both accuracy and reaction time data in the word-level did not correlate with any condition in the phoneme-level.

However, these two phoneme discrimination levels were not alike in terms of the orthography effect. As mentioned above, the findings of the current study show that orthographic input type affected the performance of the participants only in word-level discrimination but not in phoneme-level discrimination. It was found that the learning of participants who were presented with a familiar script accompanied by a novel script was hindered in some cases, and generally they had the lowest performance among the orthographic input-type groups, whereas none of the three orthographic input-type groups were significantly different in their phoneme-level discrimination.

Phonological development is somehow related to lexical development in L2 learning because having difficulties in recognizing words, or parts of words, accurately is sometimes caused by having contrasts that distinguish these words which are difficult for learners to discriminate in perception tasks (Escudero, 2015). In contrast, if the words, either novel or familiar, are distinguished by contrasts which learners can readily discriminate, then distinguishing these words will not necessarily be difficult, as evidenced by the findings of Broersma (2005) and Escudero et al. (2013). As shown in Broersma (2005), which was conducted on both L1 Dutch learners of L2 English and L1 English naïve listeners of Dutch, and Escudero et al. (2013), which was conducted on L1 Spanish learners of L2 Dutch, there was a continuity between phoneme discrimination and the word recognition as participants in both studies were unable to discriminate minimal pairs that were contrasted by perceptual difficult phonemes but, in contrast, they were able to successfully discriminate minimal pairs (novel or familiar) that were contrasted by easy perceptual phonemes. However, studies have also shown that, in some cases, L2 learners can distinguish minimal pairs that are contrasted by perceptually difficult phonemes even though they are not able to discriminate these contrasting phonemes (e.g., Weber & Cutler, 2004). Escudero et al. (2008) attribute this to the fact that it can happen because of another source of information, which is orthographic input.

In other cases, however, contrasts which are revealed by orthographic input might cause more difficulty in word recognition than in phoneme-level discrimination. By way of illustration, the participants of Escudero et al. (2014), who were both learners and naïve listeners of Dutch, failed to discriminate vowel duration contrast due to the availability of orthography at the word level more than when these contrasts occurred at the phoneme level. Therefore, Escudero (2015) concluded that having orthographic input during a word recognition task might lead to a discontinuity between word recognition and phoneme perception as learners are able to discriminate words but not contrasting phonemes of these words.

As for correlation, it was found that there were no correlations in the three conditions of L1'/L2 phoneme similarity levels between phoneme-level and word-level discrimination in terms of both accuracy and reaction time data. Generally, it is important to bear in mind that the Yes-No Identification Task used in the word-level discrimination was found to require more analytical answers. This is because it does not provide direct comparison which is more disadvantage in the case of the perceptual difficult contrasts. On the other hand, in the Same-Different Task, which was used in the phoneme-level discrimination, participants' task was only to decide whether the two sounds are the same or different without the need for identifying what the difference was, if there was a difference (McGuire, 2010). Apart from this, one possible reason for the lack of correlation between the two discrimination levels is the participants' working memory. According to Sweller (2011), working memory has less processing capacity than its storage capacity as it is able to process no more than three to four items of information at the same time. Sweller (2011) defines the processing here as the ability to combine, contrast, or deal with multiple elements. Looking at the participants' task in both level of discriminations, it is obvious that their memory load was increased in the word-level discrimination as they were required to use more processing when doing word-phoneme discrimination than when doing phoneme-level discrimination. This is because, in the former, they were asked to relate the presented picture to the auditory form they heard, whereas in the latter, they were just asked to decide whether the two phonemes sounded the same or differently without any visual input presented. As mentioned in Subsection 5.3.1, having spoken input only results in the visual channel having less work to do. In this case only the auditory channel needs to process spoken text, which in turn decreases the cognitive load and eventually enhances participants to process information (Sweller, 2011). There are other possible reasons that might be applicable in some conditions but not in others. These are discussed below.

Looking at the identical-phoneme condition, performance in one condition, the Romanised-script group was at the chance level in both the Criterion Test Phase and the Phonological Test Phase. This indicates that the participants in this condition did not learn the meanings of the non-words included in this condition of L1/L2 phoneme similarity. Therefore, this affected their performance in word-level discrimination (Phonological Test Phase) negatively but did not necessarily indicate that they were not able to discriminate the target phonemes, especially as these two phonemes do exist in the participants' L1 (identical phonemes). In other words, they were essentially just guessing in the word-level discrimination, which can be the reason for the lack of the association with the performance in the phoneme-level discrimination. This led to participants having lower performance in word-level discrimination and higher performance in phoneme-level discrimination and there being no correlation between these two cases.

It is worth highlighting here that the discrimination between two new phonemes in the Phonological Test Phase, was significantly better than the discrimination between two identical phonemes in both the Arabic and No-script groups. This condition did not correlate. However, as mentioned above, word-level discrimination did not correlate with phoneme-level discrimination in the case of discriminating between two similar phonemes. The reason behind that might relate to the perceptual difficulty of the contrast between these two phonemes. Participants basically failed to discriminate this contrast, as shown in their performance in the Phonological Test Phase where their answers were slightly above chance in many conditions, including the Arabic-script group and the Romanised-script group. Another possible reason might be the fact that their performance was negatively affected by the pharyngealisation or emphasis effect. This is evidenced by their performance in the Pre-Sound Recognition Phase which preceded the Learning Phase. As mentioned in Section 5.3, their Performance when these phonemes occurred in real-word contexts, like in the second task of the Pre-Sound Recognition Phase, differed from their performance when discriminating between phonemes uttered in isolation. The pharyngealization spread as a phonological cue might have had a negative effect on their performance in the word-level discrimination due to their unfamiliarity of Arabic phonological system. This might have led them to think that the difference occurs on vowels rather than consonants resulting in giving wrong answers.

5.3.5 Fifth Research Question

*RQ*5: Is the effect of orthographic availability influenced by the combination of phonemic coding ability and L1/L2 phoneme similarity?

To answer this question, all the participants of this study took an aptitude test, to establish their phonemic coding ability level, which was Llama-E (Meara, 2005; Meara & Rogers, 2019). This test took place before the main experiment of the study. It consisted of two short phases: learning and testing. In the learning phase, which lasted for two minutes, participants were provided with 24 recorded syllables along with their unfamiliar written representations. Participants spent two minutes familiarizing themselves with all the syllables by clicking on each one to hear its auditory form. The participants moved immediately to the testing phase which consisted of 20 trials. This phase tested the participants by presenting a bisyllabic word orally accompanied by 20 written words to choose from. The participants' task was to guess the spelling of the word they heard depending on how its combining syllables were pronounced. Their score was calculated by counting the number of correct answers out of 20.

Before moving to the main analysis of this research question, which is multiple linear regression, a correlational analysis between the phonemic coding ability scores, L1/L2 phoneme similarity, and orthographic input type with the Phonological Test and Post-Sound Recognition scores was carried out. As no correlation was found between the phonemic coding ability scores and the accuracy data of the Post-Sound Recognition scores, as well as reaction time data in both the Phonological Test and Post-Sound Recognition Phases, multiple linear regression was only carried out in the Phonological Test Phase, on the accuracy data. Having different results in the correlations between phonemic coding ability scores and scores in Post-Sound Recognition and Phonological Test Phases is not surprising. As mentioned in Subsection 5.3.4, contrasts which are revealed by orthographic input might cause more difficulty in word recognition than in phoneme-level discrimination (Escudero et al., 2014). Having no correlation between phonemic coding ability and these measures indicates the poor effect of phonemic coding ability in their scores of these measures. Therefore, the following discussion is restricted to the accuracy data in the word-level discrimination.

Based on the literature which shows that phonemic coding ability plays a great role in helping learners to structure words into smaller units (Carroll, 1993; Meara, 2005), retain and analyse novel sound patterns (Saito, 2017), and develop input processing strategies (Reynolds, 2002), it was expected that participants with higher scores in the Llama-E test would have higher performance in the phoneme discrimination tasks which followed. In contrast, those who obtained lower Llama-E scores were expected to have lower scores in the phoneme discrimination tasks which followed, because it has been found that learners who are weak in their phonemic coding ability face difficulties in remembering phonetic forms and mimicking speech sounds (Carroll, 1962), and they are unable to do rapid and efficient sound and word decoding (Sparks et al., 2011). However, the findings were not in line with expectations, as phonemic coding ability was not related to and had no effective role in participants' performance in perceptual tasks.

The null effect of phonemic coding ability on participants' performance in the perceptual tasks in the current study can be attributed to many possible reasons. According to Reynolds (2002), the second section in the Modern Language Aptitude Test (MLAT-II), which is designed to test phonemic coding ability, requires the test-taker to go through several procedures simultaneously. These include perceiving nonsense syllables which are presented orally. Test-takers must retain these syllables in their phonological short-term memory while they are taking the test. At the same time, they have to segment the phonological units forming these syllables. Finally, they must associate these phonological units with their graphemic symbols.

Given that the Llama test battery is very similar to, and based on, the MLAT (Saito, 2019), it has very similar procedures when testing phonemic coding ability. Therefore, it is claimed that phonemic coding ability is componential in nature. This is because, according to Skehan (1999, as cited in Reynolds, 2002), phonemic coding ability determines to what extent test-takers are able to exploit a combination of oral and written input at the same time. Therefore, Reynolds (2002) argues that phonemic coding ability consists of at least three subcomponents: phonological short-term memory (which is used during the participants' storing of phonological information), phonological awareness (which is used in the participants' segmentation of phonological units), and grapho-phonemic awareness (which is used in the participants' association of phonemes to their corresponding graphemes). Phonological short-term memory refers to being able to maintain phonological information in the short-term store (Kogan, 2020). Phonological awareness, on the other hand, refers to being able to divide words into smaller phonological units that do not have any semantic value (Reynolds, 2002).

As for the first subcomponent of phonemic coding ability, according to Kogan (2020), greater phonological short-term memory capacity is found to have a relationship with L2 learners' perceptual accuracy and it is the most researched subcomponent of phonemic coding ability. Reynolds (2002) argues that the major effect of phonological short-term memory, as evidenced by a number of studies (Kogan, 2020), is that phonological short-term memory can be an essential component that undergoes language aptitude. On the other hand, including phonological awareness as a subcomponent of phonemic coding ability is evidenced by Carroll (1993), who conducted a study of phonological awareness in his factor-analytic language studies (Stanovich, Cunningham, & Feeman, 1984, as cited in Carrol, 1993), to examine the construct of phonemic coding ability. Nevertheless, it is argued that, as phonological awareness plays a greater role in predicting reading in the case of alphabetic writing systems (Reynolds, 2002), it cannot be seen as a direct component of language aptitude.

According to Reynolds (2002), a number of studies have shown that phonological awareness is susceptible to training, which in turn excludes it being a subcomponent of language aptitude, as language aptitude has been argued to be not trainable and is not affected by outside influences (Carroll, 1981). Reynolds (2002) claims that results obtained from tests that rely on phonological awareness, such as MLAT-II and Llama-E, might not be precise enough to test language aptitude.

Therefore, these tests need to be revised to include testing phonological short-term memory separately from phonological awareness, which is because the susceptibility of phonological awareness to training. Rogers et al. (2017) concluded that prior experience and/or training have an effect on the Llama test results which, according to them, is not surprising based on the conceptualization of language aptitude by Granene (2013), who divided language aptitude into explicit language aptitude including Llama-B, E, and F and implicit language aptitude including Llama-D (refer to Subsection 2.3.3.2.1 for what these tests involve). According to Rogers et al. (2017), test-takers who have had prior instruction in L2 tend to outperform others and, therefore, results for the two populations should not be compared. Thus, the findings of the current study might have been different if the study had been directed at testing the participants' phonological short-term memory independently from their phonological awareness, or if this was discarded as a component of language aptitude, and they were trained before the treatment to segment words into smaller phonological units.

It is important to bear in mind that the possible bias in these results is caused by the phonemic coding ability test used in the study. Overall, despite the fact that the Llama test battery was found to have acceptable-to-good stable test-retest reliability on the whole, and all four subcomponents of the test showed acceptable reliability (Granena, 2013), it is claimed by a number of researchers that the current aptitude tests are still not sensitive enough to examine all language abilities (Artieda & Muñoz, 2016; Robinson, 2005, 2013). Artieda and Muñoz (2016) argue that, notwithstanding that the Llama test battery does not claim to measure all language-learning-related cognitive abilities for advanced learners, there are a number of cognitive abilities that are not tested with intermediate learners. Robinson, (2005, 2013) agrees with that, pointing out that, to date, there is no aptitude test that taps into all the different components of language aptitude that play an essential role in different stages of L2 learning. Granena (2013) mentions that the Llama tests are still not extensively standardised and consequently should be avoided in high-stakes situations, which is also mentioned in the Llama manual (Meara, 2005).

Apart from this, despite the fact that the Llama test battery was designed on a language neutrality basis (Granena, 2013, 2018; Rogers et al, 2017), it is argued that language aptitude differs across languages, i.e., it is possible for someone to succeed in learning a certain second language, depending on his/her language aptitude, but fail in another language (Kogan, 2020; Reynolds, 2002). MacWhinney (1995 as cited in Kogan, 2020) argues that this variety of the effect of language aptitude across languages might be caused by the nature of the learner's abilities, including his/her strengths and weaknesses in a large range of tasks. In addition, the variety of the effect of language aptitude across languages might be attributed to the linguistics characteristics that characterise a given language such as the orthographic depth or complexity of the inflectional system of that language. According to Reynolds (2002), research has shown that coming from different L1 backgrounds leads to having different strategies and skills in learning an L2. These differences occur in different areas of language learning, among them are orthographic processing strategies, such as decoding skills and reading comprehension, and speech segmentation processes, such as deletion, addition, or repetition of speech segments, which have been found to be affected by having alphabetic literacy (Koda, 1998; Read, Zhang, Nie, & Ding, 1986). These orthographic processing strategies and speech segmentation processes, according to Reynolds (2002), have a direct relationship with phonemic coding ability. Thus, the results obtained from tests such as MLAT-II and Llama-E might have a different relationship with language aptitude when testing different populations, especially as Llama-E test consists of graphemic and numeric symbols taken from a Roman script (Rogers et al., 2017). However, notwithstanding that the participants of the current study came from alphabetic writing system background, the target language used in this study does not use an alphabetic writing system, as discussed below.

Given that the target language in the current study is Arabic, which has an abjad writing system, and the participants' L1 language is English, which has an alphabetic writing system, it is important to take Reynolds' (2002) claim into consideration. The large distance between the Arabic and English orthographic systems in terms of their directionality, graphemic symbols, and type of writing system, might lead to weakening the role of phonemic coding ability in the participants' performance in perceptual tasks. This is because in abjad writing systems, only consonants have orthographic values which is not the case with the alphabetic writing systems where both vowels and consonants are represented in orthography. In addition, having different direction of script and joining graphemes together may have caused a number of difficulties in processing and decoding the Arabic script. This is supported by Reynolds (2002) who claims that most of the studies that have used phonemic coding ability tests were directed at European languages that used an alphabetic writing system, with a small number of studies being directed at languages with logographic writing systems, which were found to have lower predictive value of phonemic coding ability for L2 proficiency. This is because of having phonological awareness as a subcomponent of phonemic coding ability, which is affected by training and learning, as discussed above. It has been shown that illiterate L1 Portuguese adults (Morais, Bertelson, Cary, & Alegria, 1986) and literate L1 Chinese who never learned alphabetic orthography (Read et al., 1986) were equal in their inability in phonological awareness tasks, such as deleting and reversing phonemes. This led Reynolds (2002) to suggest avoiding using tests that involve phonological awareness in situations including non-alphabetic orthography in either L1 or L2 with learners who have low or no alphabetic literacy due to the low validity and reliability of the scores of these tests.

Considering Reynold's claim, different writing systems may cause different results of phonemic coding ability effects. However, given that Arabic has Abjad alphabetic writing system, but with different direction from that of English, further investigation may be needed to test the effect of phonemic coding ability on the orthographic availability effect.

As discussed in Subsection 5.4.1, there is a possibility in the current study that participants directed their attention to the wrong direction, as the Arabic writing system runs in the opposite direction to English. It is possible that the participants who had Arabic orthographic input (Arabic-script group and Romanised-script group) were trying to attend to orthography but just could not override the directionality. Therefore, they got useless information by looking at what they thought was the onset of a word and it was not. In addition to that, the fact that the Romanised script starts from the other direction would have been an additional factor biasing them to attend to the wrong part of the Arabic word. According to DeKeyser and Koeth (2011), specific areas of language aptitude depend largely on the context such as teaching methods or techniques in the field of education. Therefore, in the current context, the role of phonemic coding ability in the participants' performance in perceptual tasks might not have been the same if the participants were trained in basic information about the Arabic writing system in terms of its directionality and/or graphemic symbols. In other words, participants might not be given the opportunity to use their phonemic coding ability in the tasks used in the current study because they could not link the graphemes to their phonemes due to the opposite direction of the script.

5.4 Summary of the Chapter

The aim of this chapter was to conduct a detailed discussion of the study findings. It was shown in this review that orthographic input type only had an effect on word-level discrimination, and even then, the effect was only significant in the accuracy measure: the effect of orthographic input type was significant in neither word-level discrimination reaction times, nor in both accuracy and reaction times in phoneme-level discrimination. The effect of orthographic input was restricted to the case where participants were provided with both Romanised and Arabic scripts at the same time, where it was found to have a negative effect. The reason behind there being no effect of unfamiliar script alone could be the unfamiliarity and total foreignness of the writing system in terms of its graphemic symbols and directionality.

Having a negative or null effect from providing familiar and/or unfamiliar scripts might be attributed to many reasons. These include being unfamiliar with the phonological system of Arabic, having an additional input source which might have distracted the participants' attention, using digraphs to represent target sounds in the familiar script, and having the lowest phonemic coding ability scores. Apart from this, the literature shows that having an opaque orthography background, which was the case for the participants in this study, was found to decrease the participants' reliance on orthographic input (Bassetti et al. 2020; Burnham, Tyler, & Horlyck, 2002; Nimz & Khattab, 2020), which explains why orthographic input (familiar and unfamiliar) was found to be unhelpful in this study.

On the other hand, the performance of the participants in terms of accuracy and reaction time was affected by whether the phonemes were identical, similar, or novel compared to the participants' L1 phonemes. This effect occurs on both discrimination levels: word-level and phoneme level discrimination. In phoneme-level discrimination, discriminating between two identical or two novel phonemes was not significantly different, and both were better than discriminating between two similar phonemes. However, in word-level discrimination, discriminating between two new phonemes was easier than discriminating between two identical or two similar phonemes, and there was no significant difference between them.

Having no significant difference in discrimination between two identical and two similar phonemes in word-level discrimination was not due to difficulty in discriminating the contrast between the two phonemes in each pair, but instead because of the learning hindrance in the case of identical phonemes, as evidenced by the participants' performance in phoneme-level discrimination. Having the lowest performance in the similar-phoneme condition is supported by many L2 speech models, such as the SLM-r (Flege & Bohn, 2021), PAM (Best, 1995), and PAM-L2 (Best & Taylor, 2007), which argue that similar phonemes tend to be more difficult to discriminate than different ones. Finally, having the best performance in the case of new phonemes is also supported by the previously mentioned speech perception models which claim that they are easy to discriminate because it is easier for listeners to notice the difference between two phonemes. The participants' performance in the three levels of L1/L2 phoneme similarity might have been affected by the fact that the participants were not active learners of Arabic and their minimal experience of Arabic hindered their ability to categorise and discriminate some phonetic contrasts.

Furthermore, it was shown in this chapter that the two factors (orthographic input type and L1/L2 phoneme similarity) did not interact in some cases. These cases include providing a novel script only to participants, i.e., the case of the Arabic-script group. In this case, their performance could only be attributed to the perceptual difficulty caused by L1/L2 phoneme similarity. However, in the case of providing both novel and familiar scripts to participants, they had the lowest performance when discriminating between two similar phonemes. This can be attributed to the perceptual difficulty of the phonemes as well as lacking orthographic cues, as both were represented by the grapheme <s> and distinguished by (') which does not mark a contrast in the participants' L1 orthography. However, as mentioned above, these findings should be treated with caution as the initial analysis showed no significant effect.

The participants' ability to discriminate phonemes in word-level discrimination did not correlate. This can be because of the different analytical abilities used in tasks of the two discrimination levels (McGuire, 2010) or the increase of the short-memory load caused by having different stimuli in tasks of the two discrimination levels (Sweller, 2011). Moreover, participants' answers were at the chance level in one of the orthographic input-type group in the identical-phoneme condition, which decreases the possibility of a correlation with the other phoneme discrimination level. On the other hand, in the case of similar phonemes, one of the possible explanations was the perceptual difficulty of this pair as the performance was either at the chance level or slightly above the chance level in some conditions indicating that participants were guessing, which clearly accounts for the fact there was no correlation. In addition, participants unfamiliarity of the Arabic phonological system might have affected their performance negatively.

Apart from this, phonemic coding ability was found to be unrelated and to play no effective role in the participants' performance in perceptual tasks. The reason for the null effect of phonemic coding ability might be due to the fact that the test used in this study included phonological awareness as a subcomponent which was found to be affected by prior instruction and training, and consequently its results might not be accurate. This is because it has been evidenced that people with prior instruction in L2 tend to perform better than others, and therefore results of people who had prior instructional experience should not be compared with those who have no prior instructional experience. (Rogers et al., 2017). Furthermore, it is argued that language aptitude differs across languages (Kogan, 2020; Reynolds, 2002). Having a large distance between the Arabic and English orthographic systems in many aspects, including its directionality, might have led to weakening the role of phonemic coding ability in the participants' performance in perceptual tasks. Moreover, despite the acceptable reliability of the Llama tests (Granena, 2013), it is argued that current language aptitude tests are still not sensitive enough to examine all language abilities, and the Llama test battery, in particular, needs to have extensive standardization, as acknowledged by its designer, Meara (2005).

6 Conclusion

6.1 Overview

This study was designed to determine the relationship between orthographic input type and L1/L2 phoneme similarity in discriminating L2 Arabic phonemes by L1 English listeners. In addition, it looked at the role of the participants' phonemic coding ability in modulating the effect of orthographic input and L1/L2 phoneme similarity in discriminating these phonemes.

The results of this investigation showed that orthographic input type impacted on the participants' performance. This effect can be summarised thus: when it was only unfamiliar script, i.e., Arabic script, it did not have any effect, whereas when it was a combination of both familiar and unfamiliar scripts it affected participants' performance negatively. The best performance, however, was by those who were not exposed to any orthographic input. This effect was only in discriminating phonemes embedded in words, not when they were presented in isolation.

The clearest finding to emerge from this study is that L1/L2 phoneme similarity played an essential role in the participants' performance. Their performance when they were asked to discriminate between two novel or two identical phonemes was better than their performance when asked to discriminate between two similar phonemes, whether these phonemes were uttered in isolation or embedded in words. These findings are supportive evidence for the SLM-r (Flege & Bohn, 2021) and PAM-L2 (Best & Taylor, 2007), as these models argue that similar phonemes tend to be more difficult to discriminate than different ones.

The relationship between the two factors (orthographic input type and L1/L2 phoneme similarity) is shown in the following points:

- The lowest performance was reported by those who were presented with a familiar and an unfamiliar script at the same time in the case of discriminating similar phonemes.
- This lowest performance was attributed to the perceptual difficulty of the phonemes, as well as the lack of orthographic cues, as both were represented with <s> and distinguished by (’), which does not mark a contrast in their L1.
- There was no correlation between participants’ performance in word-level and phoneme-level discrimination in all cases of L1/L2 phoneme similarity.
- Having no correlation between word-level and phoneme-level discrimination can be attributed to the different analytical abilities and/or the increase of short-term memory load in the tasks of the two levels of discrimination.
- Having no correlation between word-level and phoneme-level discrimination in the case of identical phonemes can be attributed to the fact that word learning was impaired as participants’ responses were at the chance of level in this condition.
- Having no correlation between word-level and phoneme-level discrimination in the case of similar phonemes can be attributed to the perceptual difficulty of similar phonemes as the participants’ discrimination of the contrast between the two similar phonemes was the lowest among other conditions. In addition, pharyngealisation spread might have had a negative effect on participants’ performance due to their unfamiliarity of the phonological system of Arabic.

Apart from L1/L2 phoneme similarity, phonemic coding ability was found to be unrelated and have no effect on the participants’ performance in perceptual tasks.

This study is one of the first attempts to examine these three factors together, namely: orthographic input type; L1/L2 phoneme similarity; and phonemic coding ability. The findings of this study contribute in several ways to our understanding of L2 phonological development and provide essential information about the role of L1/L2 phoneme similarity in the acquisition of L2 phonemes. This is made clear by the fact that new and identical phonemes were easier to discriminate than similar phonemes. The study contributes to our understanding of the dominant speech perception models, such as the SLM-r (Flege & Bohn, 2021), PAM (Best, 1995), and PAM-L2 (Best & Taylor, 2007), as the findings of this study support what these models hypothesise. These models claim that similar sounds tend to be more difficult to discriminate than different ones, especially when two similar phonemes in L2 map to only one L1 phoneme, which was supported by the findings of this study. On the other hand, the findings of this study add to the growing body of research that indicates that the effect of orthographic input on L2 phonological development needs more systematic investigation in terms of script familiarity, L1 background, and phoneme perceived difficulty. It was found that having a familiar script with an unfamiliar script had a negative effect and having unfamiliar script alone had no effect. It is, then, worth suggesting that in the context of distant languages such as Arabic and English, that it is better to avoid orthographic input during phonological learning, especially in the case when the perceived difficulty is evidenced to be high. Apart from this, the findings go some way towards enhancing our understanding of the weak role of phonemic coding ability in helping participants to benefit from an unfamiliar script and/or a familiar script in discriminating L2 phonemes. However, as discussed in the previous chapter, this might be attributed to the susceptibility of some of the subcomponents of the phonemic coding ability test used in this study or the large distance between the two language orthographic systems. This means that if the participants had had better phonological training or had come from different L1 background they might have performed differently.

6.2 Pedagogical Implications

Notwithstanding that participants in this study were naïve listeners - i.e., not learners - of Arabic, the aim of doing such research is to shed light on language learning. According to Binasfour (2018), research on L2 Arabic learning and teaching has tended to be directed to the design of pedagogical materials that focus on grammar, reading, and writing rather than phonology. This increases the demand for developing new methodologies and approaches to train sound perception which might eventually result in having accurate production as well, because accurate L2 production is based on accurate L2 perception (Chan, 2012; Flege, 1995; Kogan, 2020). The acquisition of Arabic emphatic consonants could be facilitated by directing learners' attention to noticing the differences between them and their non-emphatic counterparts based on surrounding vowels which are affected by pharyngealised spread, as it is evidenced that even L1 Arabic speakers rely on surrounding vowels to discriminate this contrast (Alhumaid, 2019; Jongman et al, 2011). According to Schmidt (1990, as cited in Binasfour, 2018), noticing is an essential technique in L2 learning because it helps to identify errors and correct them. This leads Binasfour (2018) to suggest including explicit training in perception and production in L2 Arabic phonology because not all learners have the ability to notice the essential features of Arabic sounds implicitly. In addition, according to Sweller (2011), one essential way of facilitating learning is by providing explicit instruction. Noticing these features also helps to expand learners' ability to acquire more vocabulary through their ability to discriminate the contrasts in minimal pairs because, according to Odisho (2005, as cited in Hayes-Harb & Durham, 2016), the inability to discriminate consonant contrasts results in losing the ability to understand what thousands of Arabic words mean.

Furthermore, one of the issues that emerged from the findings of this study is that visual word learning is affected negatively by the number and type of orthographic inputs. This is evidenced as the learning of lexical representations occurring successfully in all nine conditions of the experiment except in the identical phoneme condition by participants who had Romanised along with Arabic scripts. In this condition, the participants' responses were at the chance level. Notwithstanding that the phonemes in this condition were identical to two English phonemes, indicating that the contrast between the two phonemes in this pair was already acquired, their word learning was impaired. In addition, the performance of the group that had Romanised along with Arabic scripts was lower in general than the other groups. Sweller (2011) argues that the storage capacity and duration of working memory are limited when dealing with novel information when it can store no more than around seven items plus or minus two (Miller, 1956) for no more than 20 seconds. Its processing capacity, however, is less than when it can process three to four items of information at the same time. Therefore, it is recommended to take the capacity and duration limits of working memory into consideration when providing instructional procedures. This is because increasing the load of working memory may lead to constraining the ability to learn. Therefore, according to cognitive load theory (Sweller, 1988 as cited in Sweller 2011), the redundancy effect, which occurs when additional information that redescribes other information is provided, causes learners to attempt to process two elements of information, and they might attempt to associate these two elements together, leading to having an increased extraneous cognitive load, which has a negative effect on the occurrence of learning. Given that two different direction scripts (left-to-right script and right-to-left script) were provided to the participants in this group, this may have confused them more, especially if they were trying to associate two elements. This leads to negatively affecting the occurrence of learning. However, the performance in other conditions indicated that adult UK learners had the ability to learn lexical representations rapidly as they needed to be exposed to 10 exposures or fewer in order to build new lexical items phonologically.

Moreover, phonemic coding ability playing no effective role in participants' performance in perceptual tasks might have implications for the findings. It is probably the case that the distant orthographic systems of English and Arabic in terms of their directionality, graphemic symbols and type of writing systems might have weakened this role. This in turn challenges the concept of the language neutrality of the Llama test battery. According to Kogan (2020), the predictability of language aptitude tests is not directed at specific languages, but instead to any foreign language regardless of what phonetic and linguistic properties it has. Kogan (2020) suggests that if a learner has the opportunity to choose an L2 (for academic or military purposes), looking at the phonetic distance between his/her L1 and L2 in addition to his/her psychoacoustic profile, which includes his/her language aptitude, might provide a useful guide for better L2 selection. However, the role of phonemic coding ability in an L2 that is very distant from the L1 in its orthographic system, directionality, and graphemic symbols needs more systematic and focused investigation, which was not possible to consider in this study due to the time limitation and resource constraints.

6.3 Limitations of the Study

Due to practical constraints, this thesis could not conduct a comprehensive review on L2 Arabic learners population because the participants of the current study were not active learners of Arabic, instead they were naïve listeners. As mentioned in the previous chapter, naïve listeners tend to have difficulty in discriminating contrasts, especially those that do not exist in their L1, more than L2 learners which is because of their minimal experience in the L2 (Best & Taylor, 2007). Therefore, if this study was carried out on L2 Arabic learners, different findings might be obtained.

However, because it was necessary to consider the unfamiliarity of Arabic sounds and script and due to the Covid-19 pandemic, it was not possible to have access to institutions that offer Arabic language courses, and to access learners who are in the initial stages of learning Arabic, during the intended timeframe of data collection. That is why the sampling method used in this study was convenience sampling. This might have had a negative effect on the participants' motivation, which in turn affected their performance negatively.

Moreover, as discussed in Subsection 5.4.5, it is argued that the results of tests that include phonological awareness strategies, such as the Llama-E test, are not accurate enough to assess language aptitude because it has been found that it is possible to train phonological awareness. Therefore, there is a need to revise these tests (Reynolds, 2002). This is supported by Rogers et al. (2017), who claim that the Llama test results are affected by prior experience. This implies that the effect of phonemic coding ability might not be the same if participants are tested on their phonological short-term memory separately from their phonological awareness, if they are exposed to some training, or if they are tested using a different test.

Furthermore, as discussed in Subsections 5.3.1 and 5.3.2, the effect of L1/L2 phoneme similarity was not the same in the three conditions. This effect is in line with many previous studies and supports the key claims of models such as the SLM-r (Flege & Bohn, 2021) and PAM (Best, 1995). However, their performance in the three conditions might have been affected by the order effect. This, especially, might have affected their performance in the first session (Identical-phoneme condition) negatively due to their unfamiliarity with what would come next in the experiment, which was not the case with the other two conditions. This can be avoided by counterbalancing the sessions, so such an effect does not occur. Having the same order in the three phoneme conditions (identical, followed by similar, followed by new phoneme condition) aimed to control the difficulty level equally for all participants, starting with what was thought to be the easiest as they were mapped to two L1 phonemes, then moving gradually to the more difficult contrasts. However, this approach was found to have a negative effect on the participants' performance because, as mentioned above, they were unfamiliar with what would come next in the first phoneme condition, causing them to have lower performance in the word-level discrimination.

Apart from this, given that there were violations of some of the assumptions underlying the use of parametric tests, non-parametric tests were also conducted to validate the significant findings. Because of the current awareness of the replicability crisis, only significant findings were further investigated with follow-up non-parametric tests and discussed in this thesis as an attempt to be as conservative as possible about claiming significant effects. Despite the fact that only two cases where discrepancies between parametric and non-parametric tests were found, Thus, the findings of the two cases where discrepancies did occur should be treated with caution and verified with further investigations.

In addition, having no significant difference between Arabic-script group and No-script group might be strongly attributed to the fact that the Arabic script was ignored due to its novelty, or participants in the Arabic-script group were trying to attend to orthography but they just could not cope with the directionality. As mentioned previously, they were not explicitly told that the Arabic script reads from right to left. Given the fact that the instructions did not have an effect on the participants' performance in the studies of Alhumaid (2019) and Showalter and Hayes-Harb (2015), and the procedure was done online due to the Covid 19 pandemic, the introduction about how Arabic is written was removed. The participants' performance might have been different if they had been explicitly instructed about the direction of Arabic script.

Furthermore, as mentioned in Subsection 5.3.1, the null effect of orthographic input in some cases (accuracy in phoneme-level discrimination and reaction time in both phoneme-level and word-level discrimination) might be due to low power which might be caused by the small sample size. Therefore, ideally, it would be better if a power analysis was carried out prior to the study to work out what the minimum sample size was to give a sufficient power (.80) to detect a significant effect, if such an effect is true.

6.4 Future Directions

Strategies to enhance the research on the role of language aptitude in the effect of orthographic input on L2 phonological perception might involve investigating other language aptitude subcomponents rather than phonemic coding ability. For example, associative memory is defined as the ability to associate sounds with meanings and to retain these associations in memory (Moskovsky et al., 2015; Saito, 2019), which might have affected the participants' performance in perceptual tasks. Associative memory, which is tested by Llama-B, or the interaction between the two subcomponents: phonemic coding ability and associative memory, might play a greater role than phonemic coding ability alone. Apart from associative memory, the ability to recognise sequences of sounds may lead to easier and faster L2 learning (Meara, 2005). This ability, however, is not tested by Llama-E but rather Llama-D. Notwithstanding that Llama-D has been found to be related to implicit language learning and processing (Granene, 2013; Artieda & Muñoz, 2016), Granena (2018) argues that implicit memory ability, involving the ability to retrieve incidentally previously learned information, helps learners to recall foreign language words and relate their meaning.

Saito (2019) also argues that learners' ability to recognise sequences of sounds is found to be strongly associated with the extent to which they are successful in learning L2 in certain contexts. That is to say that these components of language aptitude might have stronger effect if they interact together on the participants' ability to discriminate phonemes in different conditions of L1/L2 phoneme similarity and orthographic input. Referring to Skehan's (2016) model of language aptitude and the stages of L2 learning, Saito (2019) emphasises that the interaction between aptitude and different stages of L2 learning can occur with more than one aptitude component at the same time, instead of having one aptitude component in a given stage of L2 learning. This is because all aspects of a L2 learner's developing system are related to each other. However, a full discussion of language aptitude is beyond the scope of this study. Therefore, it might be worth investigating other components of language aptitude to enhance our knowledge of the role of language aptitude in L2 phonological development.

Moreover, another possible area of future research would be to investigate the role of phonemic coding ability in the case of an L2 that belongs to the same orthographic system. As discussed in Subsection 5.4.5, according to Reynolds (2002), it is found that different L1 backgrounds lead to the use of different strategies and skills in learning an L2. These differences include orthographic processing strategies and speech segmentation processes, which according to Reynolds (2002), have a direct relationship with phonemic coding ability. This raises a question, if this study was carried out on participants who have a smaller cross-language orthographic distance between their L1 and L2, would their phonemic coding ability play the same role as in the current study? If no, how and why would this role differ from the findings of the current study?

Apart from this, this study could be extended to include L2 production in addition to perception to test whether orthographic input, L1/L2 phoneme similarity, and phonemic coding ability have the same impact on perception and production or if it is different when involving production. Despite the fact that the acquisition of production is based on the acquisition of perception (Chan, 2012; Flege, 1995; 2016; Kogan, 2020), different findings might be found with production, especially in the case of new phonemes. This is because they were found to be easily discriminated perceptually in the current study, which does not, however, necessarily mean that they will be easily produced. In other words, when it comes to production, participants might find it more difficult as they do not exist in their L1. This is because, according to Flege (2016), it takes time to transfer properties from perceptual phonetic categories to phonetic implementation rules, which means that differences between perception and production found at a given time might not be found later at another time. Thus, Flege (2016) argues that having accurate perception does not necessarily lead to having accurate production, but accurate production is based on accurate perception.

In summary, it has been shown by this study that the orthography effect on L2 phonological development is not modulated by phonemic coding ability, but it is by L1/L2 phoneme similarity. It is hoped that the findings reported by this study will contribute to the investigation of the role of orthography in L2 phonological development.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, 30(4), pp.481-509.

Abu-Rabia, S., & Sammour, R. (2013). Spelling errors' analysis of regular and dyslexic bilingual Arabic-English students. *Open Journal of Modern Linguistics*, 3(1), pp.58-68.

Alghamdi, M. (2015). *الصوتيات العربية والفونولوجيا* [Arabic Phonetics and Phonology]. King Fahad National Library.

Alhumaid, H. (2019). *Orthography in L2 phonology acquisition: The role of phonemic coding ability* (Unpublished doctoral dissertation). University of Reading, Reading, United Kingdom.

Ali, S. H., Carr, P., & Ruit, K. (2016). Validity and reliability of scores obtained on multiple-choice questions: why functioning distractors matter. *Journal of the Scholarship of Teaching and Learning*, 16(1), pp.1-14.

Aljasser, F., & Vitevitch, M.S. (2018). A web-based interface to calculate phonotactic probability for words and nonwords in Modern Standard Arabic. *Behavior Research Methods*, 50, pp.313-322.

Alkhatib, S. (2008). *On the directionality of emphasis spread*. Paper presented at the Proceedings of the 2008 annual conference of the Canadian Linguistic.

Almahmoud, M., (2013). Discrimination of Arabic contrasts by American learners. *Studies in Second Language Learning and Teaching*, 3(2), pp.261-292.

Amayreh, M., & Dyson, A. (1998). The acquisition of Arabic consonants. *Journal of Speech, Language, and Hearing Research*, 41(3), pp.642–653.

Artieda, G., & Muñoz, C. (2016). The Llama tests and the underlying structure of language aptitude at two levels of foreign language proficiency. *Learning Individual Differences*, 50, pp.42-48.

Bani Salameh, M. (2015). The phonetic nature of consonants in Modern Standard Arabic. *English Linguistics Research*, 4(3), pp.30-57.

Bassetti, B. (2005). Effects of writing systems on second language awareness in English learners of Chinese as a foreign language. In: P. Cook & B. Bassetti (Eds), *Second language writing systems* (pp. 335-356). Retrieved from: http://ebookcentral.proquest.com Created from reading on 2019-4-26.

Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language and Literacy*, 9(1), pp.1-12.

Bassetti, B. (2007). Bilingualism, Biliteracy and Metalinguistic Awareness: Word Awareness in English and Japanese Users of Chinese as a Second Language. *BISAL*, 2(1), pp.1-21.

Bassetti, B. (2008) Orthographic input and second language phonology. In T. Piske & M. Young-Scholten (Eds), *Input matters in SLA* (pp.191-206). Clevedon, UK: Multilingual Matters.

Bassetti, B. (2017). Orthography affects second language speech: Double letters and geminate production in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(11), 1835-1842.

Bassetti, B., & Atkinson, N. (2015). Effects of orthographic forms on pronunciation in experienced instructed second language learners. *Applied Psycholinguistics*, 36(1), pp.67–91.

Bassetti, B., Cerni, T., & Masterson, J. (2022). The efficacy of grapheme-phoneme correspondence instruction in reducing the effect of orthographic forms on second language phonology. *Applied Psycholinguistics,* 43(3), pp. 386-705.

Bassetti, B., Escudero, P., & Hayes-Harb, R. (2015). Second language phonology at the interface between acoustic and orthographic input. *Applied Psycholinguistics*, 36, pp.1-6.

Bassetti, B., Mairano, P., Masterson, J., & Cerni, T. (2020). Effects of orthographic forms on second language speech production and phonological awareness, with consideration of speaker-level predictors. *Language Learning,* 70(4), pp.1218-1256.

Bassetti, B., Sokolović-Perović, M., Mairano, P., & Cerni (2018) Orthography-induced length contrasts in the second language phonological systems of L2 speakers of English: Evidence from minimal pairs. *Language and Speech,* 61(4), pp.577-597.

Best, C. (1995). A direct realist view of cross-language speech perception. *Speech perception and linguistic experience: Issues in cross-language research*, pp.171- 204.

Best, C., & Taylor, M. (2007). Nonnative and second-language speech perception: commonalities and complementarities. In: O. Bohn, & M. J. Munro (Eds), *Language Experience in Second Language Speech Learning: in honour of James Emil Flege* (pp. 13-45). John Benjamins Publishing Company.

Binasfour, H. (2018). *Investigating the Perception and Production of the Arabic Pharyngealised*

Sounds by L2 Learners of Arabic (PhD thesis). University of Reading, Reading, United Kingdom.

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America*, 117(6), pp.3890-3901.

Card, E. (1983). *A phonetic and phonological study of Arabic emphasis* (PhD thesis), Cornell University, New York, USA.

Carroll, J. (1962). The prediction of success in intensive foreign language training. In: R. Glaser (Ed.), *Training Research and Education* (pp.87-136). New York: Wiley.

Carroll, J. (1981). Twenty-five years of research on foreign language aptitude. In: K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp.83-118). Rowley, Mass: Newbury House.

Carroll, J. (1993). *Human cognitive abilities: a survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Cerni, T., Bassetti, B., & Masterson, J. (2019). Effects of orthographic forms on the acquisition of novel spoken words in a second language. *Frontiers in Communication*, 4(31), pp.1-14

Chan, A. (2012). Cantonese English as a second language learners' perceived relations between "similar" L1 and L2 speech sounds: a test of the speech learning model. *The Modern Language Journal*, 96(1), pp.1-19.

Cutler, A. (2015). Representation of second language phonology. *Applied Psycholinguistics*, 36(1), pp.115-128.

Davis, S., & Ragheb, M. (2014). Geminate representation in Arabic. In: Farwaneh, F. and Ouali, H. (Eds), *Perspectives on Arabic Linguistics XXIV-XXV (*pp.3-19). Philadelphia/Amsterdam: John Benjamins Co.

DeKeyser, R., & Koeth, J. (2011). Cognitive aptitudes for second language learning. In: E. Hinkle (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp.395-406). Routledge: ProQuest Ebook.

Dörnyei, Z. (2005). *The psychology of language learner: individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dörnyei, Z. (2007). *Research methods in Applied Linguistics; Quantitative, Quantitative and Mixed methods*. Oxford: Oxford University Press.

Ehri, L. (2005). Learning to Read Words: Theory, Findings, and Issues. *Scientific*

Studies of Reading, 9(2), pp.167-188.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Embarki, M., Yeou, M., Guilleminot, C., & Al Maqtari, S. (2007). An acoustic study of coarticulation in modern standard Arabic and dialectal Arabic: pharyngealized vs non-pharyngealized articulation. *ICPhS XVI*, pp.141-146.

Erdener, D. & Burnham, D. (2005). The role of audiovisual speech and orthographic information in non-native speech production. *Language Learning*, 55(2), pp.191-228.

Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition: Explaining the* *attainment of optimal phonological categorization* (PhD thesis). University of Utrecht, Utrecht, The Netherlands.

Escudero, P. (2015). Orthography plays a limited role when learning the phonological forms of new words: the case of Spanish and English learners of novel Dutch words. *Applied Psycholinguistics*, 36(1), pp.7-22.

Escudero, P. & Wanrooij, K. (2010). The Effect of L1 Orthography on Non-native Vowel Perception. *Language and Speech*, 53, pp.343-365.

Escudero, P. & Williams, D. (2012). Native dialect influences second-language vowel perception: Peruvian versus Iberian Spanish learners of Dutch. *The Journal of the Acoustical Society of America*, 131(5), pp.406-412.

Escudero, P., Broersma, M., & Simon, E. (2013). Learning words in a third language: Effects of vowel inventory and language proficiency. *Language and Cognitive Processes*, 28(6), pp.746-761.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, 36(2), pp.345-360.

Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn’t always help. *Bilingualism: Language and Cognition*, 17(2), pp.384-395.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. SAGE Publications Ltd.

Flege, J. (1986). Effects of equivalence classification on the production of foreign language speech sounds. In: A. James & J. Leather (Eds), *Sound Patterns in Second Language Acquisition* (pp.9-40). De Gruyter Mouton.

Flege, J. (1991). Perception and Production: the relevance of phonetic input to L2 phonological learning. In: T. Huebner & C. Ferguson (Eds), *Crosscurrents in Second Language Acquisition and Linguistic Theories* (pp.249-289). John Benjamins Publishing Company.

Flege, J. (1995). Second language speech learning: theory, finding, and problems. In: W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross language research* (pp.233-277). York: York Press.

Flege, J. (2016). The role of phonetic category formation in second language speech acquisition. Talk presented at New Sounds, Aarhus University, Aarhus, Denmark.

Flege, J., & Bohn, O. (2021). The revised Speech Learning Model (SLM-r). In Wayland, R. (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3-83). Cambridge: CUP.

Flege, J., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese Adults' perception of /ɹ/ and /l/. *The Journal of the Acoustical Society of America*, 99(2), pp.1161-1173

Frost, R., Repp, B., & Katz, L. (1988). Can speech perception be influenced by simultaneous presentation of print? *Journal of Memory & Language*, 27(6), pp.741-755.

Geva, E., & Zadeh, Z. Y. (2006). Reading efficiency in native English speaking and English-as-a-second-language children: the role of oral proficiency and underlying cognitive-linguistic processes. *Scientific Studies of Reading*, 10(1), pp.31-57.

Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA language aptitude test. In: G. Granena & M. H. Long (Eds., *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp.105-129). Amsterdam, The Netherlands: John Benjamins.

Granena, G. (2018). Cognitive aptitudes and L2 speaking proficiency: links between Llama and Hi-lab. *Studies in Second Language Acquisition*, pp.1-24.

Granena, G., & Long, M. (2013). Age of onset, length of residence, aptitude and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29(3), pp.311-343.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America, 107*(5), pp.2711-2724.

O'Leary-Kelly, S., & Vokurka, R. (1998). The empirical assessment of construct validity. *Journal of Operations Management,* 16(4), pp.387-405.

Hammersley, M. & Atkinson, P. (1995). *Ethnography: Principles in Practice* (2nd ed.). London: Routledge.

Han, J-I., & Oh, S. (2018) The role of phonetic similarity and orthographic information in asymmetrical lexical encoding in second language. *Journal of Psycholinguistic Research*, 47(5), pp. 1015-1033.

Hanulíkuvá, A., Dediu, D., Fang, Z., Bašnaková, J., & Huettig, F. (2012). Individual differences in the acquisition of complex L2 phonology: a training study. *Language Learning*, 62(2), pp.79-109.

Harnsberger, J. (2001). The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: a multidimensional scaling analysis. *Journal of Phonetics*, 29(3), pp.303-327.

Hayes-Harb, R. & Barrios, S. (2021). The influence of orthography in second language phonological acquisition. *Language Teaching*, 54(3), pp. 297-326.

Hayes-Harb, R. & Cheng, H-W. (2016). The influence of the Pinyin and Zhuyin writing systems on the acquisition of Mandarin word form by native English speakers. Frontiers in Psychology, 7(785).

Hayes-Harb, R. & Durham, K. (2016). Native English speakers' perception of Arabic emphatic consonants and influence of vowel context. *Foreign Language Annals*, 49(3), pp.557-572.

Hayes-Harb, R. & Hacking, J. (2015). The influence of written stress marks on native English speakers' acquisition of Russian lexical stress contrasts. *The Slavic and East European Journal*, 59(1), pp. 91-109.

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech*, 53(3), pp.367–381.

Hellmuth, S. (2014). *Towards a research-led approach to the teaching of Arabic pronunciation*. Paper presented at the Proceedings of the International Symposium on the Acquisition of

Second Language Speech Concordia Working Papers in Applied Linguistics.

Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. London, UK: Longman.

Holmes, V., & Ng, E. (1993). Word-specific knowledge, word-recognition strategies, and spelling ability. *Journal of Memory and Language*, 32, pp.230-257.

Hu, W., Tao, S., Li, M., & Liu, C. (2019). Distinguishing and assimilation in vowel perception in a second language. *Journal of Speech, Language, and Hearing Research*, 62(12), pp 4534-4543.

Hu, X., Ackermann, H., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2012). Language aptitude for pronunciation in advanced second language (L2) Learners: Behavioural predictors and neural substrates. *Brain and Language*, 127(3), pp.366-376.

Ibrahim, R. (2011). Literacy problems in Arabic: sensitivity to diglossia in tasks involving working memory. *Journal of Neurolinguistics*, 24, pp.571-582.

Jackson, J. (2016). The effects of novel orthographic input and phonetic instruction in second language phonological acquisition (MA dissertation). University of Utah, Salt Lake City, Utah, USA.

Jongman, A., Herd, W., Al-Masri, M., Sereno, J., & Combest, S. (2011). Acoustic and perception of emphasis in Urban Jordanian Arabic. *Journals of Phenetics*, 39(1), pp.85-95.

Iskarous, K., & Mooshammer, C. (2021). Coarticulation. In R. Knight & J. Setter (Eds.), *The Cambridge Handbook of Phonetics* (Cambridge Handbooks in Language and Linguistics, pp. 106-132). Cambridge: Cambridge University Press.

Koda, K. (1998). The role of phonemic awareness in second language reading. *Second Language Research*, 14(2), pp. 194-215.

Kogan, V. (2020). *The effect of first language perception on the discrimination of a non-native vowel contrast: investigating individual differences* (PhD thesis). University of Barcelona, Barcelona, Spain.

Ladefoged, P. & Ferrari, D. (2012). *Vowels and Consonants* (3rd ed.). Blackwell Publishing Ltd.

Lado, R. (1957). *Linguistics across cultures: applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

Li, S. (2016). The construct validity of language aptitude. *Studies in Second Language Acquisition*, 38(4), pp.801-842.

Loewen, S., & Plonsky, L. (2015). *An A–Z of Applied Linguistics Research Methods*. Macmillan International Higher Education.

Maddieson, I., (2013) Vowel Quality Inventories. In: Dryer, Matthew S. & Haspelmath, Martin (Eds), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at: http://wals.info/chapter/2, Accessed on 2021-4-3.)

Major, R. & Kim, E. (1999). The similarity differential rate hypothesis. *Language Learning*, 49(1), pp.151-183.

Major, R. (2001). *Foreign accent: the ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Erlbaum.

Massaro, D., Cohen, M., & Thompson, L. (1990). Visible language in speech perception: Lipreading and reading. *Visible Language*, 22(1), pp.8-31

Mathieu, L. (2014). *The influence of unfamiliar orthography on L2 phonolexical acquisition* (PhD Thesis). The University of Arizona, Tucson, Arizona.

Mathieu, L. (2016). The influence of foreign scripts on the acquisition of a second language phonological contrast. *Second Language Research*, 32(2), pp.145-170.

McAllister, R., (2007). Strategies of realization of L2-categories. In: O. Bohn, M. J. Munro (Eds), *Language Experience in Second Language Speech Learning: on honor of James Emil Flege* (pp.153-166). John Benjamins Publishing company.

McAllister, R., Flege, J., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, 30(2), pp.229-258.

McGuire, G. (2010). A brief primer on experimental designs for speech perception research. *Laboratory Report*, 77.

Meara, P. (2005). *Llama Language Aptitude Tests*. Swansea: Lognostics.

Meara, P. & Rogers, V. (2019). *The LLAMA Tests v3. Llama-B v3.* Cardiff: Lognostics.

Meara, P. & Rogers, V. (2019). *The LLAMA Tests v3. Llama-D v3.* Cardiff: Lognostics.

Meara, P. & Rogers, V. (2019). *The LLAMA Tests v3. Llama-E v3.* Cardiff: Lognostics.

Meara, P. & Rogers, V. (2019). *The LLAMA Tests v3. Llama-F v3.* Cardiff: Lognostics.

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), pp. 81-97.

Morais, J., Bertelson, P., Carey, L., & Algeria, J. (1986). Literacy training and speech segmentation. *Cognition*, 24(1-2), pp. 45-64.

Moskovsky, C., Alshahrani, M., Ratcheva, S., & Paolini, S. (2015). Aptitude as a predictor of second language achievement: an investigation in the Saudi Arabian context. *Arab World English Journal*, 6(1), pp.3-21.

Mubarak, H., Abdeali, A., Darwish, K., Eldesouki, M, Samih, Y., & Sajjad, H. (2019). *A system for diacritizing four varieties of Arabic*. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations. Hong Kong, China.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34(4), pp.520-531.

Nation, P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, 26(2), pp.1-16.

Nimz, K. & Khattab, G. (2020). On the role of orthography in L2 vowel production: the case of Polish learners of German. *Second Language Research*, 36(4), pp.623-653.

O'Leary-Kelly, & Vokurka, R. (1998). The empirical assessment of construct validity. *Journal of Operations Management*, 16(4), pp.387-405.

Pasta D. (2009), "Learning when to be discrete: continuous vs. categorical predictors", Proceedings of the SAS Global Forum 2009, http://support.sas.com/resources/papers/proceedings09/248-2009.pdf.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, & E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* 10.3758/s13428-018-01193-y

Pikulski, J. J., & Chard, D. J. (2005). Fluency; bridge between decoding and reading comprehension. *International Reading Association*, 58(6), pp.510-519.

Piske, T. (2007). Implications of James E. Flege's research for the foreign language classroom. In O. Bohn, & M. J. Munro (Eds), *Language Experience in Second Language Speech Learning: in honor of James Emil Flege* (pp.302-314). John Benjamins Publishing company.

Pisoni, D., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Percept Psychology*, 15(2), pp.285-290.

Read, C., Yun-Fei, Z., Hong-Yin, N., & Bao-Qing, D. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition*, 24(1-2), pp. 31-44.

Reynolds, B. (2002). *Phonetic Coding in Language Aptitude: Different Roles for Different Languages.* Retrieved from: https://www.tuj.ac.jp/tesol/publications/working-papers/vol-14/reynolds.html.

Roach, P. (2000). *English Phonetics and Phonology* (3rd ed.). Cambridge: Cambridge University Press.

Roach, P. (2008). British English: Received Pronunciation. J*ournal of International Phonetic Association*, 34(2), pp. 239-245.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, 25, pp.46-73.

Robinson, P. (2013). Aptitude in second language acquisition. In C. Chappelle (Ed.), *The Encyclopedia of Applied Linguistics.* Blackwell Publishing.

Rogers, V., Meara, P., Branett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of European Second Language Association*, 1(1), pp.49-60.

Saiegh-Haddad, E., & Henkin-Roitfarb, R. (2014). The Structure of Arabic Language and Orthography. In: E. Saiegh-Haddad, & M. Joshi (Eds), *Handbook of Arabic Literacy: Insights and Perspectives* (pp.3-28).

Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, 67(3), pp. 665-693.

Saito, K. (2019). The role of aptitude in second language segmental learning: the case of Japanese learners' English /ɹ/ pronunciation attainment in classroom settings. *Applied Psycholinguistics*, 40(1), pp.183-204.

Saito, K., Suzukida, Y., & Sun, H. (2018). Aptitude experience and second language pronunciation proficiency development in classroom settings: a longitudinal study. *Language Teaching Research*, 22, pp.398-417.

Saito, K., Sun, H., & Tierney, A. (2019). Explicit and implicit aptitude effects on second language speech learning: scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. *Bilingualism: Language and Cognition*, 22(5), pp.1123-1140.

Salameh, M., & Abu-Melhim, A. (2014). The phonetic nature of vowels in Modern Standard Arabic. *Advances in Language and Literacy Studies*, 5(4), pp.60-67.

Sasaki, M. (2012). The modern language aptitude test (Paper and pencil version). *Language Testing*, 29(2), pp.315-321.

Schmidt, A. M., (2007). Cross-language consonant identification: English and Korean. In: O. Bohn, & M. J. Munro (Eds), *Language Experience in Second Language Speech Learning: on honor of James Emil Flege* (pp.185-200). John Benjamins Publishing company.

Schwartz, G., & Kaźmierski, K. (2020). Vowel dynamics in the acquisition of L2 English – an acoustic study of L1 Polish learners. *Language Acquisition*, 27(3), pp.227-254.

Shar, S., & Ingram, J. (2010). *Pharyngealization in Assiri Arabic: an acoustic analysis.* Paper presented at the 13th Australasian International Conference on Speech Science and Technology, Melbourne, Australia.

Shariq, M. (2015). Arabic and English consonants: a phonetic and phonological investigation. *Advances in Language and Literary Studies*, 6(6), pp.146-152.

Showalter, C. (2012). *The influence of novel orthographic information on second language word learning: The case of native English speakers learning Arabic* (unpublised master's dissertation). The University of Utah, UT, USA.

Showalter, C., & Hays-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: a novel lexicon study. *Second Language Research*, 29(2), pp.185-200.

Showalter, C., & Hayes-Harb, R. (2015). Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition. *Applied Psycholinguistics*, 36, pp.23-42.

Simon, E, Chambless, D., & Kickhöfel Alves, U. (2010). Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Science*, 32(3), pp.380-394.

Simon, G., Bernard, C., Lalonde, R., & Rebaï, M., (2006). Orthographic transparency and phoneme-grapheme conversion: An ERB study in Arabic and French readers. *Brain Research*, 1104(1), pp.141-152.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P. (2002). Theorizing and updating aptitude. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning* (pp. 69-64). Amsterdam, The Netherlands: John Benjamins.

Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. Jackson, & Y. Yilmaz (Eds), *Cognitive individual differences in L2 processing and acquisition* (pp.15-38). Amsterdam: John Benjamins.

Skidmore, L., & Gutkin, A. (2020). *Does a priori phonological knowledge improve cross-lingual robustness of phonemic contrasts?* Paper presented at the 22nd International Conference on

Speech and Computer, London, United Kingdom.

Sparks, R., Patton, J., Gangschow, L., & Humbach, N. (2011). Subcomponents of second-language aptitude and second-language proficiency. *The Modern Language Journal*, 95(2), pp.253-273.

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of speech, language, and hearing research: JSLHR*, 49(6), pp.1175–1192. https://doi.org/10.1044/1092-4388(2006/085)

Sun, H. Saito, K., & Tierney, A. (2021). A longitudinal investigation of explicit and implicit auditory progressing in L2 segmental and suprasegmental acquisition. *Studies in Second Language Acquisition*, 43(3), pp.551-573.

Sweller, J. (2011). Cognitive load theory. In: J. P. Mestre & B. H. Ross (Eds), *The psychology of learning and motivation: Cognition in education* (pp. 37–76). Elsevier Academic Press.

Tibi, S., Edwards, A., Schatschneider, C, Lombardino, L., Kirby, J., & Salha, S. (2021). IRT analyses of Arabic letter knowledge in kindergarten. *Reading and Writing: An Interdisciplinary Journal*, 34, pp. 791-816.

Tibi, S., Edwards, A., Kim, Y-S, Schatschneider, C, & Boudelaa, S. (2022). The contribution of letter features to Arabic letter knowledge for Arabic-speaking kindergartners. *Scientific Studies of Reading*, 26(5), pp. 357-372.

Tibi, S. & Kirby, J. (2018). Investigating phonological awareness and naming speed as predictors of reading in Arabic. *Scientific Studies in Reading*, 22(1), pp. 70-84.

Tsai, C. H., McConkie, G. W., & Zheng, X. J. (1998). *Lexical parsing by Chinese readers.* Poster session presented at the Advanced Study Institute on Advances in Theoretical Issues and Cognitive Neuroscience Research of the Chinese Language, University of Hong Kong.

Tsukada, K., Cox, F., Hajek, J., & Hirata, Y. (2018). Non-native Japanese learners' perception of consonant length in Japanese and Italian. *Second Language Research*, 34(2), pp.179-200.

Van Leussen, J-W. & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology*, 6(1000).

Vaid, J., Chen, H-C., & Rao, C. (2022). Biscriptal bilingualism differentially affects segmentation of cross-language homophones: evidence from Hindi and English users. *International Journals of Bilingualism*, 26(1), pp. 13-30.

Watson, J. (2002). *The phonology and Morphology of Arabic*. Oxford University Press.

Wayland, R. (2007). The relationship between identification and discrimination in cross-language perception: the case of Korean and Thai. In: O. Bohn, M. J. Munro (Eds), *Language Experience in Second Language Speech Learning: in honor of James Emil Flege* (pp.201-218). John Benjamins Publishing company.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), pp.1-25.

Weir, C. J. (2004). Language testing and validation: An evidence-based approach. Retrieved from: http://ebookcentral.proquest.com

Wen, Z., Bierdroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, 50(1), pp.1-31.

Winke, P., Sydorenko, T., & Gass, S. (2013). Factors influencing the use of captions by foreign language learners: an eye-tracking study. *The Modern Language Journal*, 97(1), pp.254-275.

Young-Scholten, M. (2002). Orthographic input in L2 phonological development. In: P. Burmeister, T. Piske, & A. Rohde, (Eds), *An integrated view of language development: Papers in honour of Henning Wode* (pp.263-279). Trier, Germany: Wissenschaftlicher Verlag Trier.

Zitouni, I. (2014) (Ed.). *Natural Language Processing of Semitic Languages*. Washington, USA: Redmond.

Appendix A: Llama-B Test

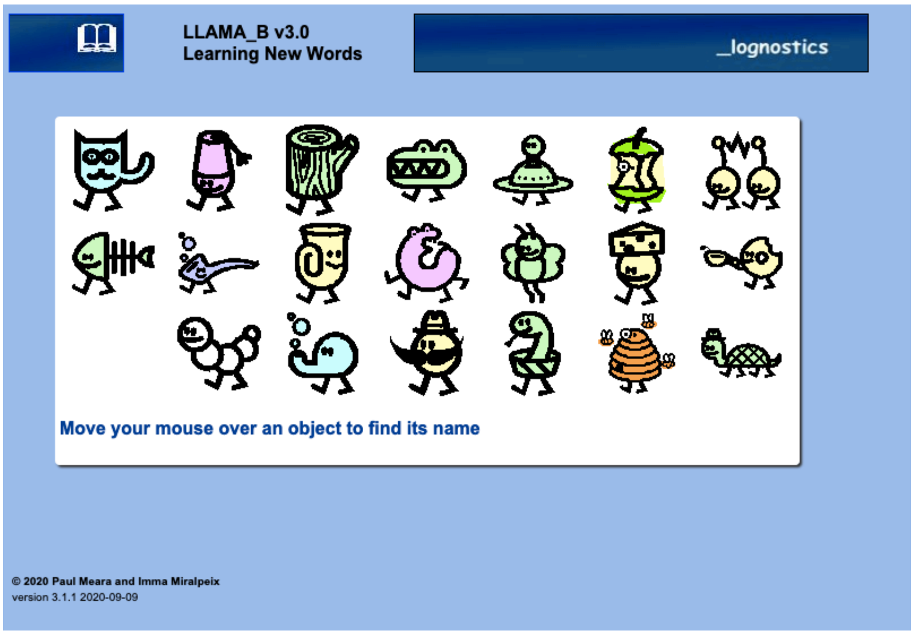


Figure A1 Llama-B learning phase (Meara & Rogers, 2019)

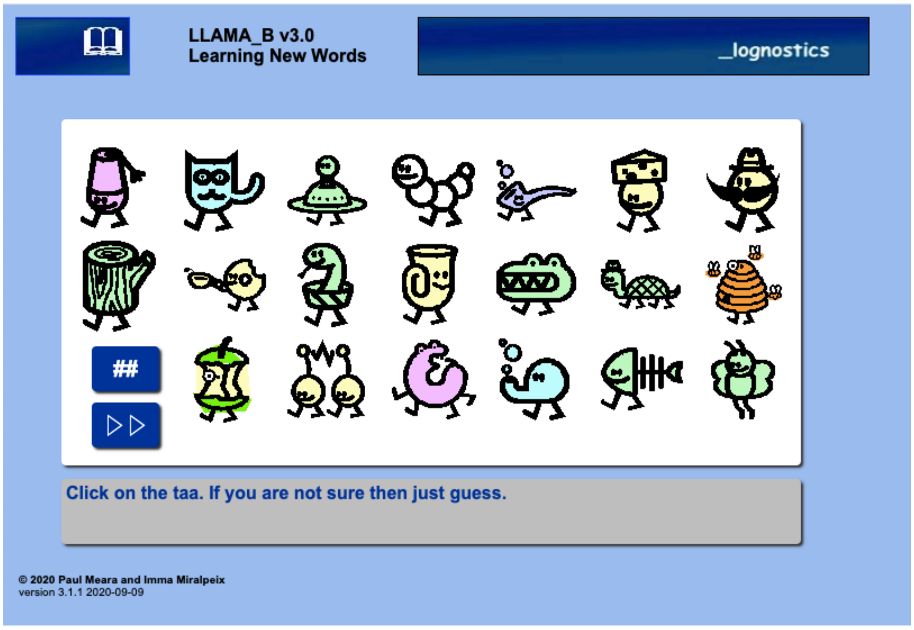


Figure A2 Llama-B testing phase (Meara & Rogers, 2019)

Appendix B: Llama-D Test

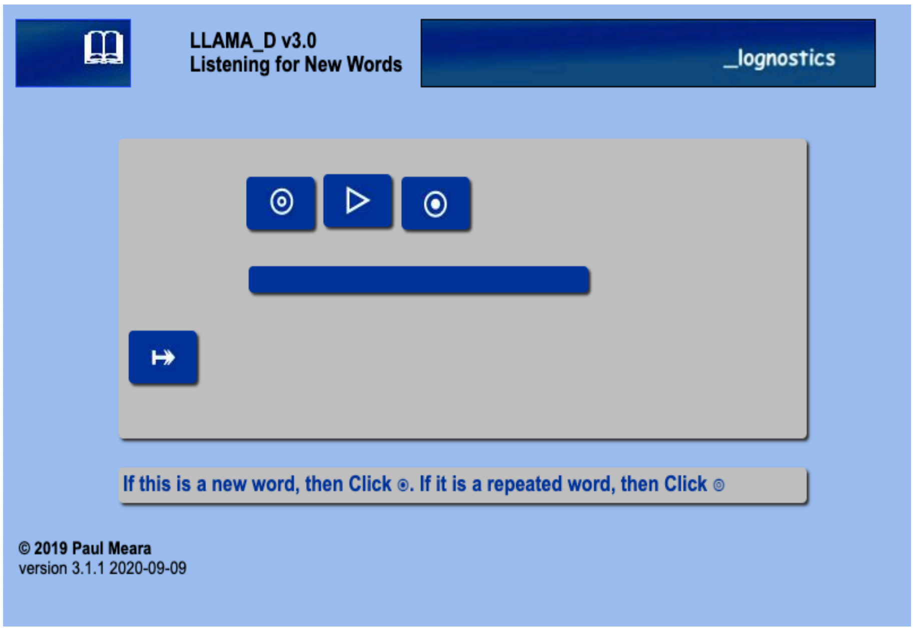


Figure B1 Llama-D testing phase (Meara & Rogers, 2019)

Appendix C: Llama-E Test

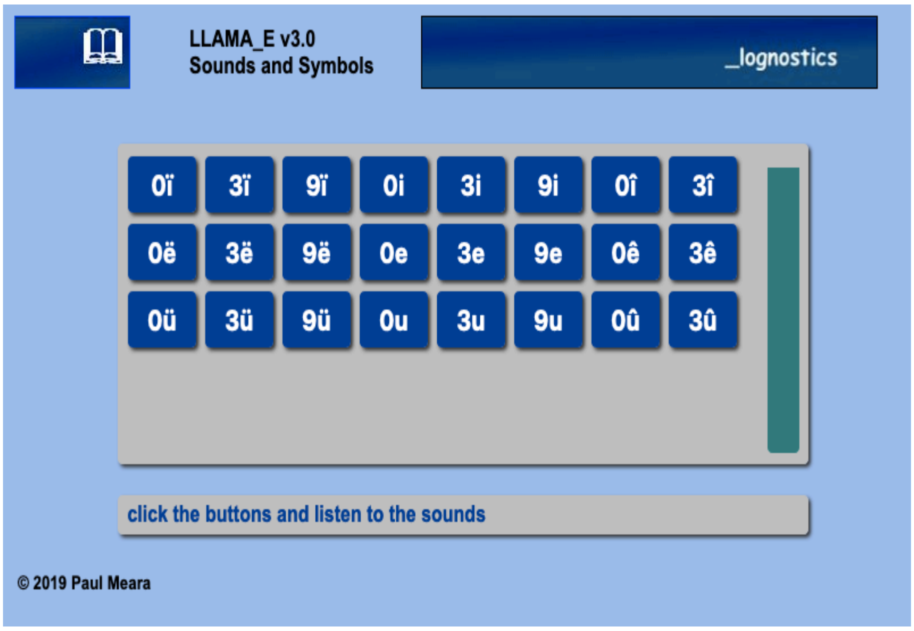


Figure C1 Llama-E learning phase (Meara & Rogers, 2019)

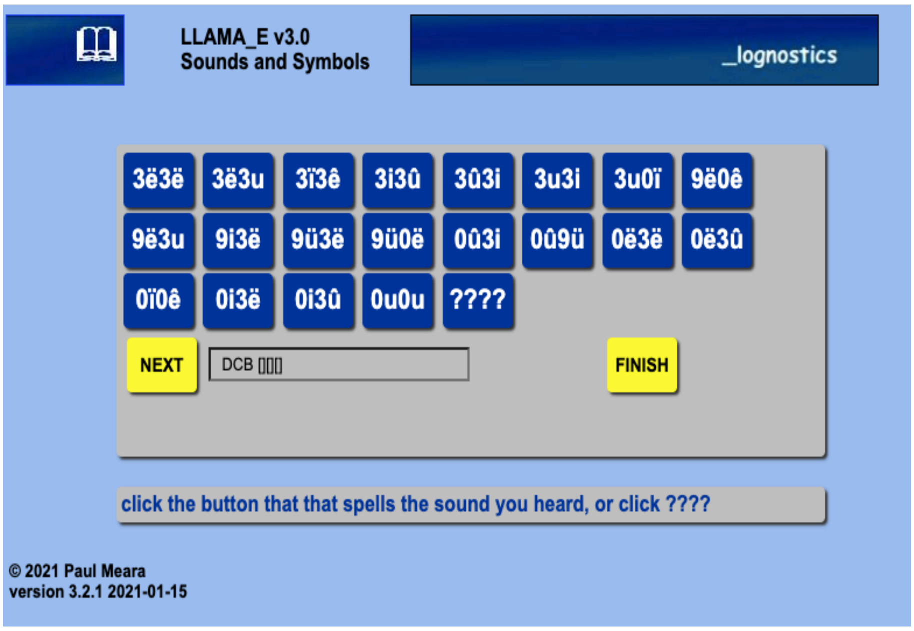


Figure C2 Llama-E testing phase (Meara & Rogers, 2019)

Appendix D: Llama-F Test

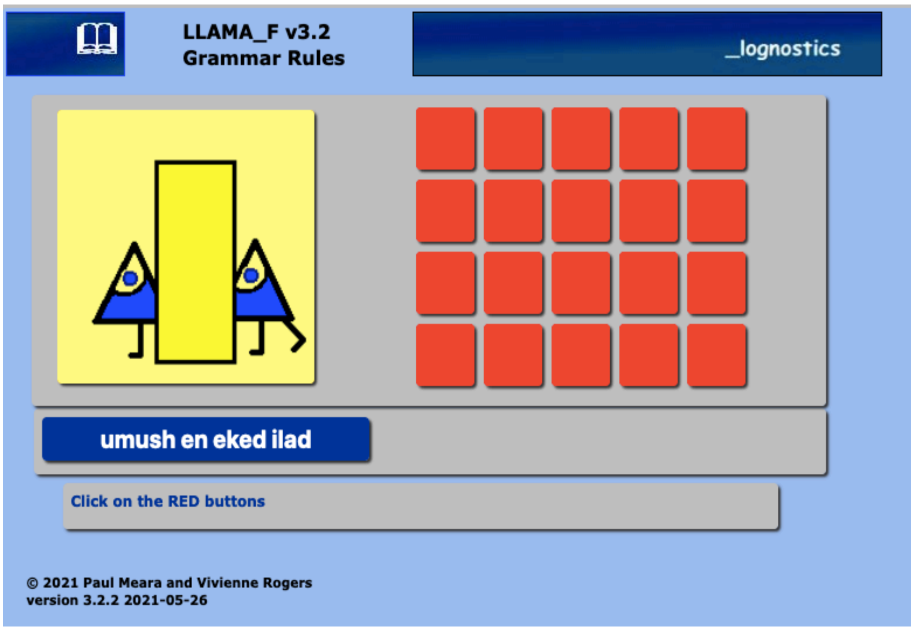


Figure D1 Llama-F learning phase (Meara & Rogers, 2019)

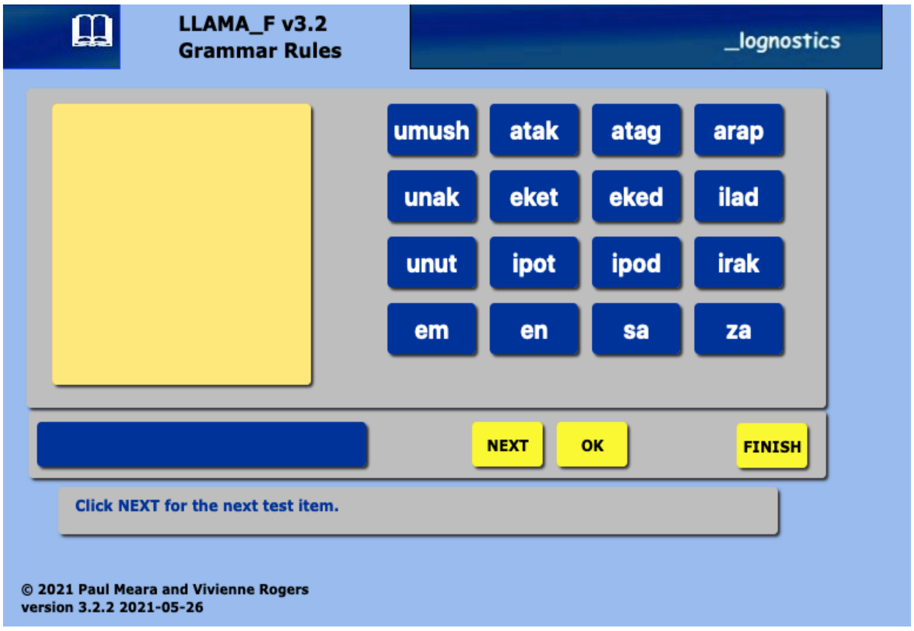


Figure D2 Llama-F testing phase (Meara & Rogers, 2019)

Appendix E: Submission of the Project

School of Literature and Languages
Department of English Language and Applied Linguistics

University of Reading

ETHICS COMMITTEE

Project Submission

Note All sections of this form must be completed.

Principal Investigator (Supervisor): Prof. Jane Setter

Student name: Haifa Alhumaid

Department: Department of English Language and Applied Linguistics

Title of Project: Orthography in L2 Phonology Acquisition: the Role of Similarity between L1 and L2 Sounds and Phonemic Coding Ability

Proposed starting date: 15/07/2020

Number of participants that you require consent from (approximate): 75 participants

I confirm that to the best of my knowledge the Ethics Committee have been made aware of all relevant information. I undertake to inform the Committee of any such information which subsequently becomes available whether before or after the research has begun.

I confirm that a list of the names and contact details of the participants in this project will be compiled and that this, together with signed Consent Forms, will be retained by the researcher under secure storage. All (or in large sample cases a selection) of the signed copies will be submitted with a copy of the dissertation.

Signed:

(Supervisor) Date: 17/06/2020

(Student) Date: 15/06/2020

Appendix F: Project Description

ETHICS COMMITTEE

University of Reading

Project Description

The study aims to investigate the role of how exposure to orthographic input may or may not influence the ability to distinguish two different L2 phonemes. The similarity between the phonemes in L1 and L2 will be considered as a variable in this study. Therefore, the L2 phonemes are divided into three pairs in terms of similarity with L1 phonemes; identical, similar, and new phonemes. Phonemic coding ability also will be investigated as a variable in this study in which participants' scores in Llama E will be correlated with their performance in the experiment. This will be done by examining L2 Arabic phonological acquisition by L1 English participants. The study will be conducted using experimental design and will follow a quantitative method approach. It has two stages: an aptitude pre-test stage (Llama-E) that examines the participants' phonemic coding ability and an experimental stage that examines the influence of the availability of orthography. The study includes six Arabic fricatives that occur initially in Arabic: /ð/ and /θ/, /s/ and /sˤ/, and /ʁ/ and /χ/.

Participants will be divided based on their Llama-E test scores into Arabic-script group, Arabic Romanized-script group, and no-script group. In the second stage, participants will be presented with different types of input depending on which group they belong to. An experiment, adopted from Showalter and Hayes-Harb (2015) with some adjustments, will be conducted. The experiment is divided into five phases: pre-sound recognition phase, word learning phase, criterion test phase, final test phase, and post-sound recognition phase. In the first phase, pairs of Arabic phonemes and minimal pairs will be presented to participants and participants' task is to decide whether they are the same or different. In the second phase, the Arabic-scrip group will be provided with Arabic-like non-words in audio and written forms as well as the picture of the word, the Romanized-script group will have the same except that the written form will be Romanized, whereas the no-script group will listen to the word and see its picture without any script. In the third phase, participants will be exposed to these non-words orally along with the pictures in random order. There will be no scripts. They will have to respond either with 'Yes' or 'No' to the presented picture. In the fourth phase, the same procedure will be repeated except that it will involve a minimal pair that has similar phonemes and participants will be examined in their discrimination between these two minimal pairs by clicking on either 'Yes' or 'No' buttons. The last phase will be a repetition of the first one.

For confidentiality purposes, participants will be assigned codes instead of their real names in order to keep their data records anonymous. This information will be saved to the researcher's personal computer protected by password to which no one else has access. Participants will give informed consent and their data will be deleted if they wish to withdraw from the project.

Appendix G: Information Sheet

Researcher:
Haifa Alhumaid
Email: h.s.alhumaid@pgr.reading.ac.uk

Supervisor:
Prof. Jane Setter
Phone: 0118 3786089
Email: j.e.setter@reading.ac.uk

Department of Englis
School of Literature and Languages Edith Morley Building
PO Box 218
Whiteknights Reading RG6 6AA

Phone 01183788141

+44 (0)118 378 6472 +44 (0)118 975 6506
Email appling@reading.ac.uk
p.a.thompson@reading.ac.uk

INFORMATION SHEET

The main aim of the study is to find out whether access to written forms affects the ability to perceive second language (L2) phonemes, in this case, Arabic phonemes.

There are three sessions in this the research that will take place on three different days. Session one lasts for 40 minutes and begins with an aptitude pre-test (LLAMA-E), which takes under 10 minutes to complete. Participants will then take part in a 30-minute sequence of activities involving Arabic written and audio forms, including computer-based same-different matching exercises and an Arabic word-learning task. Participants will undertake similar activities in the second and third sessions but with different sounds. The second and third sessions are each 30 minutes in length.

For confidentiality purposes, participants will be assigned codes instead of their real names in order to keep their data records anonymous. This information will be saved to the researcher's personal computer protected by a password to which no one else has access. It will be used for only academic purposes and for the researcher's PhD research and will be destroyed after successful examination of the thesis. Only the researcher and the researcher's supervisors will have access to the data.

Participants will be provided, if they so wish, with research results after finishing the project. However, if you wish to withdraw at any stage, please feel free to contact my supervisor at the address above or by email at [j.e.setter@reading.ac.uk]

This project has been subject to ethical review by the School Ethics Committee, and has been allowed to proceed under the exceptions procedure as outlined in paragraph 6 of the University's *Notes for Guidance* on research ethics.

If you have any queries or wish to clarify anything about the study, please feel free to contact my supervisor at the address above or by email at [j.e.setter@reading.ac.uk].

Signed

Haifa Alhumaid

Appendix H: Consent Form

School of Literature and Languages
Department of English Language and Applied Linguistics

University of Reading

ETHICS COMMITTEE

Consent Form

Project title: Orthography in L2 Phonology Acquisition: the Role of Similarity between L1 and L2 Sounds and Phonemic Coding Ability

I understand the purpose of this research and understand what is required of me; I have read and understood the Information Sheet relating to this project, which has been explained to me by Haifa Alhumaid agree to the arrangements described in the Information Sheet in so far as they relate to my participation.

I understand that my participation is entirely voluntary and that I have the right to withdraw from the project at any time.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:

Signed:

Date:

Appendix I: Demographic Questionnaire

School of Literature and Languages
Department of English Language and Applied Linguistics

University of Reading

Project title: Orthography in L2 Phonology Acquisition: the Role of Phonemic Coding Ability

Researcher: Haifa Alhumaid

Supervisor: Prof. Jane Setter & Dr. Daisy Powell

Contact information:

Name: ....

Email Address: ....

Phone number: ....

Please answer the following questions to the best of your knowledge:

Part I: General information

Age: ....

Gender: ....

Education: ....

Part II: Linguistic background

A) What is your first language?

....

B) What Language(s) do you usually use at home? Please specify?

....

C) Have you ever been exposed to Arabic language (written or spoken)?

....

If your answer to the previous question is YES, please answer the following question. If your answer is NO, please skip it

D) How do you rate your proficiency in Arabic language, Please follow the following scale in specifying your level

1.Very poor	2.Poor	3.Functional	4.Good	5.Very good	6.Native-like

Reading proficiency	Writing proficiency	Listening proficiency	Speaking proficiency

E) Have you ever learned a foreign language in the past? If yes please specify?

....

....

If your answer to the previous question is YES, please answer the following two questions. If your answer is NO, please skip them

F) How did you learn your foreign language(s)?

a. Instructional environment (e.g. attending formal language classes)

b. Naturalistic environment (e.g. watching T.V, or practicing by chatting with friends)

If you learned more than one foreign language, please add how you learned them

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

G) What age were you when you learned this language, and did you grow up speaking it?

…………………………………………………………………………………………………

If you learned more than one foreign language, please add what age were you when learned them and whether you grew up speaking them

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

H) How is your proficiency level in the foreign language(s)? Please follow the following scale in specifying your level

1.Very poor	2.Poor	3.Functional	4.Good	5.Very good	6.Native-like

Language	Reading proficiency	Writing proficiency	Listening proficiency	Speaking proficiency

End of the questionnaire
Thank you ☺

Appendix J: Pre-Sound Recognition Phase

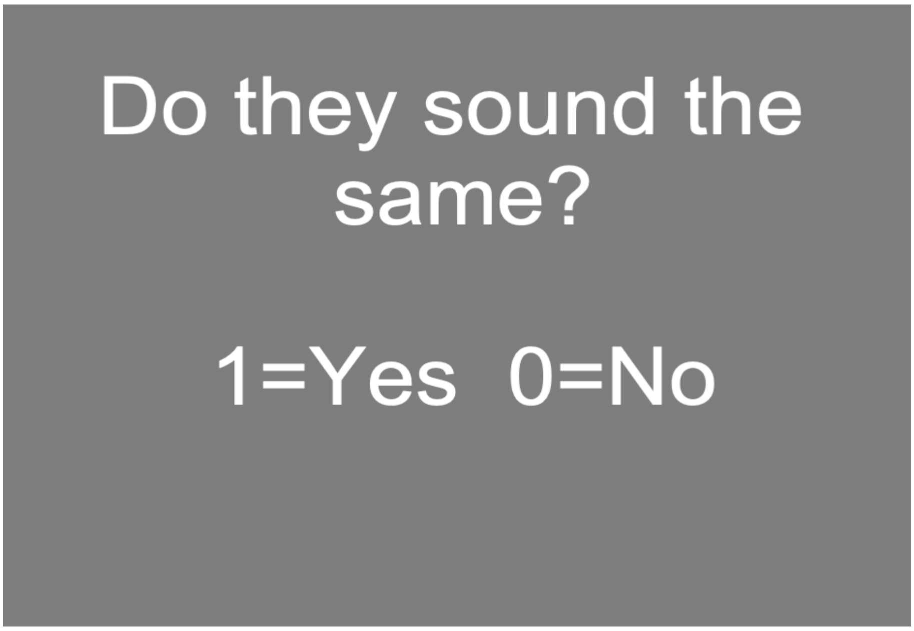


Figure J1 Pre-Sound Recognition Phase: Task 1

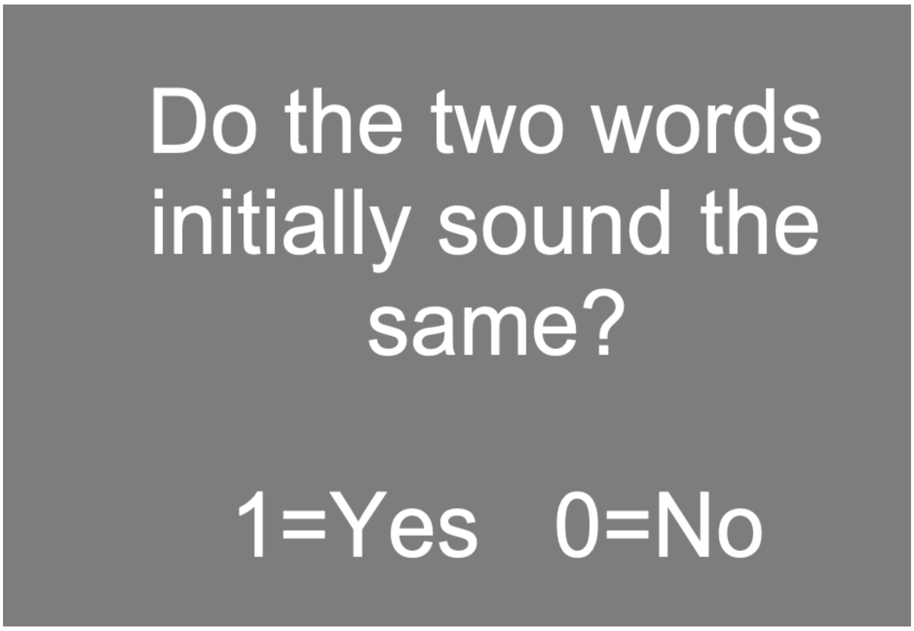


Figure J2 Pre-Sound Recognition Phase: Task 2

Appendix K: Word Learning Phase

Figure K1 Word Learning Phase: Arabic-script group

Figure K2 Word Learning Phase: Romanised-script group

Figure K3 Word Learning Phase: No-script group

Appendix L: Criterion Test Phase

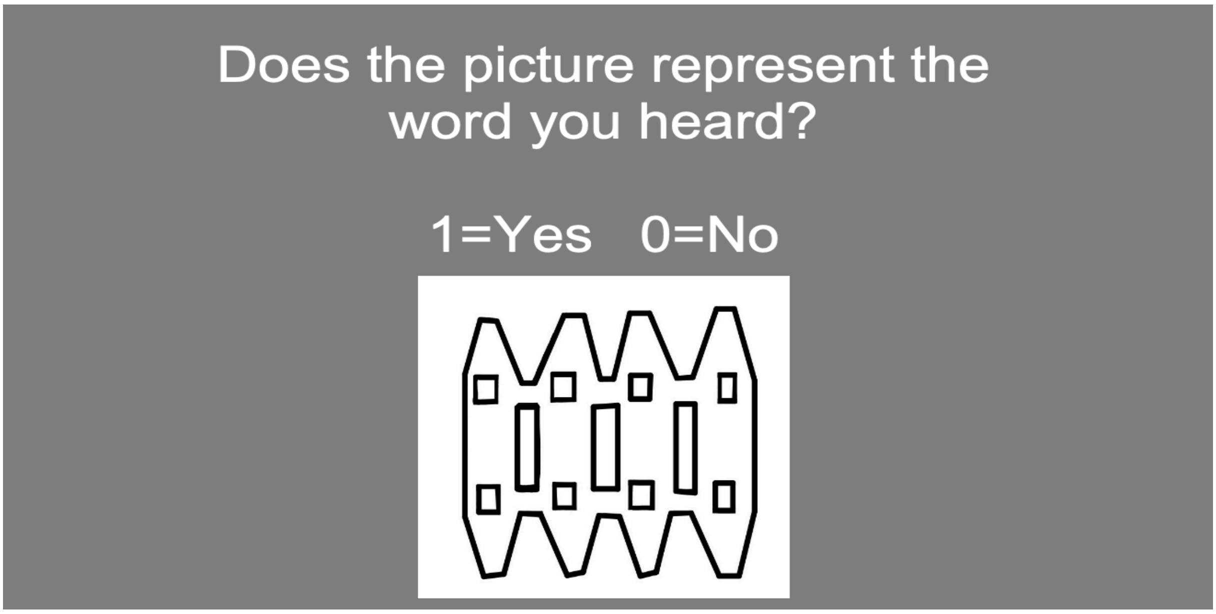


Figure L1 Criterion Test Phase: Yes – No Identification Task

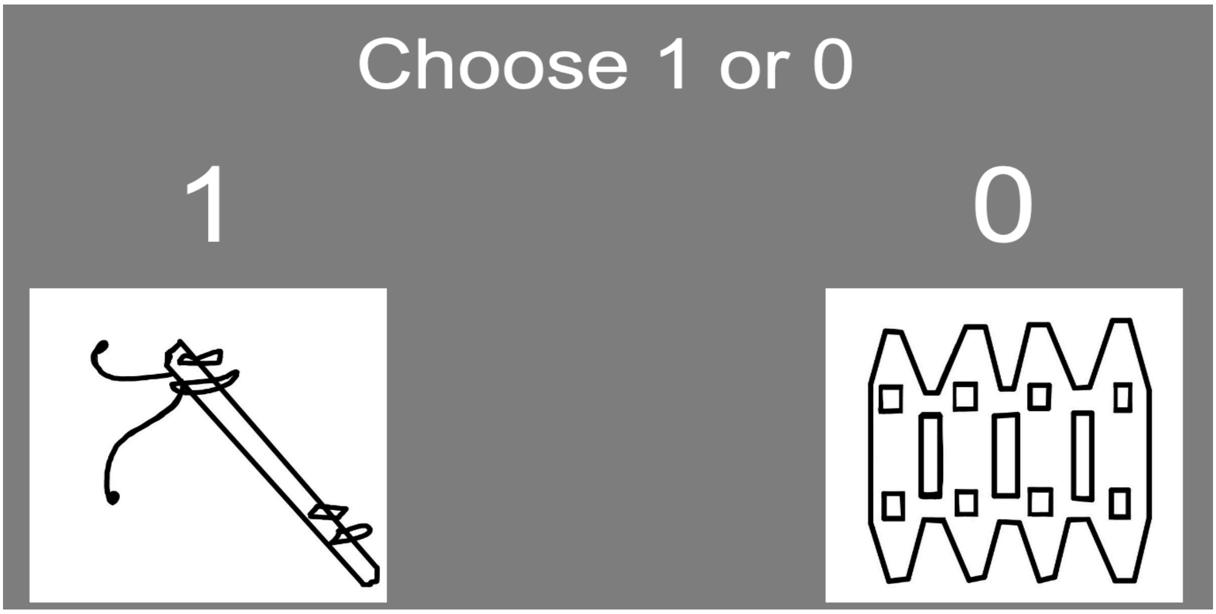


Figure L2 Criterion Test Phase: Forced Choice Identification

Appendix M: Phonological Test Phase

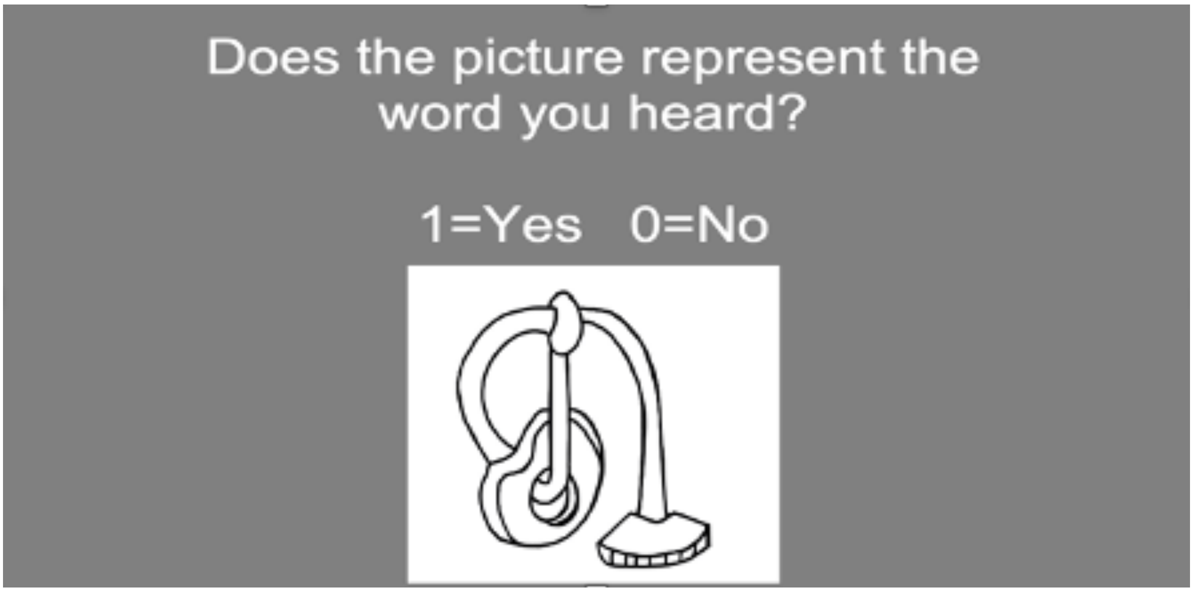


Figure M1 Phonological Test Phase: Yes – No Identification Task

Appendix N: Data Set

Table N1: *Data Set of Arabic-Script Group*

Participant's code	HSH01	HSH03	HSH06	HSH09	HSH13	HSH24	HSH25	HSH28	HSH31
Phonemic coding score	14	9	1	7	16	19	1	2	17
honeme similarity conditio	dentica	dentica	dentica	dentica	dentica	dentica	dentica	dentica	dentica
Pre-Sound Recgnition T1	11	12	11	12	12	12	12	11	12
Pre-Sound Recgnition T1-RT	1.181	1.462	2.174	0.903	0.919	1.078	0.489	1.589	0.632
Pre-Sound Recgnition T2	11	12	12	12	12	12	12	10	12
Pre-Sound Recgnition T2-RT	1.631	0.742	1.052	0.564	0.78	1.333	0.431	1.414	1.009
Criterion Test T1	6	11	9	11	10	11	12	7	7
Criterion Test T1- RT	0.406	0.546	1.106	0.869	0.801	2.336	0.61	2.702	1.879
Criterion Test T2	11	12	10	12	12	10	12	9	10
Criterion Test T2-RT	1.095	1.107	1.421	1.064	1.523	3.454	0.523	2.671	2.235
Phonological Test	11	11	8	10	11	10	10	6	8
Phonological Test-RT	1.807	2.675	1.735	2.193	3.166	2.309	2.872	3.193	2.679
Post-Sound Recgnition	12	12	12	12	12	12	12	10	12
Post-Sound Recgnition-RT	1.826	1.517	1.407	1.196	1.71	2.093	1.055	1.018	0.564
honeme similarity conditio	Similar	Similar	Similar	Similar	Similar	Similar	Similar	Similar	Similar
Pre-Sound Recgnition T1	12	11	7	10	11	11	9	12	9
Pre-Sound Recgnition T1-RT	0.432	0.593	0.281	0.534	0.613	0.398	1.022	0.309	0.365
Pre-Sound Recgnition T2	9	9	11	10	11	8	6	12	10
Pre-Sound Recgnition T2-RT	0.551	0.452	0.654	0.278	0.472	0.239	0.506	0.146	0.48
Criterion Test T1	12	7	10	11	9	11	11	11	12
Criterion Test T1- RT	0.531	0.482	0.62	0.823	0.597	0.624	0.312	0.5	0.452
Criterion Test T2	12	9	11	11	12	12	12	10	11
Criterion Test T2-RT	0.44	0.362	0.198	0.374	0.477	0.3	0.517	0.235	0.39
Phonological Test	6	11	9	11	6	5	7	10	12
Phonological Test-RT	1.297	2.025	2.678	2.121	2.362	2.2	3.111	2.964	2.989
Post-Sound Recgnition	12	12	12	10	12	10	8	12	10
Post-Sound Recgnition-RT	0.849	1.508	1.05	0.56	1.169	0.319	2.349	0.818	0.977
honeme similarity conditio	New	New	New	New	New	New	New	New	New
Pre-Sound Recgnition T1	12	12	12	11	12	12	12	12	12
Pre-Sound Recgnition T1-RT	0.278	0.582	0.955	0.552	0.559	0.733	0.831	2.025	0.572
Pre-Sound Recgnition T2	11	10	12	11	11	11	9	11	11
Pre-Sound Recgnition T2-RT	0.633	0.515	0.604	0.592	0.503	0.951	0.321	2.41	0.94
Criterion Test T1	12	10	10	11	10	12	12	6	11
Criterion Test T1-RT	0.491	0.468	0.371	0.428	0.598	0.951	0.468	3.112	2.191
Criterion Test T2	11	11	11	11	12	11	12	8	11
Criterion Test T2-RT	0.218	0.695	0.825	0.318	0.521	2.755	0.602	2.77	2.684
Phonological Test	10	11	12	10	11	12	12	7	12
Phonological Test-RT	1.533	0.427	0.384	2.158	3.36	2.255	2.862	4.169	3.099
Post-Sound Recgnition	12	12	12	12	12	12	12	12	12
Post-Sound Recgnition-RT	1.181	1.286	1.704	0.603	0.832	1.32	0.796	0.739	0.552

Table N1 Cont.: *Data Set of Arabic-Script Group*

Participant's code	HSH35	HSH38	HSH41	HSH45	HSH46	HSH47	HSH48	HSH52	HSH59
Phonemic coding score	18	19	13	0	9	7	19	20	17
honeme similarity conditic	dentica	dentica	dentica	dentica	dentica	dentica	dentica	dentica	dentica
Pre-Sound Recgnition T1	10	11	12	12	12	12	12	12	11
Pre-Sound Recgnition T1-RT	1.398	1.978	1.243	1.721	2.107	1.972	1.561	1.59	1.632
Pre-Sound Recgnition T2	10	11	12	12	12	12	12	11	11
Pre-Sound Recgnition T2-RT	1.494	2.126	1.314	1.499	3.061	1.742	1.904	1.49	1.938
Criterion Test T1	8	11	9	7	7	9	10	7	9
Criterion Test T1- RT	2.758	3.002	2.192	2.938	2.07	2.788	1.945	2.24	2.133
Criterion Test T2	12	12	7	6	9	9	11	11	10
Criterion Test T2-RT	3.371	2.544	2.135	2.081	2.097	2.332	2.118	2.309	2.111
Phonological Test	6	12	7	5	8	9	6	11	8
Phonological Test-RT	2.644	2.303	3.523	2.207	3.371	2.367	3.15	2.734	0.815
Post-Sound Recgnition	12	12	10	12	12	12	12	12	12
Post-Sound Recgnition-RT	1.174	1.699	0.73	1.503	2.023	1.34	1.427	0.879	0.752
honeme similarity conditic	Similar	Similar	Similar	Similar	Similar	Similar	Similar	Similar	Similar
Pre-Sound Recgnition T1	10	11	10	8	9	11	9	11	11
Pre-Sound Recgnition T1-RT	0.625	1.017	1.187	2.411	0.334	0.551	0.127	0.782	0.568
Pre-Sound Recgnition T2	11	11	11	11	11	10	10	10	10
Pre-Sound Recgnition T2-RT	0.361	0.84	0.652	1.185	0.394	0.62	0.714	0.248	0.377
Criterion Test T1	12	11	10	12	9	10	9	12	12
Criterion Test T1- RT	0.561	0.733	0.84	1.334	0.33	0.787	0.498	0.23	0.163
Criterion Test T2	12	12	9	11	11	11	5	12	12
Criterion Test T2-RT	0.412	0.44	0.645	1.445	0.225	1.058	1.535	0.271	0.383
Phonological Test	5	11	8	6	8	10	6	10	9
Phonological Test-RT	1.368	2.384	2.064	1.375	2.253	1.932	2.586	2.055	2.67
Post-Sound Recgnition	8	12	8	8	8	8	10	12	8
Post-Sound Recgnition-RT	1.271	0.98	0.808	1.366	0.808	0.562	0.097	0.483	0.628
honeme similarity conditic	New	New	New	New	New	New	New	New	New
Pre-Sound Recgnition T1	12	12	12	11	12	12	9	12	12
Pre-Sound Recgnition T1-RT	1.038	1.579	0.846	1.5	1.773	1.801	1.383	1.103	1.446
Pre-Sound Recgnition T2	11	10	12	12	10	11	11	12	11
Pre-Sound Recgnition T2-RT	1.208	1.412	0.969	1.764	0.936	1.586	1.653	1.386	1.524
Criterion Test T1	11	10	12	11	11	11	2	12	12
Criterion Test T1-RT	3.099	2.831	2.689	3.25	1.906	2.508	4.673	2.573	2.262
Criterion Test T2	11	12	12	11	11	12	3	12	12
Criterion Test T2-RT	4.005	2.428	2.582	3.647	2.114	2.383	2.68	2.637	2.322
Phonological Test	11	12	12	12	9	10	7	12	11
Phonological Test-RT	2.217	3.259	3.975	3.535	2.015	2.144	2.932	2.333	0.647
Post-Sound Recgnition	12	12	12	12	12	12	10	12	12
Post-Sound Recgnition-RT	0.783	1.592	0.874	1.202	1.686	1.293	1.41	0.674	1.205

Table N1 Cont.: *Data Set of Arabic-Script Group*

Participant's code	HSH61	HSH65	HSH70	HSH72	HSH80	HSH84	HSH90
Phonemic coding score	3	17	3	17	17	7	19
honeme similarity conditio	dentica	dentica	dentica	dentica	dentica	dentica	dentica
Pre-Sound Recgnition T1	12	11	11	11	12	11	12
Pre-Sound Recgnition T1-RT	2.548	1.769	1.336	1.925	1.917	2.024	1.619
Pre-Sound Recgnition T2	7	11	9	12	12	11	12
Pre-Sound Recgnition T2-RT	1.86	1.832	1.425	1.596	1.496	2.389	1.473
Criterion Test T1	9	5	10	6	9	9	7
Criterion Test T1- RT	2.865	2.826	2.26	2.884	2.7	2.807	1.981
Criterion Test T2	7	8	10	12	12	5	7
Criterion Test T2-RT	3.087	3.07	2.267	2.642	2.488	2.474	1.928
Phonological Test	6	7	7	10	10	5	7
Phonological Test-RT	0.592	0.966	1.35	1.173	0.85	2.755	2.983
Post-Sound Recgnition	12	12	12	12	12	12	12
Post-Sound Recgnition-RT	0.553	0.492	0.939	0.363	1.115	1.53	1.466
honeme similarity conditio	Similar	Similar	Similar	Similar	Similar	Similar	Similar
Pre-Sound Recgnition T1	9	10	11	10	10	11	6
Pre-Sound Recgnition T1-RT	0.162	0.896	0.187	0.155	0.317	2.076	0.678
Pre-Sound Recgnition T2	11	12	10	10	10	10	12
Pre-Sound Recgnition T2-RT	0.204	0.919	0.309	0.249	0.201	1.063	0.697
Criterion Test T1	9	12	10	8	8	8	9
Criterion Test T1- RT	0.709	0.306	0.251	0.238	0.414	0.911	0.501
Criterion Test T2	12	12	11	10	12	12	12
Criterion Test T2-RT	0.244	1.119	0.172	0.329	0.397	0.659	0.436
Phonological Test	8	9	5	5	6	6	3
Phonological Test-RT	0.537	0.902	0.253	0.58	1.731	2.521	3.973
Post-Sound Recgnition	10	8	6	10	10	12	10
Post-Sound Recgnition-RT	0.95	1.819	0.31	0.53	0.425	1.234	0.874
honeme similarity conditio	New	New	New	New	New	New	New
Pre-Sound Recgnition T1	12	12	12	12	12	12	12
Pre-Sound Recgnition T1-RT	1.462	1.19	1.145	1.235	1.554	1.33	1.528
Pre-Sound Recgnition T2	12	11	8	11	11	10	10
Pre-Sound Recgnition T2-RT	1.433	1.381	1.194	1.481	1.158	1.452	1.593
Criterion Test T1	11	12	6	4	12	7	10
Criterion Test T1-RT	2.455	1.905	2.542	2.262	2.743	3.243	2.8
Criterion Test T2	10	12	5	6	12	10	12
Criterion Test T2-RT	2.711	3.152	3.155	3.475	2.374	3.769	2.695
Phonological Test	8	11	9	3	11	7	12
Phonological Test-RT	0.813	0.631	0.778	1.829	0.431	2.972	2.683
Post-Sound Recgnition	12	12	12	12	12	12	12
Post-Sound Recgnition-RT	0.253	0.487	0.609	0.38	1.488	1.247	1.674

Table N2: *Data Set of Romanised-Script Group*

Participant's code	HSH02	HSH05	HSH12	HSH04	HSH17	HSH20	HSH21	HSH30	HSH32
Phonemic coding score	6	6	2	15	14	2	0	18	2
oneme similarity conditi	dentica	dentica	dentica	dentica	dentica	dentica	dentica	dentica	dentica
Pre-Sound Recgnition T1	12	10	7	9	11	12	12	12	12
re-Sound Recgnition T1-R	2.61	1.22	2.08	1.32	0.63	1.57	0.46	1.13	1.86
Pre-Sound Recgnition T2	12	11	12	11	11	12	12	12	11
re-Sound Recgnition T2-R	2	0.78	1.13	0.6	0.69	1.44	0.98	1.23	1.66
Criterion Test T1	6	6	6	5	11	7	11	5	3
Criterion Test T1- RT	3.83	1.06	1.72	2.51	1.03	0.55	0.81	2.68	2.13
Criterion Test T2	9	7	10	9	12	5	10	4	5
Criterion Test T2-RT	3.96	1.28	1.25	1.64	1.23	2.01	1.11	3.05	2.12
Phonological Test	6	5	8	4	12	5	5	2	5
Phonological Test-RT	4.22	1.31	0.62	1.31	1.13	0.25	1.03	2.11	2.59
Post-Sound Recgnition	12	12	12	10	10	12	12	12	12
Post-Sound Recgnition-RT	1.53	0.45	0.94	0.96	0.51	0.55	0.72	0.88	1.95
oneme similarity conditi	Similar	Similar	Similar	Similar	Similar	Similar	Similar	Similar	Similar
Pre-Sound Recgnition T1	6	10	11	10	11	10	10	9	11
re-Sound Recgnition T1-R	0.58	0.59	0.57	1.13	0.34	0.8	0.73	1.27	1.18
Pre-Sound Recgnition T2	11	11	9	7	10	11	12	11	7
re-Sound Recgnition T2-R	1.52	0.47	0.39	0.62	0.63	0.82	0.34	1.39	2.27
Criterion Test T1	9	7	12	8	12	4	12	5	10
Criterion Test T1- RT	2.5	0.81	0.52	1.56	0.82	0.28	0.93	1.84	1.93
Criterion Test T2	9	6	12	6	12	9	12	9	8
Criterion Test T2-RT	2.4	0.18	0.5	1.41	0.36	0.76	1.39	1.58	2.13
Phonological Test	8	6	4	3	9	5	7	7	7
Phonological Test-RT	3.18	0.35	0.33	1.75	0.26	0.6	1.37	2.39	2.23
Post-Sound Recgnition	12	8	10	10	6	12	10	10	10
Post-Sound Recgnition-RT	1.01	0.65	0.56	1.45	0.35	0.6	0.64	0.57	0.82
oneme similarity conditi	New	New	New	New	New	New	New	New	New
Pre-Sound Recgnition T1	12	12	12	6	11	12	12	12	12
re-Sound Recgnition T1-R	1.09	0.5	0.53	1.2	0.91	0.22	0.39	0.64	1.75
Pre-Sound Recgnition T2	12	12	10	9	10	10	12	10	12
re-Sound Recgnition T2-R	1.26	0.42	0.71	0.4	1.18	0.77	0.34	1.15	1.54
Criterion Test T1	9	11	5	8	12	11	12	3	9
Criterion Test T1-RT	3.32	0.34	0.54	1.41	0.28	0.38	0.65	1.42	2.63
Criterion Test T2	9	9	7	10	12	12	11	8	10
Criterion Test T2-RT	4.03	0.11	0.7	1.08	0.31	0.35	0.65	2.48	2.83
Phonological Test	9	9	10	7	11	11	12	5	7
Phonological Test-RT	3.13	1.04	0.51	1.64	0.69	0.4	0.82	1.86	2.34
Post-Sound Recgnition	12	12	12	6	12	12	12	12	12
Post-Sound Recgnition-RT	1.25	0.31	0.61	1.04	0.12	0.25	0.63	0.67	1.42

Table N2 Cont.: *Data Set of Romanised-Script Group*

Participant's code	HSH33	HSH36	HSH39	HSH42	HSH43	HSH51	HSH53	HSH54	HSH5(
Phonemic coding score	9	19	16	1	2	17	18	13	11
ioneme similarity conditi	denticа	denticа	denticа	denticа	denticа	denticа	denticа	denticа	dentic
Pre-Sound Recgnition T1	12	12	12	12	10	12	12	12	12
're-Sound Recgnition T1-R	2.52	2.33	1.78	2.13	1.52	2.98	0.95	0.97	1.17
Pre-Sound Recgnition T2	9	11	12	10	12	12	12	11	11
're-Sound Recgnition T2-R	1.9	1.51	2.93	1.81	1.68	2.1	1.26	1.28	1.3
Criterion Test T1	3	9	6	8	4	8	7	9	10
Criterion Test T1- RT	2.13	3.33	2.29	3.92	3.36	2.46	3.66	3.02	2.77
Criterion Test T2	8	10	7	7	6	3	11	12	12
Criterion Test T2-RT	1.92	2.39	2.15	3.64	2.51	0.32	3.38	2.58	3.46
Phonological Test	7	10	6	6	5	9	9	5	5
Phonological Test-RT	1.69	3.25	2.23	3.67	2.66	2.86	3.54	4.12	3.44
Post-Sound Recgnition	12	12	12	12	12	12	12	12	10
Post-Sound Recgnition-RT	1.57	1.44	1.5	2.04	1.59	2.61	0.56	1.63	1.14
ioneme similarity conditi	Similar	Similar	Similar	Similar	Similar	Similar	Similar	Similar	Simila
Pre-Sound Recgnition T1	11	12	10	10	10	10	10	9	10
're-Sound Recgnition T1-R	0.98	1.12	0.82	1.12	1.06	1.24	0.92	0.97	0.6
Pre-Sound Recgnition T2	9	12	11	10	10	10	12	7	9
're-Sound Recgnition T2-R	2.13	2.08	1.73	1.88	1.45	1.23	1.31	1.28	1.95
Criterion Test T1	12	12	12	8	11	12	10	12	9
Criterion Test T1- RT	1.92	2.22	1.95	3.99	2.85	1.87	2.82	3.02	2.56
Criterion Test T2	12	12	12	11	10	8	12	12	11
Criterion Test T2-RT	1.67	2.38	1.64	3.86	3.2	2.36	2.83	2.58	3.05
Phonological Test	6	9	11	6	6	6	8	7	8
Phonological Test-RT	1.95	2.94	1.73	3.44	2.14	1.53	3.11	4.12	3.32
Post-Sound Recgnition	12	12	12	10	8	8	10	8	8
Post-Sound Recgnition-RT	0.96	0.81	0.71	0.83	0.78	1.27	0.6	0.84	0.4
ioneme similarity conditi	New	New	New	New	New	New	New	New	New
Pre-Sound Recgnition T1	12	12	12	11	12	10	12	12	10
're-Sound Recgnition T1-R	1.48	1.06	1.61	1.51	1.6	2.44	0.52	1.28	0.96
Pre-Sound Recgnition T2	12	12	12	11	11	9	11	8	8
're-Sound Recgnition T2-R	1.64	1.21	1.22	1.2	1.81	1.31	0.99	1.66	0.95
Criterion Test T1	10	12	11	9	12	9	11	10	12
Criterion Test T1-RT	3.35	1.88	2.51	4.27	3.07	2.51	2.33	3.45	3.29
Criterion Test T2	6	12	12	9	10	5	12	10	11
Criterion Test T2-RT	2.59	1.73	2.36	4.56	3.27	2.35	2.38	2.9	3.09
Phonological Test	6	12	11	4	10	5	8	6	12
Phonological Test-RT	2.61	2.4	2.38	4.46	3.26	2	2.67	3.65	3.12
Post-Sound Recgnition	6	12	12	10	10	10	12	10	12
Post-Sound Recgnition-RT	1.29	1.14	1.2	1.24	1.36	2.08	0.55	1.37	0.8

Table N2 Cont.: *Data Set of Romanised-Script Group*

Participant's code	HSH58	HSH64	HSH69	HSH77	HSH78	HSH86	HSH8
Phonemic coding score	18	1	2	7	0	15	12
oneme similarity conditi	dentica	dentica	dentica	dentica	dentica	dentica	dentic
Pre-Sound Recgnition T1	11	12	12	12	12	12	11
're-Sound Recgnition T1-R	1.07	1.68	1.59	1.51	1.71	1.41	1.8
Pre-Sound Recgnition T2	12	12	11	12	10	12	12
're-Sound Recgnition T2-R	1.33	1.5	1.56	1.49	1.62	1.68	1.56
Criterion Test T1	4	10	0	8	8	2	8
Criterion Test T1- RT	1.94	3.31	0	3.05	2.09	2.79	3.23
Criterion Test T2	5	12	1	12	6	6	9
Criterion Test T2-RT	2.33	3.43	4.07	2.14	2.37	1.86	3.38
Phonological Test	5	6	3	12	6	6	9
Phonological Test-RT	1.81	3.84	2.88	2.22	1.98	1.97	3.15
Post-Sound Recgnition	10	12	12	12	12	10	12
Post-Sound Recgnition-RT	0.79	1.51	1.71	1.2	1.41	1.32	1.48
oneme similarity conditi	Similar	Similar	Similar	Similar	Similar	Similar	Simila
Pre-Sound Recgnition T1	10	11	10	11	10	11	10
're-Sound Recgnition T1-R	0.86	1.43	1.09	0.56	0.94	0.66	1.03
Pre-Sound Recgnition T2	11	10	11	11	9	10	10
're-Sound Recgnition T2-R	1.58	1.85	2	1.33	1.39	1.46	1.67
Criterion Test T1	12	9	12	6	4	9	11
Criterion Test T1- RT	2.51	3.46	3.12	1.93	1.31	1.59	2.91
Criterion Test T2	10	11	12	9	3	11	11
Criterion Test T2-RT	2.56	2.76	2.85	2.03	0.59	1.27	3.17
Phonological Test	7	6	7	6	7	6	8
Phonological Test-RT	2.09	3.11	3.4	1.99	0.26	0.94	4.09
Post-Sound Recgnition	8	10	8	8	6	10	12
Post-Sound Recgnition-RT	0.32	0.87	1.28	0.55	2.02	1.22	0.81
oneme similarity conditi	New	New	New	New	New	New	New
Pre-Sound Recgnition T1	12	11	12	12	8	9	11
're-Sound Recgnition T1-R	1.25	1.58	1.07	0.72	1.39	1.51	1.19
Pre-Sound Recgnition T2	12	12	9	12	10	11	10
're-Sound Recgnition T2-R	1.27	1.26	1.01	0.88	2.24	1.4	1.55
Criterion Test T1	7	6	5	12	3	8	12
Criterion Test T1-RT	2.98	2.38	2.43	2.51	2.95	1.84	3.55
Criterion Test T2	6	8	6	9	8	8	5
Criterion Test T2-RT	2.62	2.45	2.8	2.61	2.94	1.99	4.04
Phonological Test	6	5	6	10	8	6	12
Phonological Test-RT	2.34	2.79	2.48	2.06	2.34	1.54	3.25
Post-Sound Recgnition	12	12	10	12	12	12	12
Post-Sound Recgnition-RT	0.76	1.05	0.91	0.6	1.49	1.36	1.14

Table N3: *Data Set of No-Script Group*

Participant's code	HSH07	HSH15	HSH18	HSH19	HSH23	HSH27	HSH29	HSH34	HSH37
Phonemic coding score	3	18	8	16	2	11	11	15	5
honeme similarity condition	Identical	Identical	Identical	Identical	Identical	Identical	Identical	Identical	Identic
Pre-Sound Recgnition T1	12	12	12	4	12	12	12	12	10
Pre-Sound Recgnition T1-RT	0.676	1.09	1.28	2.049	0.973	1.68	1.291	1.52	1.701
Pre-Sound Recgnition T2	9	12	12	6	11	10	12	12	11
Pre-Sound Recgnition T2-RT	0.745	0.708	0.921	1.224	1.408	1.368	1.468	1.484	1.514
Criterion Test T1	11	12	12	3	12	9	11	7	12
Criterion Test T1- RT	2.099	0.596	1.104	1.248	0.799	2.64	2.553	2.301	3.243
Criterion Test T2	12	11	12	6	9	12	10	8	12
Criterion Test T2-RT	1.271	0.48	0.674	2.87	0.768	2.708	2.424	2.246	2.304
Phonological Test	11	12	9	8	12	4	7	8	11
Phonological Test-RT	0.827	0.394	1.163	2.145	0.587	2.317	2.912	2.451	2.238
Post-Sound Recgnition	10	12	12	8	12	12	12	10	10
Post-Sound Recgnition-RT	0.202	0.363	1.54	1.158	0.892	0.752	1.148	1.404	1.273
honeme similarity condition	Similar	Similar	Similar	Similar	Similar	Similar	Similar	Similar	Simila
Pre-Sound Recgnition T1	10	10	11	2	11	10	12	10	9
Pre-Sound Recgnition T1-RT	0.666	0.399	0.737	1.444	0.687	0.947	0.738	0.918	1.1
Pre-Sound Recgnition T2	10	11	10	9	10	9	11	8	10
Pre-Sound Recgnition T2-RT	0.436	0.83	1.45	1.32	1.124	1.62	1.697	1.467	1.615
Criterion Test T1	12	9	11	8	9	10	10	10	12
Criterion Test T1- RT	0.974	0.579	0.654	0.613	1.071	1.943	2.091	1.807	1.718
Criterion Test T2	12	12	12	11	12	12	11	12	12
Criterion Test T2-RT	1.069	1.258	0.683	0.593	0.504	2.529	2.056	1.797	1.384
Phonological Test	7	7	12	11	5	6	5	7	7
Phonological Test-RT	0.471	0.4	1.199	0.588	0.352	3.018	1.597	2.626	1.831
Post-Sound Recgnition	8	8	10	10	12	12	12	10	10
Post-Sound Recgnition-RT	0.417	0.964	1.242	1.575	0.979	1.406	0.483	0.894	0.6
honeme similarity condition	New	New	New	New	New	New	New	New	New
Pre-Sound Recgnition T1	12	2	12	12	12	11	12	12	12
Pre-Sound Recgnition T1-RT	0.67	2.761	0.949	0.924	1.126	1.698	1.076	1.224	1.547
Pre-Sound Recgnition T2	9	7	12	11	10	10	10	11	11
Pre-Sound Recgnition T2-RT	0.355	0.567	1.327	1.976	0.738	1.223	1.262	1.123	1.507
Criterion Test T1	12	11	12	10	12	12	11	12	11
Criterion Test T1-RT	0.492	0.823	0.776	0.669	0.965	2.239	2.341	1.946	3.088
Criterion Test T2	12	12	12	12	12	11	10	12	12
Criterion Test T2-RT	0.647	1.321	0.927	0.304	0.599	3.13	2.519	2.145	2.228
Phonological Test	12	12	10	9	12	12	7	10	7
Phonological Test-RT	0.459	0.942	0.707	0.525	0.767	2.933	2.38	1.893	2.09
Post-Sound Recgnition	12	12	12	12	12	10	12	12	12
Post-Sound Recgnition-RT	0.181	0.469	1.283	0.487	0.965	1.207	1.036	1.333	0.979

Table N3 Cont.: *Data Set of No-Script Group*

Participant's code	HSH40	HSH44	HSH49	HSH50	HSH55	HSH63	HSH66	HSH67	HSH68
Phonemic coding score	20	15	12	18	5	10	18	18	6
Phoneme similarity condition	Identical	Identical	Identical	Identical	Identical	Identical	Identical	Identical	Identical
Pre-Sound Recgnition T1	12	12	12	12	12	11	12	12	11
Pre-Sound Recgnition T1-RT	2.548	2	1.7	1.993	2.511	1.85	1.26	1.384	2.057
Pre-Sound Recgnition T2	11	12	12	10	11	10	12	12	12
Pre-Sound Recgnition T2-RT	2.44	1.898	1.734	1.732	1.627	2.116	1.158	1.791	1.972
Criterion Test T1	7	11	11	10	8	3	12	10	7
Criterion Test T1- RT	3.684	3.403	2.345	2.761	2.811	2.348	2.703	2.664	3.492
Criterion Test T2	9	10	12	9	11	6	11	10	6
Criterion Test T2-RT	3.447	2.835	2.034	2.884	2.626	2.285	2.173	2.604	3.538
Phonological Test	8	11	10	8	7	6	12	9	7
Phonological Test-RT	3.403	2.352	2.49	3.009	2.53	2.074	1.986	2.327	3.557
Post-Sound Recgnition	12	10	12	12	12	12	12	10	12
Post-Sound Recgnition-RT	1.69	1.236	0.742	1.755	1.174	0.985	0.992	2.351	2.093
Phoneme similarity condition	Similar	Similar	Similar	Similar	Similar	Similar	Similar	Similar	Similar
Pre-Sound Recgnition T1	10	10	9	11	9	10	10	8	12
Pre-Sound Recgnition T1-RT	1.791	0.902	0.406	1.189	1.588	1.355	0.815	0.877	1.373
Pre-Sound Recgnition T2	11	11	10	10	12	8	10	10	12
Pre-Sound Recgnition T2-RT	2.376	1.855	1.268	1.688	1.593	1.869	1.273	2.446	1.835
Criterion Test T1	12	8	11	10	9	7	12	8	12
Criterion Test T1- RT	2.093	2.587	1.901	2.107	1.437	1.987	2.39	2.49	2.362
Criterion Test T2	11	12	12	10	10	9	11	12	12
Criterion Test T2-RT	2.278	2.471	2.117	2.107	1.337	2.024	1.891	2.773	2.364
Phonological Test	7	8	11	8	7	6	10	4	12
Phonological Test-RT	1.985	2.748	1.579	1.841	1.144	1.315	2.391	2.96	1.918
Post-Sound Recgnition	8	12	8	10	12	10	8	12	12
Post-Sound Recgnition-RT	1.534	1.326	0.284	0.969	0.778	1.16	0.416	2.104	0.999
Phoneme similarity condition	New	New	New	New	New	New	New	New	New
Pre-Sound Recgnition T1	12	12	11	12	12	12	12	12	12
Pre-Sound Recgnition T1-RT	1.874	1.409	0.669	1.4	1.067	2.246	0.786	1.9	1.645
Pre-Sound Recgnition T2	11	11	10	10	12	7	12	11	11
Pre-Sound Recgnition T2-RT	1.69	1.411	1.029	1.69	0.974	2.011	0.974	1.693	1.474
Criterion Test T1	11	12	12	12	10	6	12	11	10
Criterion Test T1-RT	2.681	3.073	1.923	2.206	1.804	2.664	1.96	2.674	2.944
Criterion Test T2	12	12	12	12	11	7	12	8	6
Criterion Test T2-RT	3.443	3.961	2.224	2.02	2.354	3.483	1.897	2.787	3.223
Phonological Test	12	12	12	11	9	11	12	6	7
Phonological Test-RT	3.141	3.614	2.197	1.973	2.175	3.332	1.527	2.847	2.036
Post-Sound Recgnition	12	10	12	12	12	10	12	12	12
Post-Sound Recgnition-RT	1.686	1.569	0.874	1.426	0.783	1.448	0.739	1.906	1.32

Table N3 Cont.: *Data Set of No-Script Group*

Participant's code	HSH74	HSH75	HSH76	HSH81	HSH83	HSH85	HSH87
Phonemic coding score	13	10	6	18	4	0	16
Phoneme similarity condition	Identical	Identical	Identical	Identical	Identical	Identical	Identica
Pre-Sound Recgnition T1	12	11	12	12	12	12	12
Pre-Sound Recgnition T1-RT	1.565	2.263	0.951	2.896	2.158	1.962	1.735
Pre-Sound Recgnition T2	11	9	10	10	11	10	12
Pre-Sound Recgnition T2-RT	2.309	2.059	1.406	1.854	1.88	2.282	2.031
Criterion Test T1	5	8	11	9	12	10	10
Criterion Test T1- RT	2.155	2.838	2.484	2.687	3.347	2.832	2.635
Criterion Test T2	10	9	11	11	10	10	12
Criterion Test T2-RT	1.993	3.432	2.333	3.258	2.574	3.579	3.162
Phonological Test	5	5	9	11	8	10	11
Phonological Test-RT	3.603	3.166	2.184	3.235	2.603	2.801	2.597
Post-Sound Recgnition	10	12	12	12	12	12	10
Post-Sound Recgnition-RT	1.447	1.668	0.932	1.553	1.517	1.826	1.466
Phoneme similarity condition	Similar	Similar	Similar	Similar	Similar	Similar	Similar
Pre-Sound Recgnition T1	10	8	11	12	12	9	10
Pre-Sound Recgnition T1-RT	1.333	0.882	1.319	1.281	1.324	1.123	1.25
Pre-Sound Recgnition T2	11	9	9	10	11	9	11
Pre-Sound Recgnition T2-RT	1.79	1.977	1.609	1.624	2.01	1.694	1.703
Criterion Test T1	12	12	12	11	11	12	10
Criterion Test T1- RT	2.16	2.371	2.31	1.644	1.69	2.173	3.245
Criterion Test T2	11	10	12	12	12	11	11
Criterion Test T2-RT	1.693	2.966	2.217	2.749	1.549	3.392	3.46
Phonological Test	7	8	10	11	9	9	8
Phonological Test-RT	2.55	2.827	2.153	2.626	1.914	3.253	3.608
Post-Sound Recgnition	10	8	12	12	12	8	10
Post-Sound Recgnition-RT	0.956	1.133	0.609	1.05	0.989	1.04	1.163
Phoneme similarity condition	New	New	New	New	New	New	New
Pre-Sound Recgnition T1	12	12	12	12	12	12	12
Pre-Sound Recgnition T1-RT	1.63	1.924	0.904	1.355	1.242	1.434	1.456
Pre-Sound Recgnition T2	9	12	9	11	10	0	12
Pre-Sound Recgnition T2-RT	1.367	1.814	1.054	1.502	1.359	1.274	1.306
Criterion Test T1	12	12	12	12	12	12	11
Criterion Test T1-RT	2.792	2.354	2.34	2.726	2.278	2.686	2.929
Criterion Test T2	12	11	11	12	12	12	11
Criterion Test T2-RT	2.834	3.353	2.123	2.851	2.861	3.188	4.193
Phonological Test	10	12	11	12	12	11	12
Phonological Test-RT	2.417	3.267	2.098	2.435	2.684	3.096	3.633
Post-Sound Recgnition	10	12	12	12	12	12	12
Post-Sound Recgnition-RT	1.411	2.012	1.035	1.704	1.286	1.181	1.674

Appendix O: Data Screening

Exploring the assumption of linearity

Looking at the scatterplots of the Phonological Test Phase and Post-Sound Recognition Phase in Figures O1, O2, and O3 for the Phonological Test Phase; O4, O5, and O6 for the Post-Sound Recognition Phase, there are no outliers and the relationships between the variables are positive, indicating that the assumption of linearity has been met.

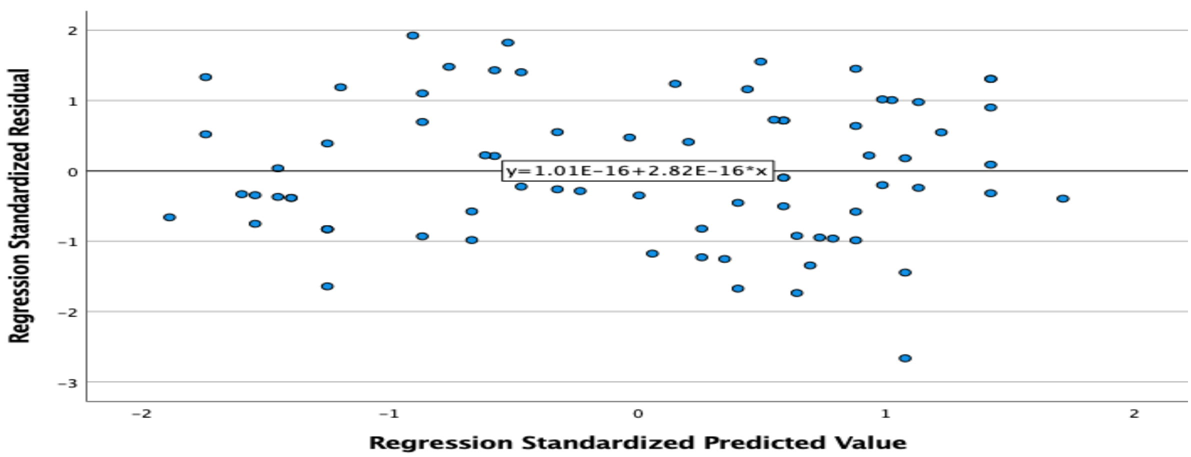


Figure O1 Plot of standarised predicted values against standarised residuals: Phonological Test Phase in Identical-phoneme condition (accuracy)

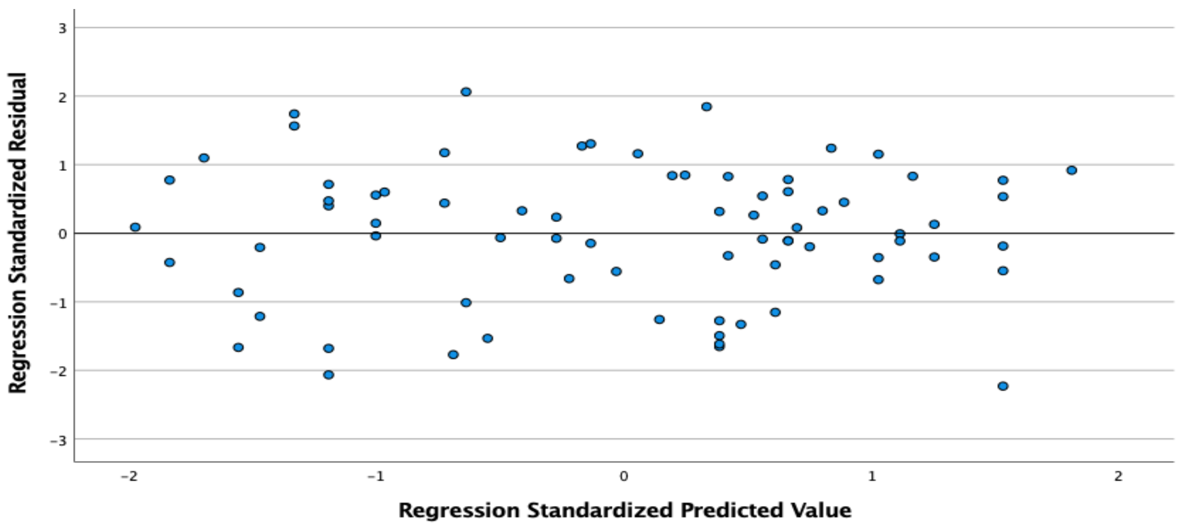


Figure O2 Plot of standarised predicted values against standarised residuals: Phonological Test Phase in Identical-phoneme condition (reaction time)

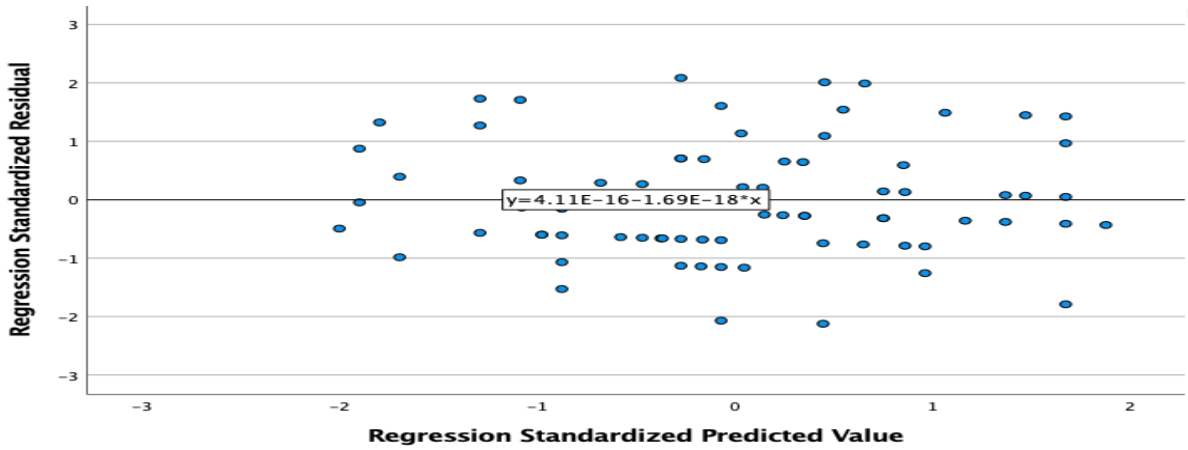


Figure O3 Plot of standarised predicted values against standarised residuals: Phonological Test Phase in Similar-phoneme condition (accuracy)

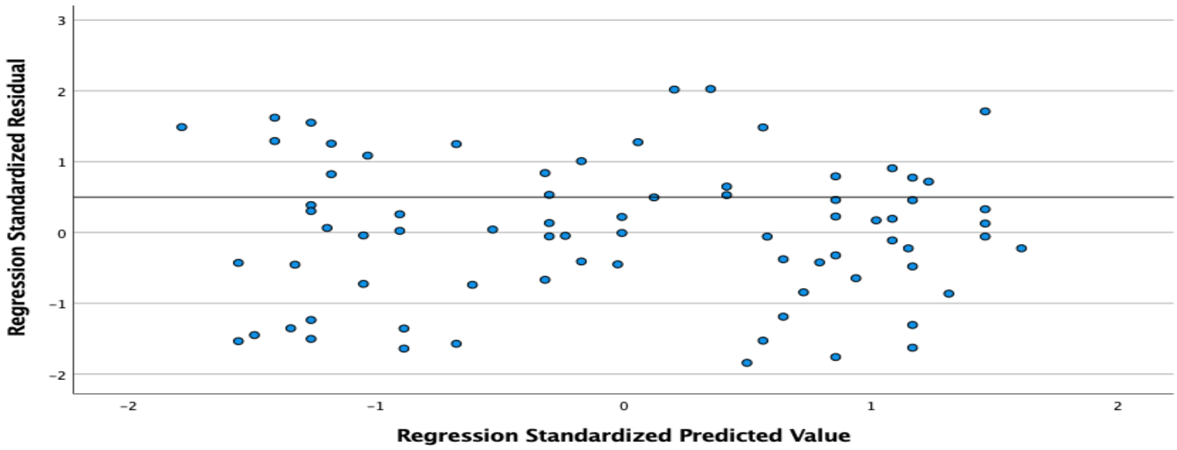


Figure O4 Plot of standarised predicted values against standarised residuals: Phonological Test Phase in Similar-phoneme condition (reaction time)

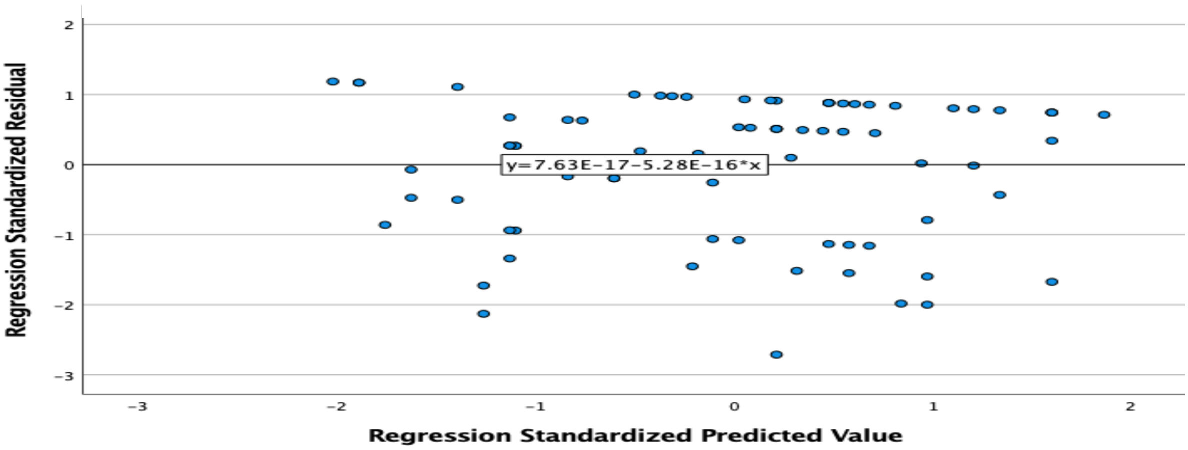


Figure O5 Plot of standarised predicted values against standarised residuals: Phonological Test Phase in New-phoneme condition (accuracy)

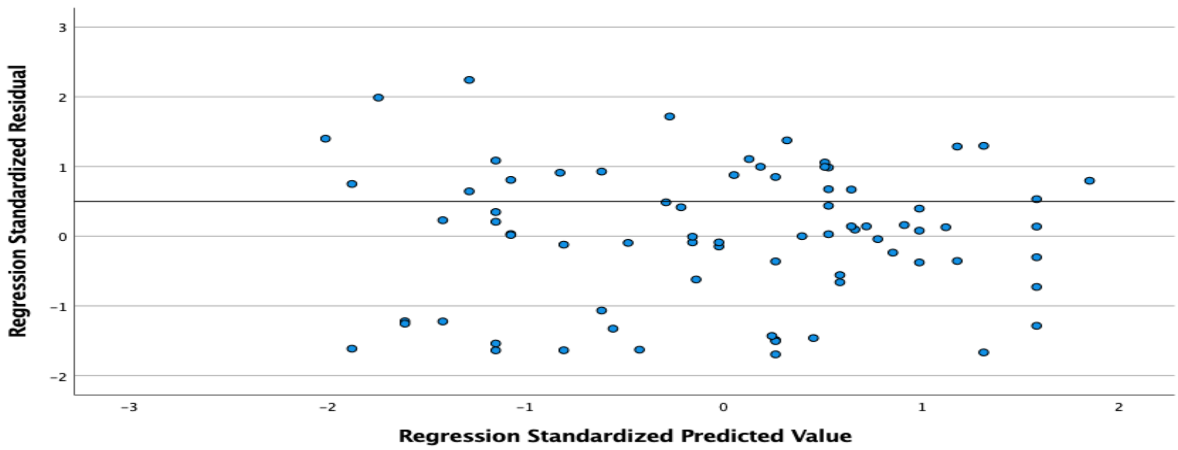


Figure O6 Plot of standarised predicted values against standarised residuals: Phonological Test Phase in New-phoneme condition (reaction time)

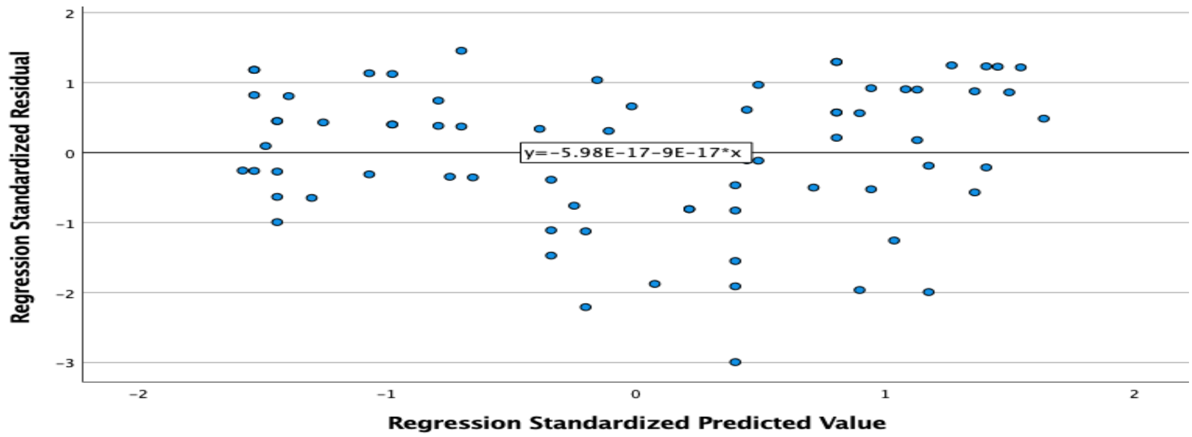


Figure O7 Plot of standarised predicted values against standarised residuals: Post-Sound Recognition Phase in Identical-phoneme condition (accuracy)

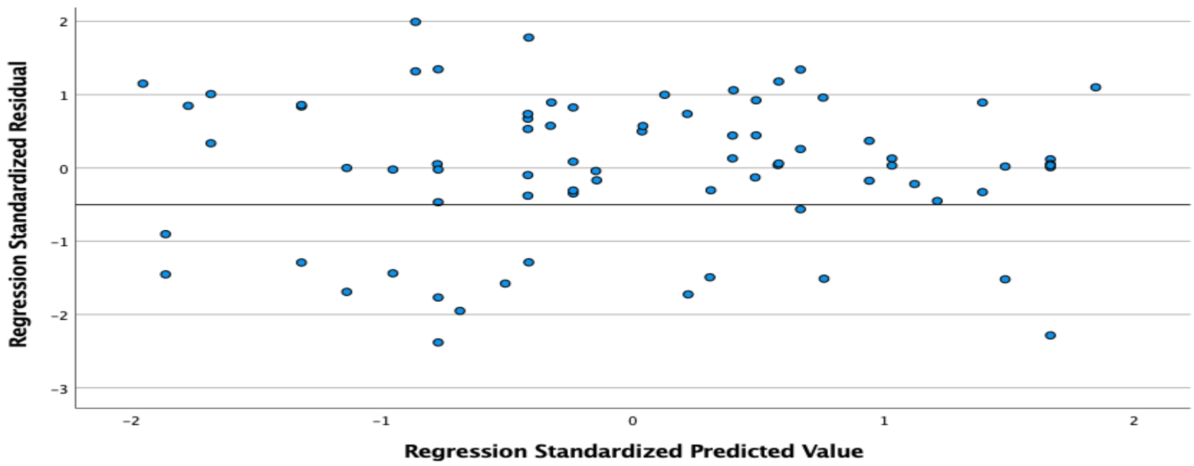


Figure O8 Plot of standarised predicted values against standarised residuals: Post-Sound Recognition Phase in Identical-phoneme condition (reaction time)

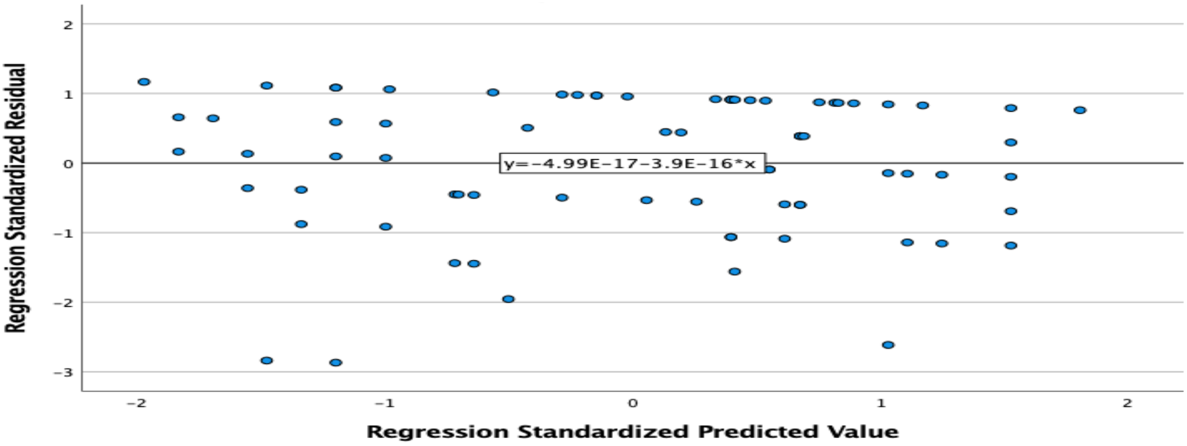


Figure O9 Plot of standarised predicted values against standarised residuals: Post-Sound Recognition Phase in Similar-phoneme condition (accuracy)

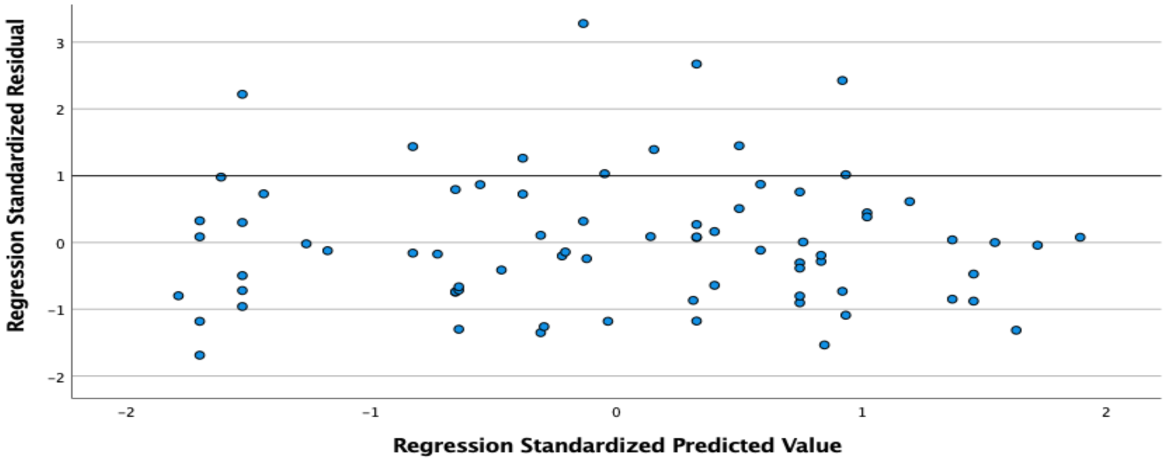


Figure O10 Plot of standarised predicted values against standarised residuals: Post-Sound Recognition Phase in Similar-phoneme condition (reaction time)

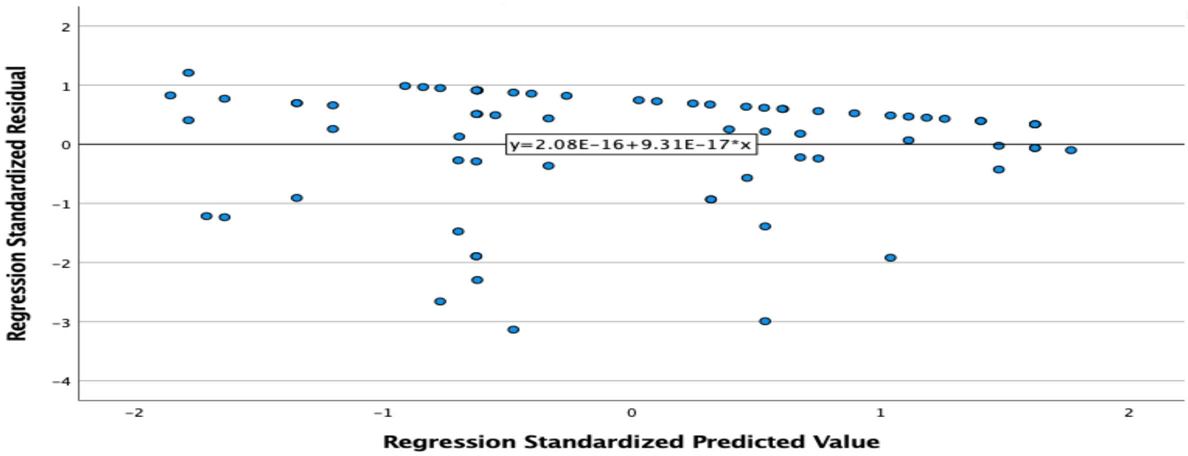


Figure 11 Plot of standarised predicted values against standarised residuals: Post-Sound Recognition Phase in New-phoneme condition (accuracy)

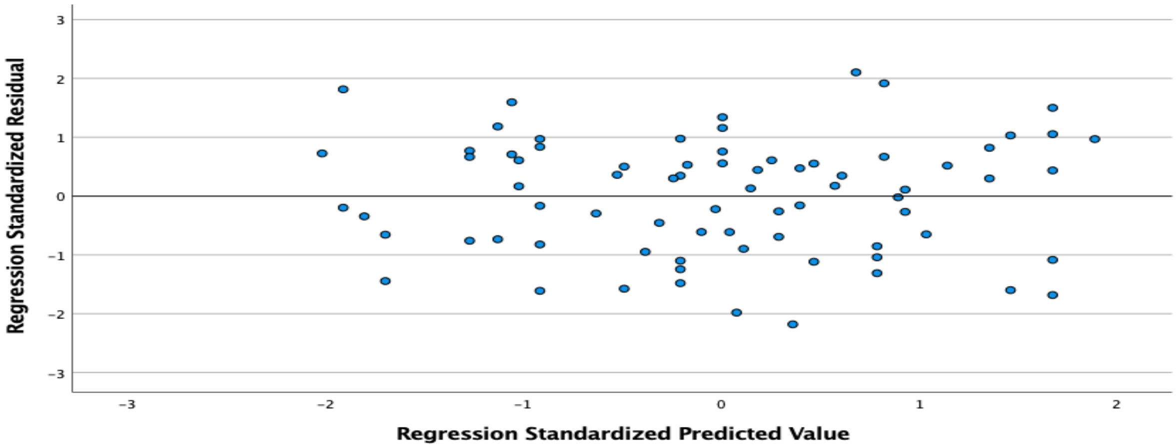


Figure 12 Plot of standarised predicted values against standarised residuals: Post-Sound Recognition Phase in New-phoneme condition (reaction time)

Exploring the normality of the distribution of data

To test whether the data were normally distributed, a Shapiro-Wilk test was performed in all tasks. It is worth noting, however, according to Field (2018), that tests that were designed to test the violation of assumptions – such as the Kolmogorov-Smirnov and Shapiro-Wilk test that test the normality of a distribution and a Levene's test that tests equality of variance and many others – do not always have the power to detect violations of assumptions. Nevertheless, a Shapiro-Wilk test was carried out and, in addition to that, the *z*-scores of both skewness and kurtosis were also calculated to measure the normality of distributions.

Table O1 reports the results of a Shapiro-Wilk normality distribution test on the phonemic coding ability scores in the three orthographic input-type groups as well as the *z*-scores for both skewness and kurtosis. Results of the Shapiro-Wilk test indicate that phonemic coding ability scores were only normally distributed in the No-script group, but not in the other two orthographic input-type groups. However, notwithstanding that the result of Shapiro-Wilk test indicated data not being distributed normally, the skewness *z*-score and the kurtosis *z*-score were less than 1.96, which is the cut off for significance and, therefore, did not appear to be a problem in the three orthographic input-type groups.

Table O1: *Shapiro-Wilk Test of Normality in Phonemic Coding Ability*

Orthography group	Shapiro-Wilk Statistic	df	Sig.	Skewness z-scores	Kurtosis z-score
Arabic	.86	25	.003	-.90	-1.62
Romanised	.87	25	.007	.09	-1.85
No	.93	25	.11	-.50	-1.36

Table O2 shows the results of the Shapiro-Wilk normality distribution test in the Pre-Sound Recognition Phase in the two tasks, as well as the *z*-scores for both skewness and kurtosis. The results indicate that the data were not normally distributed except for one case which is Task 2 in the similar-phonemes condition with the No-script group in accuracy data. This was not the case with reaction time data which were normally distributed in all cases except in Task 1 in the similar-phonemes condition with the Arabic-script group.

Table O2: *Shapiro-Wilk Test of Normality in the Pre-Sound Recognition Phase*

Task and measure	Phoneme type	Orthographic input type	Shapiro-Wilk Statistic	df	Sig.	Skewness z-scores	Kurtosis z-scores
Task 1 (Accuracy)	Identical phonemes	Arabic	.96	25	<.001	-2.01	-.003
		Romanised	.75	25	<.001	-5.35	7.24
		No	.63	25	<.001	-9.34	22.19
	Similar phonemes	Arabic	.88	25	.01	-2.29	1.30
		Romanised	.72	25	<.001	-4.59	9.04
		No	.73	25	<.001	-5.54	10.86
	New phonemes	Arabic	.35	25	<.001	-8.27	17.51
		Romanised	.63	25	<.001	-4.81	5.60
		No	.25	25	<.001	-10.46	26.52
Task 1 (Reaction time)	Identical phonemes	Arabic	.98	25	.89	-.72	-.14
		Romanised	.98	25	.95	.75	-.03
		No	.99	25	.99	.58	-.29
	Similar phonemes	Arabic	.77	25	<.001	5.29	9.24
		Romanised	.97	25	.72	-.34	.75
		No	.97	25	.73	-.08	-.53
	New phonemes	Arabic	.97	25	.73	-.25	-.88
		Romanised	.97	25	.57	.57	.35
		No	.96	25	.33	1.7	1
Task 2 (Accuracy)	Identical phonemes	Arabic	.56	25	<.001	-4.92	6.56
		Romanised	.73	25	<.001	-3.01	1.92
		No	.78	25	<.001	-3.72	4.74
	Similar phonemes	Arabic	.84	25	<.001	-3.17	4.20
		Romanised	.87	25	.006	-1.86	.32
		No	.92	25	.58	-.36	-.41
	New phonemes	Arabic	.84	25	<.001	-2.30	1.84
		Romanised	.84	25	.002	-1.26	-.91
		No	.68	25	<.001	-6.25	11.76
Task 2 (Reaction time)	Identical phonemes	Arabic	.96	25	.40	.95	1.41
		Romanised	.94	25	.17	1.56	2.63
		No	.97	25	.70	-.79	-.42
	Similar phonemes	Arabic	.93	25	.09	1.82	.16
		Romanised	.94	25	.13	-1.04	-.77
		No	.95	25	.20	-1.33	1.91
	New phonemes	Arabic	.96	25	.34	.44	.2
		Romanised	.97	25	.68	.1	.42
		No	.98	25	.79	-.89	.25

Table O3 reports the results of the Shapiro-Wilk normality distribution test in the Criterion Test Phase in two tasks as well as the *z*-scores for both skewness and kurtosis. As shown in Table O3, there were problems with normality, in terms of both skew and kurtosis, and also in terms of normality tests, for some but not all cases.

Table O3: *Shapiro-Wilk Test of Normality in the Criterion Test Phase*

Task and measure	Phoneme type	Orthographic input type	Shapiro-Wilk Statistic	df	Sig.	Skewness z-scores	Kurtosis z-scores
Yes – No (Accuracy)	Identical phonemes	Arabic	.94	25	.19	-.20	-.96
		Romanised	.96	25	.61	-.88	-.29
		No	.85	25	.003	-2.36	.53
	Similar phonemes	Arabic	.89	25	.01	-.78	-1.13
		Romanised	.83	25	<.001	-1.92	-.34
		No	.86	25	.004	-1.27	-.89
	New phonemes	Arabic	.74	25	<.001	-3.65	2.41
		Romanised	.86	25	.004	-1.91	-.25
		No	.59	25	<.001	-6.46	12.23
Yes – No (Reaction time)	Identical phonemes	Arabic	.86	25	.002	-1.79	-.74
		Romanised	.95	25	.26	-1.37	-.32
		No	.9	25	.02	2.04	.68
	Similar phonemes	Arabic	.94	25	.12	2.16	1.96
		Romanised	.98	25	.89	-.08	-.53
		No	.94	25	.13	-1.02	.08
	New phonemes	Arabic	.9	25	.02	-.34	-.45
		Romanised	.93	25	.09	-.96	-.82
		No	.87	25	.005	-1.96	.35
Forced Choice	Identical phonemes	Arabic	.87	25	.006	-1.62	-.48
		Romanised	.94	25	.16	-.66	.76
		No	.86	25	.003	-2.06	.31
	Similar phonemes	Arabic	.64	25	<.001	-5.78	9.77
		Romanised	.81	25	<.001	-3.06	2.11
		No	.74	25	<.001	-2.65	.95
	New phonemes	Arabic	.66	25	<.001	-4.49	4.13
		Romanised	.92	25	.07	-.54	-1.08
		No	.59	25	<.001	-4.74	4.53
Forced Choice (Reaction time)	Identical phonemes	Arabic	.96	25	.43	-.77	-.04
		Romanised	.97	25	.6	-.05	-.064
		No	.91	25	.03	-1.87	.34
	Similar phonemes	Arabic	.76	25	<.001	3.79	2.56
		Romanised	.97	25	.51	-.49	-.89
		No	.98	25	.85	-.19	-.47

	New phonemes	Arabic	.88	25	.008	-1.45	-.63
		Romanised	.94	25	.13	.63	-.41
		No	.96	25	.32	-1.07	-.39

Table O4 reports the results of the Shapiro-Wilk normality distribution test in the Phonological Test Phase, as well as the *z*-scores for both skewness and kurtosis. As shown in Table O4, there were problems with normality, in terms of both skew and kurtosis, and also in terms of normality tests, for some but not all cases in accuracy data. However, data were normally distributed in reaction time data except in one case which is the identical-phonemes condition with the No-script group, but the skewness *z*-score and the kurtosis *z*-score was less than 1.96, and, therefore, did not appear to be a problem.

Table O4: *Shapiro-Wilk Test of Normality in the Phonological Test Phase*

Measure	Phoneme type	Orthographic input type	Shapiro-Wilk Statistic	df	Sig.	Skewness z-scores	Kurtosis z-scores
Accuracy	Identical phonemes	Arabic	.92	25	.08	.05	-1.43
		Romanised	.904	25	.02	1.68	.35
		No	.94	25	.16	-.70	-.85
	Similar phonemes	Arabic	.94	25	.17	.18	-1.14
		Romanised	.94	25	.14	.34	1.63
		No	.94	25	.22	.51	-.77
	New phonemes	Arabic	.79	25	<.001	-3.47	3.08
		Romanised	.91	25	.03	.07	-1.62
		No	.76	25	<.001	-2.56	.18
Reaction time	Identical phonemes	Arabic	.93	25	.08	-1.18	-.94
		Romanised	.98	25	.8	-.22	-.93
		No	.92	25	.04	-1.84	-.51
	Similar phonemes	Arabic	.97	25	.65	-.51	.07
		Romanised	.95	25	.19	-.17	-1.13
		No	.96	25	.47	-.51	-.84
	New phonemes	Arabic	.94	25	.13	-.25	-1.20
		Romanised	.97	25	.61	-.2	-.15
		No	.94	25	.14	-.93	-.75

Table O5 reports the results of the Shapiro-Wilk normality distribution test in the Post-Sound Recognition Phase, as well as the *z*-scores for both skewness and kurtosis. As shown in Table O5, looking at accuracy data, there were problems with normality in terms of normality tests. However, not all cases are problematic in terms of skew and kurtosis. This was not the case with reaction time case, as data were normally distributed except in one condition which is the similar-phonemes condition with the Romanised-script group

Table O5: *Shapiro-Wilk Test of Normality in Post-Sound Recognition Phase*

Measure	Phoneme type	Orthographic input type	Shapiro-Wilk Statistic	df	Sig.	Skewness z-scores	Kurtosis z-scores
Accuracy	Identical phonemes	Arabic	.30	25	<.001	-7.10	10.68
		Romanised	.49	25	<.001	-3.44	.65
		No	.56	25	<.001	-4.01	3.25
	Similar phonemes	Arabic	.85	25	.002	-.41	-1.16
		Romanised	.87	25	.006	-.35	-.86
		No	.78	25	<.001	-.51	1.68
	New phonemes	Arabic	.20	25	<.001	-10.77	27.71
		Romanised	.62	25	<.001	3.60	9.59
		No	.44	25	<.001	-4.25	2.28
Reaction time	Identical phonemes	Arabic	.98	25	.77	-.05	-.89
		Romanised	.96	25	.35	.72	.13
		No	.99	25	.99	-.29	.26
	Similar phonemes	Arabic	.94	25	.19	2.16	1.76
		Romanised	.9	25	.02	2.96	2.95
		No	.96	25	.4	1.04	1
	New phonemes	Arabic	.95	25	.2	-.09	-1.34
		Romanised	.96	25	.46	.34	-.03
		No	.98	25	.93	-.75	-.16

Exploring the equality of variance

As shown in Table O6 and based on the median, as it is less biased by outliers (Field, 2018), Levene's tests showed that the assumption of equality of variance was met in the accuracy data in the Pre-Sound Recognition Phase and one condition of the Post-Sound Recognition Phase as the variances for the three orthographic input types were not significantly different. This was not the case with the Criterion Test Phase and Phonological Test Phase, and the two conditions of the Post-Sound Recognition Phase because in some conditions, as shown in that table, the variances were found to be significantly different, indicating that the assumption of the equality of variance was not met. However, the data of reaction time met the assumption except in three conditions which are Task 2 of the Pre-Sound Recognition Phase and Task 1 and 2 of the Criterion Test Phase in the similar-phoneme condition

Table O6: *Levene's Test Results*

Measure	Phase	Phoneme type	Task 1				Task 2			
			Levene statistic	df 1	df 2	Sig	Levene statistic	df 1	df 2	Sig
Accuracy	Pre-Sound Recognition	Identical	.106	2	72	.98	1.01	2	72	.36
		Similar	1.41	2	72	.24	.66	2	72	.51
		New	1.14	2	72	.32	1.91	2	72	.15
	Criterion Test	Identical	1.62	2	72	.203	5.54	2	72	.006
		Similar	4.22	2	72	.01	3.42	2	72	.03
		New	4.27	2	72	.01	2.01	2	72	.14
	Phonological Test	Identical	.05	2	72	.94				
		Similar	3.64	2	72	.03				
		New	2.56	2	72	.08				
	Post-Sound Recognition	Identical	1.41	2	72	.24				
		Similar	.04	2	72	.95				
		New	4.12	2	72	.02				
Reaction time	Pre-Sound Recognition	Identical	.39	2	72	.68	.2	2	72	.82
		Similar	.7	2	72	.5	3.77	2	72	.03
		New	.05	2	72	.96	.62	2	72	.54
	Criterion Test	Identical	1.37	2	72	.26	1.11	2	72	.34
		Similar	9.76	2	72	<.001	10.59	2	72	<.001
		New	1.32	2	72	.27	.05	2	72	.95
	Phonological Test	Identical	1.78	2	72	.18				
		Similar	1.88	2	72	.16				
		New	.78	2	72	46				
	Post-Sound Recognition	Identical	.08	2	72	.92				
		Similar	.77	2	72	.47				
		New	.09	2	72	.91				

Exploring the assumption of independence

As the study was conducted with every participant individually, the assumption of independence was met. To test the assumption of independent errors, a Durbin-Watson test was performed in all tasks and phases. According to Field (2018), if the test value is less than 1 or greater than 3, then the assumption is violated. As seen in Table O7, none of the tasks in the accuracy data scored less than 1 or more than 3, meaning that the assumption of independence was met. However, this was not the case with reaction time data as there are three conditions scored less than 1 which are Task 1 of the Criterion Test Phase in the similar and new-phoneme conditions and new-phoneme condition in the Phonological Test Phase.

Table O7: *Durbin-Watson Test Results*

Measure	Task	Phoneme type	Durbin-Watson statistics			
			Pre-Sound Recognition	Criterion Test	Phonological Test	Post-Sound Recognition
Accuracy	Task 1	Identical	1.77	1.99	1.68	2.31
		Similar	2.11	2.09	1.83	1.70
		New	1.99	1.97	1.78	2.08
	Task 2	Identical	2.01	1.95		
		Similar	1.95	1.93		
		New	2.26	1.78		
Reaction time	Task 1	Identical	1.75	1.37	1.1	1.46
		Similar	1.72	.77	1.29	2.24
		New	1.85	.86	.98	1.58
	Task 2	Identical	1.24	1.84		
		Similar	1.08	1.06		
		New	1.82	1.07		

Appendix P: Non-Parametric Test (Kruskal-Wallis Test)

This test also tested two measures of the participants' performance: accuracy and reaction time in both the Phonological Test Phase and the Post-Sound Recognition Phase.

Phonological Test Phase

As for the accuracy measure, results from Kruskal-Wallis with an adjusted *p*-value showed that orthographic input type only had a significant effect in the identical-phoneme and new-phoneme conditions, but not in the similar-phoneme condition, as shown in Table P1.

Table P1: *Kruskal-Wallis Test Summary in the Three Conditions of L1/L2 Phoneme Similarity in Participants' Performance (Phonological Test Phase)*

	Identical sound	Similar sound	New sound
Test Statistics	12.39	3.97	11.2
Sig	.002	.13	.004

Pairwise comparisons with adjusted *p*-values indicated that the Arabic-script group did not differ significantly from the No-script group in the identical-phoneme and new-phoneme conditions. However, the Arabic-script group and No-script group had significantly better performance than the Romanised-script group in the identical-phoneme and new-phoneme conditions. As explained in Table P2 and shown in Figures P1, the effect of orthographic input type is consistent across the different levels of L1/L2 phoneme similarity, e.g., similar phonemes always had a lower performance in all three orthographic input-type groups than the other two conditions of L1/L2 phoneme similarity.

Table P2: *Pairwise Comparisons of the Three Orthographic Input Types in the Accuracy Measure (Phonological Test Phase)*

L1/L2 phoneme similarity	Task	Source	Test Statistics	Std Error	Std. Test Statistics	Adj. Sig
Identical phonemes	Yes-No	Arabic / Romanised	16.72	6.118	2.733	.01
		Arabic / No	-3.4	6.118	-.556	1.000
		Romanised / No	-20.12	6.118	-3.289	.003
New phonemes	Yes-No	Arabic / Romanised	15.44	6.020	2.565	.03
		Arabic / No	-3.5	6.020	-.581	1.000
		Romanised / No	-18.94	6.020	-3.146	.005

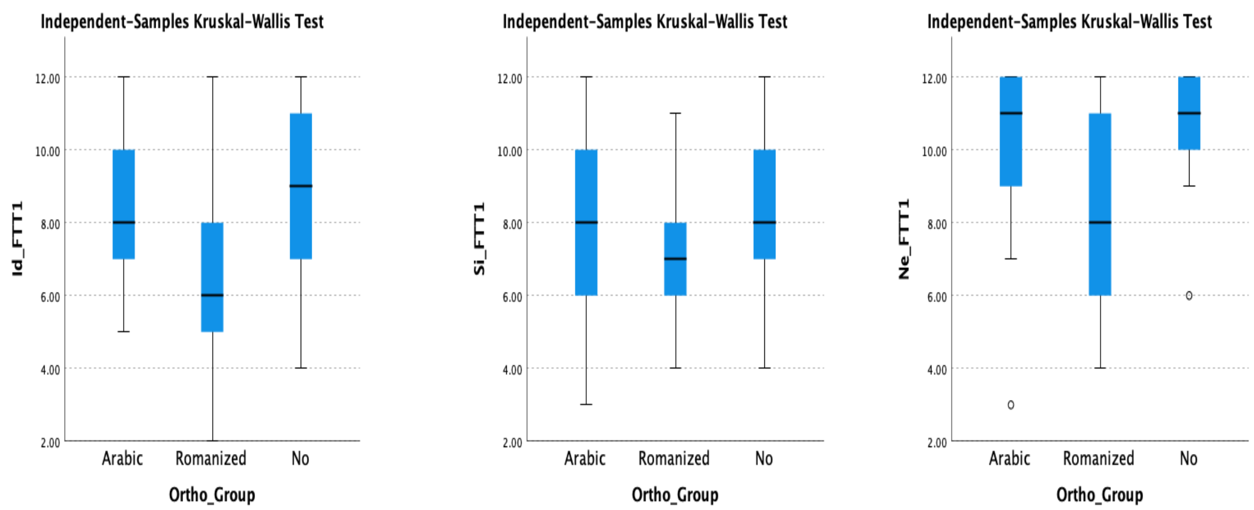


Figure P1 Performance in the three orthographic input types in identical-phoneme condition (left side), similar-phoneme condition (middle), and new-phoneme condition (right side) (Phonological Test Phase)

As for the reaction time measure, results from Kruskal-Wallis with an adjusted *p*-value showed that orthographic input type had no significant effect in both tasks, as shown in Table P3.

Table P3: *Kruskal-Wallis Test Summary in the Three Conditions of L1/L2 Phoneme Similarity in Participants' Reaction Time (Phonological Test Phase)*

	Identical sound	Similar sound	New sound
Test Statistics	.14	.30	.11
Sig	.93	.85	.94

Post-Sound Recognition Phase

As for the accuracy measure, results from Kruskal-Wallis with an adjusted *p*-value showed that orthographic input type had no significant effect in all three conditions of L1/L2 phoneme similarity, identical-phoneme, similar-phoneme, and new-phoneme conditions, as shown in Table P4.

Table P4: *Kruskal-Wallis Test Summary in the Three Conditions of L1/L2 Phoneme Similarity in Participants' Performance (Post-Sound Recognition)*

N		Identical sounds condition	Similar sounds condition	New sounds condition
75	Test Statistics	2.48	1.80	7.10
	Sig	.28	.40	.15

As for the reaction time measure, results from Kruskal-Wallis with an adjusted *p*-value showed that orthographic input type had no significant effect in all three conditions of L1/L2 phoneme similarity, as shown in Table P5.

Table P5: *Kruskal-Wallis Test Summary in Three Conditions of L1/L2 Phoneme Similarity in Participants' Reaction Time in Post-Sound Recognition*

N		Identical sounds condition	Similar sound condition	New sound condition
75	Test Statistics	.31	3.13	3.35
	Sig	.85	.20	.18

Appendix Q: Non-Parametric Test (Friedman Test)

This test also tested two measures of the participants' performance: accuracy and reaction time in both the Phonological Test Phase and the Post-Sound Recognition Phase.

Phonological Test Phase

As for the accuracy measure, results from a Friedman's ANOVA test showed that the performance of the participants was significantly affected by L1/L2 phoneme similarity, $\chi^2(2)$ = 37.8, p = <.001. Pairwise comparisons with adjusted p-values indicated that the performance of the participants in the identical-phoneme condition did not significantly differ from their performance in the similar-phoneme condition. However, this was not the case with the new-phoneme condition which was significantly better than that in the identical-phoneme condition and in the similar-phoneme condition, as shown in Table Q1.

Table Q1: *Pairwise Comparisons of L1/L2 Phoneme Similarity in Participants' Performance (Phonological Test Phase)*

Task	Source	Test Statistics	Std Error	Std. Test Statistics	Adj. Sig
Yes-No	Identical / Similar	-.027	.163	-.163	1.000
	Identical / New	-.833	.163	-5.103	<.001
	Similar / New	-.807	.163	-4.940	<.001

Friedman's ANOVA also showed that L1/L2 phoneme similarity had a significant simple effect in only two orthographic input-type groups: Arabic-script and No-script groups (p <.001), whereas it did not significantly differ in the Romanised-script group (p = .08).

New phonemes were always significantly better than identical phonemes and similar phonemes in both orthographic input-type groups. Similar phonemes, on the other hand, did not significantly differ from identical phonemes in both orthographic input-type groups. This is presented in Table Q2.

Table Q2: *Simple Effect of L1/L2 Phoneme Similarity on Participants' Performance in Three Conditions of Orthographic Input Type (Phonological Test Phase)*

Orthographic input type	Task	Source	Test Statistics	Std Error	Std. Test Statistics	Adj. Sig
Arabic-script group	Yes-No	Identical / Similar	-.04	.283	-.141	1.000
		Identical / New	-.98	.283	-3.46	.002
		Similar / New	-.94	.283	-3.32	.003
No-script group	Yes-No	Identical / Similar	.16	.283	.566	1.000
		Identical / New	-.94	.283	-3.32	.003
		Similar / New	-1.10	.283	-3.88	<.001

As for the reaction time measure, results from a Friedman's ANOVA test showed that the reaction time of the participants was significantly affected by L1/L2 phoneme similarity, $\chi^2(2) = \chi^2(2) = 11.98$, $p = .003$. Pairwise comparisons with adjusted *p*-values indicated that the reaction time of the participants was significantly different in the three conditions of L1/L2 phoneme similarity. The similar-phoneme condition had a significantly faster reaction time than the identical-phoneme condition. The similar-phoneme condition was not significantly different from the new-phoneme condition. The identical-phoneme condition and the new-phoneme condition did not significantly differ, as shown in Table Q3.

Table Q3: *Pairwise Comparisons of L1/L2 Phoneme Similarity in Participants' Reaction Time (Phonological Test Phase)*

Task	Source	Test Statistics	Std Error	Std. Test Statistics	Adj. Sig
Yes-No	Identical / Similar	.560	.163	3.42	.002
	Identical / New	.340	.163	2.08	.11
	Similar / New	-.220	.163	-1.34	.53

Post-Sound Recognition Phase

As for the accuracy measure, results from a Friedman's ANOVA test showed that the performance of the participants was significantly affected by L1/L2 phoneme similarity, $\chi^2(2) = 60.87$, $p = <.001$. Pairwise comparisons with adjusted *p*-values indicated that the performance of the participants in the identical-phoneme condition did not significantly differ from their performance in the new-phoneme condition. However, this was not the case with the similar-phoneme condition which was significantly lower than that in the identical-phoneme condition and in the new-phoneme condition, as shown in Table Q4.

Table Q4: *Pairwise Comparisons of L1/L2 Phoneme Similarity in Participants' Performance (Post-Sound Recognition Phase)*

Source	Test Statistics	Std Error	Std. Test Statistics	Adj. Sig
Identical / Similar	.853	.163	5.22	<.001
Identical / New	.027	.163	.163	1.000
Similar / New	-.827	.163	-5.06	<.001

As for the reaction time measure, results from a Friedman's ANOVA test showed that the reaction time of the participants was significantly affected by L1/L2 phoneme similarity, $\chi^2(2) = 33.38$, $p < .001$. Pairwise comparisons with adjusted *p*-values indicated that the reaction time of the participants was significantly different in the three conditions of L1/L2 phoneme similarity. The similar-phoneme condition was not significantly different from the new-phoneme condition. However, the identical-phoneme condition was significantly slower in their reaction time than both the new-phoneme condition and similar-phoneme condition, as shown in Table Q5.

Table Q5: *Pairwise Comparisons of L1/L2 Phoneme Similarity in Participants' Reaction Time (Post-Sound Recognition Phase)*

Source	Test Statistics	Std Error	Std. Test Statistics	Adj. Sig
Identical / Similar	.933	.163	5.71	<.001
Identical / New	.587	.163	3.59	.001
Similar / New	-.347	.163	-2.12	.10

Appendix R: Descriptive Statistics

Table R1: *Descriptive Statistics for all the Data in the Study / Accuracy Data*

Orthographic input type	L1/L2 phoneme similarity		Pre-Sound Recognition		Criterion Test		Phonological Test	Post-Sound Recognition
			Task 1	Task 2	Task 1	Task 2		
All	Identical	*M*	11.48	11.16	8.17	9.24	7.85	11.62
		SD	1.21	1.18	2.74	2.58	2.51	0.85
	Similar	*M*	9.97	10.12	10.06	10.80	7.52	9.86
		SD	1.54	1.27	2.00	1.78	2.15	1.78
	New	*M*	11.49	10.49	10.12	10.18	9.66	11.54
		SD	1.5	1.75	2.54	2.28	2.46	1.16
Arabic-script	Identical	*M*	11.56	11.28	8.64	9.84	8.36	11.84
		SD	0.58	1.20	1.86	2.13	2.11	0.55
	Similar	*M*	9.96	10.24	10.20	11.40	7.68	9.84
		SD	1.45	1.30	1.55	1.56	2.42	1.81
	New	*M*	11.80	10.84	9.92	10.44	10.16	11.92
		SD	0.64	0.98	2.73	2.39	2.26	0.40
Romanised-script	Identical	*M*	11.40	11.40	6.56	7.92	6.44	11.60
		SD	1.22	0.81	2.84	3.13	2.51	0.81
	Similar	*M*	10.12	10.04	9.60	10.00	6.80	9.52
		SD	1.09	1.45	2.67	2.36	1.63	1.85
	New	*M*	11.16	10.68	9.16	9.00	8.32	10.04
		SD	1.51	1.34	2.88	2.27	2.64	1.74
No-script	Identical	*M*	11.48	10.80	9.32	9.96	8.76	11.44
		SD	1.63	1.41	2.71	1.88	2.33	1.08
	Similar	*M*	9.84	10.08	10.40	11.36	8.08	10.24
		SD	1.99	1.07	1.58	0.86	2.21	1.66
	New	*M*	11.52	9.96	11.28	11.12	10.52	11.68
		SD	2.00	2.49	1.3	1.66	1.93	0.74

Table R2: *Descriptive Statistics for all the Data in the Study / Reaction Time Data*

Orthographic input type	L1/L2 phoneme similarity		Phonological Test	Post-Sound Recognition
All	Identical	*M*	2.33	1.26
		SD	0.93	0.49
	Similar	*M*	2.03	0.91
		SD	1.00	0.43
	New	*M*	2.19	1.06
		SD	1.03	0.45
Arabic-script	Identical	*M*	2.25	1.21
		SD	0.87	0.48
	Similar	*M*	2.03	0.90
		SD	0.88	0.50
	New	*M*	2.13	1.03
		SD	1.16	0.43
Romanised-script	Identical	*M*	2.39	1.27
		SD	1.1	0.53
	Similar	*M*	2.1	0.83
		SD	1.2	0.38
	New	*M*	2.23	0.96
		SD	1.01	0.45
No-script	Identical	*M*	2.35	1.28
		SD	0.85	0.49
	Similar	*M*	1.95	1.00
		SD	0.92	0.41
	New	*M*	2.2	1.19
		SD	0.95	0.45