Impact of Computer-Assisted Pronunciation Training on Libyan Child Learners of English

Hana Ehbara

Doctor of Philosophy (Integrated)

School of Education, Communication and Language Sciences
Newcastle University

January, 2021

Abstract

Understanding second language speech has been a pressing issue for researchers. Accounts for sources of error shown by L2 learners include age of initial exposure, relative markedness, L1 functional constraints, specifically perceptual salience and frequency (Colantoni and Steele, 2008), and perception, which is the basis for explaining cross-linguistic influence by most L2 speech learning theories (Colantoni, Steele, and Escudero, 2015), but they do not include the delay of oral production. When we look at younger beginner L2 learners, L1 influence can also be observed. The aim of the present study was to address the impact of delaying oral production and for this computer-assisted pronunciation training (CAPT) was used on Arabic-speaking children in Libya learning English as a second language with no prior instruction in English. English instruction in Libya is typically delivered by non-native teachers whose non-native input is also a possible source of L1 influence. The software provided native speaker input to address this additional aim. Within the software, test words were presented in orthographic and audio formats with pictures depicting meaning. Predictions on the role of output have varying underlying assumptions. Proponents of the importance of production practice such as Swain (1985; 1995; 2005) and Mackey (2007) argue that it is a tool for creating novel linguistic knowledge and promoting cognitive processes (see Colantoni and Steele, 2008 in their Hybrid model).

Following a three-week training with use of the software, 38 seven-year-old participants took part in picture-naming, read aloud and delayed repetition tasks in an immediate post-test and, of these, 30 took part in similar tasks for a delayed post-test 10 weeks later. The 38 participants were divided into two training conditions, Listen and Speak (n.=20) and Listen Only (n.=18) to test the role of delayed production on L2 learning. Another group of 20 aged-matched participants took part in a three-week training with use of Traditional Teaching and participated in the same tasks in an immediate post-test and of these 18 took part in similar tasks for a delayed post-test 10 weeks later. The Traditional Teaching condition was added to compare input type on participants within the same age group.

The aspects of pronunciation measured were target-likeness rating, match rating, various acoustic cues including Voice Onset Time (VOT), vowel-onset fundamental frequency, and spectral tilt (Ahi-A23). The participants' L2 values were compared to the target language and their L1 values to test predictions made by models of speech learning. The phonetic data revealed signs of merger categories between L1 and L2 corroborating the findings of Flege (1995) and MacKay, Flege, Piske, and Schirru (2001). Additionally, phonological processes were examined and compared to processes found in L1 English child phonology. The amount of lexical learning was also explored. Results for TL-likeness and match rating revealed that the experimental conditions statistically outperformed the Traditional condition in both tests. In the delayed test however, the Listen and Speak condition statistically outperformed Listen Only participants, who continued to outperform the Traditional learners. VOT and vowel-onset f0 analyses revealed that participants from all training conditions failed to establish independent L2 categories. Rather, they illustrated intermediate values resembling both native and target phonetic categories. In terms of lexical learning, the experimental conditions outperformed the Traditional condition in terms of the amount of fully learned words in the delayed repetition and picture-naming task but they all performed the same in the read aloud task. Some interlanguage processes were demonstrated by the learners in addition to the expected transfer from their Arabic variety. These varied depending on the sound class and conformed to universal language development and input from native speakers of the target language. It is concluded that the findings support the importance of output in language learning for L1 beginning-level children in the classroom as suggested by the Hybrid model (Colantoni and Steele, 2008).

Dedication

I dedicate this thesis to my husband Mohamed, my children Bashir and Yafa and my PhD baby Elena, who all made this whole journey much worthwhile.

I also dedicate it to my beloved mother Zuhra, and sisters Huwaida and Aisha.

Acknowledgements

The PhD has been an incredibly challenging journey for me. I would like to thank many people, without whom it would not have been possible to finish.

I would like to begin by thanking my supervisors, Professor Martha Young-Scholten and Dr Jalal Al-Tamimi, who each gave so generously of their time and showed great kindness, patience and support. Martha, thank you for putting a word for me with the DigLin developers, suggesting this topic for me, and setting up the reading group during the early stages of writing my proposal. You always found a way to bring peace to my troubled student soul. Jalal, thank you for always encouraging me to 'dig deeper' (this phrase will forever be associated with you!) and for bringing the best out of me. Thank you both for the time you dedicated to giving me detailed feedback which has shaped this thesis, for believing in me, and always having the right words to say to me. You have already taught me so much and I look forward to learning more in the future. Thank you for encouraging me to finish and giving me the time I needed to write.

I would also like to thank Newcastle University: CRiLLs for giving me the opportunity to enhance my skills set in organising the 8th Postgraduate Conference in Linguistics, R Meetups for giving me the opportunity to learn doing statistics using R, HaSS for providing me with training on a myriad of academic areas. I would like to thank Ghada Khattab for showing support especially after having my PhD baby and encouraging me to attend academic events. I would also like to thank my review panel Ghada GHattab, S J Hannahs, and Gary Taylor-Raebel, for providing their feedback and words of wisdom each year.

I am forever indebted to Misrata University for paving the first steps for my scholarship nomination and to the Libyan Ministry of Education for granting me this scholarship. Thank you to the Libyan Cultural Attache in London, especially Mr Husam Ounallah, Mr Faisal Antat, and Mr Adel Aldoukali for being compassionate and supportive all the way, and for Mr Ahmed Alhweik, may his soul rest in peace.

I am especially grateful for the help and patience of the assistant researchers, my sisters, Huwaida and Aisha, and my friend and colleague Hawa Esskaire. Not only did they conduct the training with utmost patience and adequacy, but also kind enough to carry out lengthy data collection sessions.

This research would have not been possible without the generosity of the participants' parents and their valued patience in persevering throughout the training period and data collection times. A massive thank you to my participants, the children attending the training and data collection sessions. The training and data collection would not have been possible without the generosity of Mr Swaib, Mr Shenina and my mother, as the headteacher of Shuhada Qaser Ahmed Primary School, whom all opened their doors for me to use their language classrooms.

To all my dear friends, Manal Ben Hmeida, Aisha Abusahmain, Alyaa Olwi, and Hanain Brohi for their emotional support. To my dear friends in 2.05 and 2.41, Nimet Copur, Shazia Tezdemir, Rana, Fatma, Nada, Badria, Yang, Hajar, Noura, Sara, Hahn, Dan, and my friends in Percy building basement, Enas Filimban, Becky Mousa, Maha Ibrahim, Natalia Pavlovskaya, Xinliang Jiang, and many others for their kind words and for sharing their wisdom and support along the way. I would also like to thank my neighbour and my close friend, Huwaida, for her support and encouragement. She always listened when I needed her. I love you all so much and I am and will be forever grateful to you for being part of my life.

My biggest thank you goes to my family: first my mum for her support, prayers, believing in me, and constantly reminding me of how proud she is of me, and my dad for recommending I study at Newcastle University. Second, my husband for believing in me more so than I believe in myself and for his emotional and financial support. Third, my children, Bashir, Yafa for being the most patient and kind, and Elena, for giving meaning to what I do. I am so proud of you all. Finally, thanks to the Facebook groups, PhD Early Career Researcher Parents, Virtual SUAW-Parents Edition, and Full Draft Club. I'd like to thank all the members including Elizabeth, Elin, and Maryam for providing support and giving me the solidarity and the sense of student community I needed.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Introduction

This chapter presents the background of the present study in section 1.2. In this section, I first present a case for the impact of age on L2 speech acquisition arguing that in many cases, it can be confounded with input type. The section also considers the role of perception and production practice in L2 speech learning highlighting some knowledge gaps in the literature. Section 1.3 provides a brief description of the setting of the study. The chapter concludes with the structure of the thesis, in section 1.4.

1.2 Background

Pronunciation in second language (L2) acquisition can be affected by a range of internal and external factors including age of onset of acquisition, quality of target language (TL) input, first language L1 knowledge, amount of continued L1 use and/or L2 input, context of learning (i.e. in the target language or elsewhere), motivation, gender, length of residence in the L2 country, aptitude, mimicry ability, educational level, formal instruction including error correction (Piske, MacKay, and Flege, 2001), social factors (Zampini, 2008) and anxiety (Krashen, 1982; Major, 2001).

Whilst the present study does not examine the effect of age, in the following section, I briefly note the role it plays in L2 pronunciation. In doing so, I echo the proposal made by Young-Scholten (1995), which argues for an input-based explanation instead of age to explain asymmetries between adult and child L2 speech performance.

1.2.1 *Earlier is not necessarily better*

The literature on second language phonology has focused to a great extent on the age of acquisition factor in phonological acquisition considering it the most crucial factor (Ioup, 2008). Research findings have led to the belief that an earlier age of acquisition correlates with more successful L2 acquisition. Early observations of this lack of success by adults instigated the proposition of a 'Critical Period Hypothesis' for acquisition, which was advanced by Lenneberg (1967). Although it was primarily an argument for first language acquisition, Lenneberg made reference to second language acquisition. Scholars who subscribe to this notion argue that there is a critical age window when it is possible for L2 learners to achieve native-like competence (DeKeyser, 2000; Lenneberg, 1967).

This window is before learners reach puberty, which is around the age of 13. This link is made on the assumption that acquisition is neurologically-conditioned and that with age comes a loss of neural plasticity as a result of maturation (Canfield and Jusko, 2009; Lenneberg, 1967). It is thus argued that post-pubescent learners are not capable of achieving native-like proficiency since the critical period is linked to puberty. Similarly, it is argued that native-like pronunciation can only be mastered if learning begins before the age of puberty.

However, the Critical Period Hypothesis received ample criticism. Research by Flege (1995) and Flege, Munro, and MacKay (1995a) established that the correlation between age of acquisition[1] and L2 phonological competence is gradient exhibiting a decline over the life span. On this basis, they argue that if Lenneberg's Critical Period was correct, we would expect to see a binary

---

[1] In most of these studies the age factor is referred to as age of arrival/acquisition, (AoA), or onset of learning.

distribution of performance whereby pre-pubescent learners would illustrate native-like competence and post-pubescent learners would exhibit non-native-like competence. In fact, evidence from research has demonstrated that adult learners are capable of reaching native-like competence based on native speaker rating or by instrumental phonetic examination in both perception (e.g. Escudero and Boersma, 2004) and production (e.g. Birdsong, 2007; Colantoni and Steele, 2006) albeit there is a rather low number of cases for such observations.

Other researchers have called for a 'sensitive period'. Advocates of the sensitive period (e.g. Hurford, 1991; Long, 1990) have also proposed multiple critical periods with each one impacting a different linguistic domain and ending at various ages. For phonological acquisition, it has been proposed that the ability to draw on abstract sound patterns regulating human language declines severely between six-seven and sixteen-seventeen years of age due to biological maturation (DeKeyser, 2000: 518-9). This proposition was narrowed more specifically by Long (1990: 280), who argued that the ability to achieve native-like phonology in many humans starts to severely decline at the cusp of seven years of age and that after the age of 12, this ability is lost regardless of how motivated a learner is or the number of opportunities to learn. He indicates that up until the age of six, an individual is more likely to acquire native-like accent that is free of any foreign accent. Researchers such as Hakuta, Bialystok, and Wiley (2003) specify that an early age of acquisition is an optimal, though not the only age window for a learner to achieve native-like attainment. More recent findings (for example Pfenninger, 2016), show that starting L2 learning in foreign language classrooms earlier does not warrant better L2 performance. Findings show that L2 learners who start learning an L2 around the same age, yet exhibit differences in their L2 pronunciation development (Flege, Munro, et al., 1995a) and those starting at different ages but manage to achieve similar levels of L2 pronunciation performance (Muñoz, 2014) indicate that age is not as a crucial factor as was once believed. Therefore, the ingredient for L2 pronunciation success cannot be determined solely by age. The age factor is confounded by the fact that child learners, in comparison to adult learners, have better opportunities reflected in ideal L2 input and better L2 interactions as pointed out by Young-Scholten (1995) and Mack (2003). It has been asserted that testing the critical or sensitive period hypothesis objectively would require a sample of children and adults receiving the same amount and quality of L2 input, with the same degree of motivation to learn the L2, matching amount of L2 use, same degree of ethnic affiliation and finally, an ongoing direct measure of alleged acquisition-related neuro-biological changes (Baker, Trofimovich, Flege, Mack, and Halter, 2008; Flege, Yeni-Komshian, and Liu, 1999). The present study however is limited to seven-year-old children.

1.2.2 *Perception and production accounts*

Krashen (1985) argues that it is comprehensible input that facilitates learning not output. He suggests that learners should be allowed time to process input without pressuring them to produce output. In Krashen's (2018: para.1, emphasis original) recent publication, his opening line states that,

> When acquirers are forced to produce language that they have not yet acquired, known as "forced speech", they often experience anxiety. I argue here that forced speech is not only uncomfortable, it makes no direct contribution to language acquisition.

In phonology, it is experience that alters phonetic perception (Kuhl, K. Williams, Lacerda, Stevens, and Lindblom, 1992).[2] In L2 acquisition, Colantoni and Steele (2008) argue that learners continue to modify their interlanguage (IL) based on feedback by comparing their output to a model input. When a learner perceives their own output as being no different from the model input, their phonological categories reach an end state as in they require no further modification. Their phonological categories become fossilised and will no longer evolve. This suggests

---

[2] This will be elaborated in Chapter 2.

the crucial role production practice plays in L2 speech learning, which is in favour of the propositions made in the Output Hypothesis (Swain, 1995; Swain, 2005). Swain (1995; 2005) also argues for the importance of output as part of the learning process and that it should not be considered merely as the product of learning. Swain proposes that language production allows learners to notice differences between their production of the L2 and the target language giving them the opportunity of conscious reflection on their learning. Whilst Swain was not specifically referring to L2 phonological/phonetic learning, her views are supported in L2 speech learning (Colantoni and Steele, 2008).

No study on L2 speech has further compared the effect of delayed production and instant production practice – which is argued to allow learners to reflect on their production and to continuously modify to match the model input – in children from a very different language, namely Arabic, just starting to learn English. Thus, the present study set out to address this gap and test the claims put forward. Based on the above, the present study asks whether and, if so, L2 pronunciation in seven-year-old children varies as a function of 1) presentation method, that is perception-only-practice vs. perception-and-production practice, and 2) input type, that is native vs. Arabic-accented. The choice of children – and not adults – is twofold. First, seven-year-old children are prepubescent, which minimises the potential effects of the age factor. Second, the chances of finding child participants with no prior instruction in English are higher than those for adults, whose aspects of L2 language can potentially be fossilised.

To this end, the present study utilised a computer-assisted pronunciation training (CAPT) system to teach English pronunciation to Arabic-speaking children providing them native English-speaker input. To further test whether input is a factor in L2 speech learning, a group of age-matched children with the same linguistic background were involved in a traditional classroom-based training. This was provided by a local teacher of English who spoke Arabic-accented English.

1.3 Setting of the study

The setting of the present study was Libya. In Libya, English was introduced in the national curriculum in 1944-1951. The syllabus comprised only vocabulary and reading and teachers adopted the grammar-translation approach (Suayeh, 1994). Throughout the 1960s and 1970s a series was introduced that is more compatible with the Arabic social and cultural contexts entitled *English for Libya* (Gusbi, 1966) and *Further English for Libya* (Gusbi and John, 1974). In 1984, English teaching was eradicated in grades seven to nine when children are aged 12 to 14 and teaching started in year ten instead when children are aged 15. In 1993, it was reintroduced for grades seven to nine (age range 12 – 14) with shorter sessions (3 sessions per weeks x 40 minutes per session = 3 hours weekly) than the period prior to its eradication (4 sessions per weeks x 45 minutes each = 4 hours weekly)[3] (UNESCO-IBE, 2007).

One of the characteristics of English teaching at the time (in the 1990s) was that teachers were advised to utilise the audio-lingual method. The method revolves around listening and speaking practice using audio recordings and language laboratory drills (Liu and Shi, 2007). It also involved dialogue memorisation through dialogue imitation and role-playing and repetition drill whereby students repeat teacher models quickly and accurately (Larsen-Freeman, 2000). Libyan teachers and students prefer such drilling, memorisation and repetition practice (Imssalem, 2002). However, they were found not to use audio recorders of native-speaker input and relied instead on teacher and peer models (Emhamed and Krishnan, 2011) – who would most probably have inaccurate pronunciation.

More recently, since 2000, there has been a curriculum reform and English language teaching is first introduced in the curriculum in year five (Aloreibi and

---

[3] The amount of formal English instruction was based on personal communication with teachers working in the profession for three to four decades.

Carey, 2017) when a pupil is 10 years of age. Despite the introduction of a curriculum reform which is based on the Communicative Language Teaching approach, the Grammar-Translation and the Audio-Lingual Methods continue to prevail in Libya (Saaid, 2010). In the English language classroom, teaching is teacher-centred, teachers continue to use the Grammar-Translation method and use Arabic to explain grammar and the meaning of texts from textbooks, whereby students are instructed to translate text from English to Arabic, or vice versa (Emhamed and Krishnan, 2011). Teaching is limited and lasts for three 40-minute sessions a week amounting to two hours of varied language tasks. Even though the new curriculum focused on everyday dialogue topics, oral practice, pronunciation drills to master the English sound system at primary level and reading aloud, these recommendations are far from being implemented in the Libyan English language classroom.

The problem with the lack of instruction efficiency in English teaching in Libya does not restrict itself to the year at which English is introduced. Teachers lack the proficiency to speak English and rely on Arabic in their instruction (see Orafi and Borg, 2009: 249 and Y. Omar, 2014: 127, ora). Orafi and Borg (2009) highlight another problem in delivering English in the language classrooms. Teachers' beliefs in their roles as teachers have not changed to match the objectives of the new system. Classroom practice and content are exam-driven (N. Saleh and Zakaria, 2013). Examination policies do not support the importance of pronunciation or oral practice. Assessment content is based on vocabulary, grammar and the memorisation of answers to pre-set questions within the curriculum. Teachers rarely use recordings and they rely on their own pronunciation, which is heavily accented (Fraser, 2000; Rababah, 2003).

Thus, introducing a programme that utilises native speaker input to teach words through speech to picture-meanings, and which allows the learner to work independently if they choose to, should prove beneficial for beginners.

1.4 Structure of the thesis

The subsequent chapters are divided as follows:

Chapter 2 provides a review of factors influencing L2 speech learning pertinent to the current study. This involves 1) formal instruction – specifically L2 pronunciation teaching, 2) literacy and orthography in L2 speech learning given that the child participants at the start of the study were illiterate in English and potentially in Arabic and that the CAPT software does not assume its users are literate and therefore aims to support users in making grapheme-phoneme correspondences in English, 3) markedness and sonority distance since they exert added and varied difficulty in perception and/or production depending on L2 sound or sound structure (Colantoni and Steele, 2008). The chapter also examines prominent models put forward to explain mechanisms involved in speech learning. These models provide various explanations based on perception, production and the role of feedback. They include the Native Language Magnet Model (NLM) and the Magnet Effect (Kuhl, 1991; Kuhl, K. Williams, et al., 1992), the Perceptual Assimilation Model (PAM) (Best, 1994; Best, 1995; Best and Tyler, 2007) and its extended version for second language (PAM-2), the Speech Learning Model (SLM) (Flege, 1995), the L2 Perception Model (L2LP) (Escudero and Boersma, 2004) and the Hybrid Model of segmental acquisition (Colantoni and Steele, 2008).

Chapter 3 provides an overview of the characteristics of the participants' L1, Libyan Arabic, its phonemic inventory, and syllable structure. It also provides an overview of the characteristics of the target language, British Received Pronunciation (RP) English. This chapter helps identify differences between the two languages, which facilitates formulation of predictions for L2 English learning by the study participants based on the models described in Chapter 2.

Chapter 4 provides an overview of child phonological acquisition in English as well as in Arabic, the learners' first language. The purpose of this chapter is to set the perspective of child phonological processes in order to see whether some errors might be developmental, rather than based on L1 influence and importantly, whether they do so more often under one of the three training conditions.

Chapter 5 describes the methodology adopted in the present study. It presents the research questions:

RQ1: Which training method will result in more target-like (and less L1 like) pronunciations of problematic sounds Libyan learners typically display learning English, including:

1. Affricates
2. CC Coda clusters
3. Diphthongs
4. Plosives
5. Voiceless dental fricatives
6. Rhotic approximants

RQ2: What developmental processes do beginner Arabic child learners of L2 English exhibit in each condition?

RQ3: Which training method will result in the most learned words?

The chapter then describes the training procedure, the CAPT programme used for training with the experimental groups, participant sample, the design, stimuli for Arabic and English, data collection and analysis procedure – including segmentation and labelling – and finally the general modelling for statistical analysis.

Chapter 6 begins by addressing RQ3 regarding word learning. This is to establish to what extent the participants fully and partially learned the words they were expected to learn. The chapter also sets out to answer the first part of RQ1 above, which links target-likeness rating and match rating to the three training conditions examined.

To support the answer to the first research question using instrumental phonetics, Chapter 7 and Chapter 8 present production results of the acoustic analyses. Chapter 7 presents the analysis and results for voicing as measured by voice onset time (VOT) and vowel-onset fundamental frequency (f0). Chapter 8 presents the acoustic analysis for place of articulation as measured by spectral tilt (Ahi-A23). The chapters conclude by providing a summary and discussion of the findings for each acoustic outcome.

Chapter 9 aims to answer the second research question set out in Chapter 5. This chapter describes the phonological processes Libyan-Arabic seven-year-old learners adopt in learning English as a second language. It specifically examines phonological processes within problematic sound classes for a typical Libyan Arabic L2 learner of English. These include affricates, bi-consonantal coda clusters, dental fricatives, diphthongs and the English rhotic approximant. It compares these processes across training conditions and discusses the findings in light of child phonology.

Chapter 10 draws on key findings in relation to the main goal of the study.

The final chapter, Chapter 11, concludes this thesis by discussing limitations and providing recommendations for further research.

Chapter 2: Second Language Speech Learning

2.1 Introduction

The following chapter gives an overview and critique of the literature surrounding L2 speech learning. It consists of five main sections. We begin with an exploration of the role of markedness and sonority and how it relates to the acquisition of (English) L2 phonology in section 2.2. In section 2.3 we examine existing experimental studies of L2 pronunciation training and instruction whilst considering type of input and mode of presentation. The subsequent section, 2.4 looks at the role of literacy and orthography and the literature surrounding its impact on L2 pronunciation. Finally, in section 2.5 we explore models of L2 speech learning.

2.2 Markedness and sonority

The L1 acquisition of sounds can be similar, to a certain degree, across languages.[4] For example /ð/ and /ɹ/ are late-acquired not only in English but in many languages of the world (Mcleod and Crowe, 2018; D. Ohala, 2008). In second language phonological acquisition, accounts making predictions and assumptions based on typological markedness such as Eckman's markedness

---

[4] Differences across languages can be explained on the basis of frequency effects (Zamuner, 2003). When the acquisition of /v/ for example was compared between English, Bulgarian, Swedish and Estonian children, it was found that it was acquired earlier in all languages but English. English children were relatively late to acquire it. Ingram (1999) concluded that the reason for this was that the occurrence of /v/ in English was relatively less frequent compared with the other languages from the study. However, the role of frequency will not be further explored in the present study.

differential hypothesis (Eckman, 1977) have argued that speech segments that are marked are difficult, if not impossible, to acquire and vice versa. Eckman's model attempts to explain relative degree of difficulty in L2 acquisition of two or more sounds based on typological markedness. L2 sounds and sound structures that do not exist in the learner's L1 phonemic inventory *and* are typologically marked are predicted to be difficult to acquire. Implicational relations, he argues, determine typological markedness. That is,

> [a] phenomenon A in some language is more marked than B if the presence of A in a language implies the presence of B; but the presence of B does not imply the presence of A (Eckman, 1977: 320).

Eckman (see 1991: Structural Conformity Hypothesis) later argued the same universal generalisations that account for L1 acquisition apply to L2 acquisition. The revised version of the Markedness Differential Hypothesis into the Structural Conformity Hypothesis in Eckman (2004) does not assume any roles for the differences between L1 and L2.

2.2.1 *Fricatives*

Maddieson (1984) surveyed 316 languages comprising UPSID (short for UCLA Phonological Segment Inventory Database). In his survey, fricatives – generally – were found in 296 (94%) of the languages in UPSID. Moreover, not all languages had a similar number of fricatives. The most frequent fricative in UPSID is the /s/ sibilant found in 266 languages. Compared to /s/, the voiceless labio-dental fricative /f/ was less frequent occurring in 135 languages but the frequency of the dental fricatives, /ð/ and /θ/ were one of the rarest found in 21 and 18 languages respectively. Several reasons have been proposed for this disparity, amongst which is differences in intensity and perceptual salience (Maddieson, 1984). Furthermore Ladefoged and Maddieson (1996) explain that the manner of articulation of fricatives in general requires a greater degree of articulatory precision compared to plosives or nasals for example. Articulation of fricatives requires the channel of air turbulence inside the oral cavity to be shaped precisely for this purpose and that this shape has to be maintained throughout the period of producing the fricative segment. In terms of voicing, the implicational relationship is that the presence of a voiced fricative in a given language's phonemic inventory implies the presence of its voiceless cognate in that inventory.[5] Consider the frequencies of various fricatives in the world's languages and the relations between voiceless and voiced fricatives from which Maddieson draws implicational relations for voicing in this sound class in table 2.1. Maddieson (1984) does not draw implicational relations by place of articulation, therefore we can rely on frequencies to establish markedness for the various places of articulation for this sound class.

Table 2.1 Frequency of fricatives in the world's languages Maddieson (1984: 45)

Voiceless	Number of languages	Percent	Voiced	Number of languages	Percent
/s/	266	89.9%	/z/	96	32.4%
/ʃ/	146	49.3%	/ʒ/	51	17.2%
/f/	135	45.6%	/v/	67	22.6%
/θ/	18	06.1%	/ð/	21	07.1%

2.2.2 *Rhotic sounds*

Rhotic sounds, based on the cross-linguistic typology in Maddieson (1984), rhotic approximants – like the ones found in British RP English – are segmentally more marked relative to rhotic taps/flaps – like the ones found in Libyan Arabic for example. Their occurrence in the world's languages (28 languages out of the 282 in UPSID amounting to 9.9%) is lower than that for tap/flap rhotics (104/282 = 36.9%) (Maddieson, 1984). It is postulated that the difficulty of rhotics in general, is due to the (tongue root movement towards the pharynx) secondary constriction in the pharynx, which is a cross-linguistic feature causing delayed articulatory development and misarticulations (Boyce,

---

[5] For exceptions, see Maddieson (1984: 47).

Hamilton, and Rivera-Campos, 2016). In terms of place of articulation, alveolar and dental rhotics are the most common across languages (86.4%). In terms of implicational relationships in rhotics, Maddieson (1984) argues that it is not a straightforward task. This is because: 1) a fifth of the world languages have two or more rhotic sounds, 2) the type of rhotics in any language are predominantly determined by its lateral inventory, and 3) the lack of certainty of phonetic realisation of the rhotics found in descriptive studies. Thus it can be said that based on the frequency of rhotic taps and rhotic approximants above that rhotic approximants are more marked than taps or trills for example.

2.2.3 *Vowels*

In UPSID, 2549 monophthongal vowels were identified (Maddieson, 1984). The frequency of these vowels is shown in table 2.2. The table shows that high-front vowels are the most frequent in the world's languages (91%). Note however, that Maddieson (1984) grouped vowels described in the original source as either located anywhere between higher-mid to lower-mid, or just mid without further specifications in one category which he annotated by inverted commas as shown in table 2.2. Note also that length differences are not considered in this frequency distribution.

By comparison, 83 diphthongal vowels were identified in UPSID found in 23 different languages, 22 of which diphthongs are found in a single language, !Xũ, part of the Khoisan language family (Maddieson, 1984: 133). Diphthongs in UPSID are rather heterogeneous and lack commonalities apart from the ones occurring in more than two languages displayed in table 2.3. Maddieson (1984) argues that diphthongs starting and ending with a high vowel element are generally favoured across world languages. He adds that this cannot be attributed to a tendency of maximising distinctiveness between diphthongs. This is because his findings show that diphthongs having short trajectories (for example /ei/ and /ie/) and large trajectories (for example /ai/ and /au/) in the vowel space are just as

Table 2.2 Most common vowel qualities in Maddieson (1984: 125)

Vowel	Number of Languages	Percent
High and low vowels		
/i/	290	91.5%
/a/	279	88.0%
/u/	266	83.9%
Vowels in the mid range		
/“o”/	139	43.8%
/“e”/	118	37.2%
/ɛ/	118	37.2%
/o/	109	34.4%
/e/	100	31.5%
/ɔ/	99	31.2%

common in the world’s languages.

Table 2.3 Common diphthongs in Maddieson (1984: 134)

Diphthong	Number of Languages	Additionally
/ei/	6	(also Burmese with additional nasalisation)
/ai/	5	(plus 2 languages with /ae/)
/au/	5	(plus 2 languages with /ao/)
/ou/	4	(also Burmese with /ɒu/ but not /ou/)
/ui/	4	
/io/	4	(including Evenki with /ioː/)
/ie/	3	
/oi/	3	

Maddieson (1984) argues that generalisations made about the world’s vowels are summarised as follows:

1. Front vowels are usually unrounded.
2. Back vowels are usually rounded.
3. Low vowels are usually central.
4. Central vowels are usually low.
5. Nearly all languages have /i, a, u/ (the minimum vowel triangle), amongst which /u/ is the least common.

6. Contrastive length correlates with distinctive vowel qualities in a given vowel inventory of a language.
7. Diphthongs with a high vowel element are favoured over those without.

The above indicates that diphthongs are generally more marked than monophthongs.

Similarly, phonological processes affecting the syllable domain as well as the segment can be shared across languages both of which reflecting what D. Ohala (2008) calls *emerging phonology*. Eckman (2008) argues for universal tendencies as an explanation for such developmental similarities. The same case has been argued for second language learning. Kløve and Young-Scholten (2008) posit that L2 learners of English simplify consonant clusters not only to conform with their L1 syllable structures, but also with language universals. However, it has been found, for example, that whilst children learning L1 English most often resort to deletion or reduction as a repair mechanism in consonant clusters (Pater and Barlow, 2003), adults learning L2 English tend to resort to vowel epenthesis for the same clusters (Broselow and Finer, 1991; Fantazi, 2003; Hancin-Bhatt and Bhatt, 1997; Karimi, 1987; Kwon, 2006; Major, 1994).[6]

Several studies have examined the postulations of Markedness Differential Hypothesis especially in phonology. For example, Anderson (1987) investigated the acquisition of English onset and coda clusters by native speakers of Arabic and two varieties of Chinese, that is Mandarin and Amoy. Anderson found that in coda clusters, the performance of Arabic learners was more target-like than that of the Chinese speakers of Mandarin and Amoy. The difference in performance correlated with the degree of markedness supporting the Markedness Differential Hypothesis. Moreover, the findings show that more errors were observed in marked final clusters than marked onset clusters.

Broselow and Finer (1991) also assert the important role markedness plays in the acquisition of certain English consonant clusters by L2 learners. They propose that some consonant clusters which are closer in the Sonority Hierarchy pose greater difficulty

---

[6] A point regarding orthographic interference relevant to this observation is made later in section 2.4. A study by Young-Scholten, Akita, and Cross (1999) compared two groups of adult L2 learners, one presented with orthographic input simultaneously with auditory input, and the other presented with auditory input only. In examining their production of L2 complex consonant clusters, participants in the former condition (auditory and orthographic input) exhibited epenthesis whereas participants from the second condition (auditory input only) exhibited productions reflecting those found in L1 child acquisition, that is cluster simplifications.

for L2 learners than those that are dispersed in sonority ranking (Sonority Dispersion Hypothesis)

2.3 Teaching L2 pronunciation

Piske, MacKay, et al. (2001) and Piske (2012) discern several factors which influence the effectiveness of L2 pronunciation teaching. These are 1) quality and quantity of the L2 input, that is the more native TL input the learners get, the more native-like their pronunciation will be, 2) L1 background, that is phonological differences between the L1 and L2 results in specific difficulties in perception and production, and finally 3) training in perception and production of an L2, that is training tasks which target structural differences between the learner's L1 and target language have been found to improve learners' pronunciation.

Input type has been identified as a factor affecting L2 pronunciation. Young-Scholten (1995) argues that one of the possible explanations that adult learners do not achieve native-like attainment in L2 phonology is the quality of input they typically receive. Other factors relating to input include the acoustic-phonetic variability of input. A number of research studies examined the impact of acoustic-phonetic variability on L2 speech learning. For example, Logan, Lively, and Pisoni (1991) showed that perceptual training using high variability phonetic tokens (involving multiple talkers as opposed to synthetic stimuli which is rather low in variability) improved Japanese learners' perception/identification of the novel L2 English phonetic contrast /l, r/, which is non-distinctive in the learners' L1. Bradlow, Pisoni, Akahane-Yamada, and Tohkura (1997), X. Wang (2002), and Thomson (2011) also used high variability phonetic training. Whilst training from stimuli with high acoustic-phonetic variability is generally thought to improve perceptual learning (Logan et al., 1991), Perrachione, Lee, Ha, and Wong (2011) found that this interacts with individual differences. They found that high-variability phonetic training improved learning only for learners who had strong perceptual abilities. By comparison, learners who had weaker perceptual abilities were at a disadvantage by high-variability training. Moreover, very few studies have demonstrated its advantageous impact on production for consonants (Bradlow, Pisoni, et al., 1997) or vowels (Thomson, 2011).

X. Wang (2002) and JY Lee (2009) on the other hand used computer-assisted pronunciation training to manipulate natural speech input by accentuating segmental cues to assist in perceptual identification. The duration of vowels was manipulated to allow learners to focus solely on quality. A number of experimental studies have used computer-aided pronunciation training software using native speaker input after which learners were asked to imitate the input (Kissling, 2013; Lord, 2005; P. Pearson, Pickering, and Da Silva, 2011; Weinberg and Knoerr, 2003). In such studies imitations were recorded to allow learners to compare their productions against native speaker models. This technique was also used to teach suprasegmentals (Chun, Jiang, and Ávila, 2013; Hardison, 2004), and global speech aspects (Tanner and Landon, 2009). In other computer-assisted pronunciation training studies, training involved automatic speech recognition feedback (for example Neri, Mich, Gerosa, and Giuliani, 2008). Training in perception and production of L2 speech has been found to have positive effects. Neufeld (1978) showed that perceptual training improved the production of articulatory and prosodic features of English adult L2 learners of three non-Indo-European languages. Trofimovich, Lightbown, Halter, and Song (2009) showed that a two-year comprehension training in listening and reading practice in the absence of any speaking practice improved L2 fluency and comprehensibility of child learners of English in grades three and four. Hardison (2005) found that focused prosody training based on learners' own oral L2 productions – integrating the video of their speech event with visual displays of pitch contour – significantly improved the use of English prosody in the novel natural discourse of advanced Chinese adult learners of English. Improvements in the production as well as perception of the L2 English /l, r/ contrast by Japanese adult learners through training have also been shown in various studies (Hazan, Sennema, Iba, and Faulkner, 2005) with successful long-term retention (Bradlow, Akahane-Yamada, Pisoni, and Tohkura, 1999). Bradlow, Pisoni, et al. (1997) investigated the effect of perception training on production. Japanese speakers received an extended period of perception training to learn the English /r/-/l/ contrast. They carried out pre- and post-tests, whereby participants produced English words contrasting /r/ and /l/. All their participants demonstrated significant perceptual learning when the post-test tokens were compared to those of the pre-test. Their findings also show that English listeners accurately identified the learners' post-test tokens more often than those of the pre-test. However, this improvement was not consistent in all their trainees, and one trainee, despite an improvement in perception, showed no improvement in production. More recently, there has been a shift in trends to test production and the effect of production training on perception and production. Hattori and P. Iverson (2009) argue that difficulties in L2 learning for Japanese learners of L2 English /r/ and /l/ are the result of single category assimilation.[7] They assessed their Japanese participants' phonemic identification of English /r/ and /l/, the assimilation of English /r/ and /l/ into the learners' L1 Japanese /ɾ/, and additionally their production of English /r/ and /l/. Their findings show that Japanese learners of L2 English assimilate L2 /l/ into their L1 /ɾ/ more often than they do with L2 /r/. The assimilation patterns were category goodness difference or uncategorised vs. categorised assimilation, but the assimilation patterns did not follow a single category assimilation as they had initially hypothesised. Furthermore, they argue that the phonetic analysis of production is one way to measure category assimilation given that a speaker's production of English /l/ is acoustically distinct from the L1 /ɾ/. They posit that single category assimilation may potentially have a significant part in influencing the production of /ɹ/-/l/. However, it can be counter-argued that English /r/ and /l/, and Japanese /ɾ/ do not have comparable acoustic dimensions nor does speech production consistently and reliably correlate with speech perception (Bradlow, Pisoni, et al., 1997). Their results provide minimal support for a strong correlation between category assimilation patterns and the perception and production of English /r/ and /l/. Kartushina, Hervais-Adelman, Frauenfelder, and Golestani (2015) presented a training method that made use of real-time acoustic analysis of vowels produced by L2 learners. The acoustic, trial-by-trial, visual feedback is accessed by the trainees to inform their articulation and allow them to compare it side-by-side with that of native speakers. They used, '[t]he Mahalanobis acoustic distance between non-native productions and target native acoustic spaces' (2015: 817) to evaluate the accuracy of their trainees' production. Their findings demonstrate that the production of four L2 Danish vowels improved after only one hour training per vowel and that the L2 Danish vowels post-training were closer to the targets compared to pre-training.

---

[7] See 2.5.2 below the Perceptual Assimilation Model predictions.

This improvement was not only seen in production but also perception albeit the correlation analyses showed no relationship between training-related changes in perception and production. They concluded that this training methods was effective in improving production, that production training improves perception, but improvements in perception and production do not proceed at equal rates within each learner. Kartushina and Martin (2019) further looked into training to produce the L2 French /e/ and /ɛ/ vowel contrast by Spanish speakers. They compared performances between a group trained with multiple-talker input and another group with single-talker input to establish the impact of variability on production learning. Their findings revealed an improved in production accuracy for both training paradigms, although the multiple-talker variability assisted the learners in establishing abstract phonemic categories in production. Furthermore, children have been found to exhibit substantially more improvement in the single talker training paradigm (Giannakopoulou, Brown, Clayards, and Wonnacott, 2017). These studies generally show that production training training benefits production but does not transfer well to perception. However, none of these studies have compared the effect of delayed production with instant production practice. Herd, Jongman, and Sereno (2013) included three training paradigms in their study: training in perception, production, and perception and production combined. All three training paradigms were carried out over six sessions over the span of two to three weeks. In each session - which lasted 20 to 30 minutes - in the perception training, the trainees took part in a forced choice task. They heard a stimulus which contained /ɾ/, /r/, or /ð/ and were simultaneously presented with two orthographic choices. Once they chose their response, they received feedback on whether their choice was ‘right’ or ‘wrong’. In the production training, the trainees were presented with 90 minimal pairs and in each session, they practised one contrast only. They saw a waveform, spectrogram, and the written form of each test word (via Praat (Boersma and Weeninck, 2016)) as produced by a native speaker. They were instructed to inspect the production, then record their production of the word when ready. The program then allows the trainee to compare the two waveforms/spectrograms and proceed to the next item when ready. The trainees were prompted to repeat their attempts to match the native speaker stimulus without being allowed to hear any of the native speaker stimuli. In this training paradigm, the first session was devoted to teaching the trainees to identify distinguish /ɾ/, /r/, and /ð/ on a waveform and spectrogram. This procedure took up half the time of the first session which lasted 60-75 minutes. The remaining five sessions lasted 35-45 minutes. The last training, the combination training paradigm, involved perceptual and production training: three perception training sessions and three production training session which took place over two to three weeks alternating in modality from one session to the next. The trainees practice one contrasted at a time, each for two times over the training period. A fourth group was a control group comprising 11 L2 Spanish learners who had no training. The results of this study showed that whilst all the three training paradigms were effective, the perception only and production only training paradigms improved the learners' perception, and the combined perception and training paradigm improved the learners' production. Another study by Sakai (2016) also tested the impact of production only training whilst examining Spanish L2 learners of English /i/ and /ɪ/. Her results revealed that the listen only training resulted in great improvements in perception, but improvements in production were small. The production only training led to medium-size improvement in perception but no great gains in production. Combined perception and production training improved production, but when compared with production only training did not show differences in perceptual gains. Herd et al. and Sakai's studies successfully isolated the effect of production only training. However, this modality is unnatural and can only be applied in laboratory settings and may not be suitable with children.

In terms of amount of L2 input studies varied drastically and generally, the aspect of L2 learning determined the duration of training. For some studies, an intervention using computer-assisted pronunciation training lasted 20 minutes only (Guilloteau, 1997). In other studies, a classroom intervention lasted 70 hours (Parlak, 2010). Duration of training in computer-assisted pronunciation training studies tends to be relatively shorter than classroom-based interventions (Thomson and Derwing, 2014). In the aforementioned studies, there was no control group to reliably account for duration as a predictor of successful L2 learning. However, Flege and Fletcher (1992) argue that amount of exposure to L2 or as they refer to it as 'number of years of English-language instruction' has an important role in L2 pronunciation albeit it only accounted for 5% of the variance in the accent ratings of Spanish learners of L2

English.

Even with exposure to a large amount of input, certain aspects of the target language develop slowly (Swain, 1988) if not at all (Akahane-Yamada, Tohkura, Bradlow, and Pisoni, 1996) especially, with forms that are hard to notice. For this reason, Sharwood Smith (1993) proposes that certain aspects of the target language input require raising the learner's meta-linguistic awareness for forms to become part of the *intake*.

Several pronunciation studies assert the role of awareness raising either through socially constructed metalanguage and critical listening (Couper, 2011), self-monitoring techniques (Ingels, 2011), hyper-pronunciation (Nagamine, 2011), or through oral practice, speech monitoring, comparing performance with other models, changing the performance to match models, and practising the changed performance aloud until fluent (Sardegna, 2011). Swain (1995; 2005) argues that one of the aspects that plays an important role in noticing is production practice. Language production enables learners to notice differences between their output and the target language input allowing them to consciously reflect on their learning (Colantoni and Steele, 2008). However, Swain was most likely referring to aspects of L2 learning in general and not L2 speech learning in particular. Sumdangdej (2007) tested the effect of meta-phonological consciousness raising on the pronunciation of Thai child learners of English (aged 6;11–11;1). His meta-phonological consciousness raising technique was providing an explanation or making salient certain key features to learners in Thai. He argued that its impact may vary between adults and children. His findings show that adding input enhancement techniques can slightly improve the impact of getting native input for syllable structure but no difference was observed for stress acquisition. Other techniques that may enhance the input include explicit instruction and feedback (Sharwood Smith, 1993). For example, Kenworthy (1987) posits that to achieve target-like pronunciation, ideally the learners are explicitly introduced to the sounds and sound patterns of the target language followed by repetition practice. Feedback through speaking practice as shown later in section 2.5.5 was argued to allow learners to reflect on their performance, which then encourages them to revise and correct their productions in a continuous process to match the target language until they feel there is an agreement between their production and the model input.

In terms of training, Kuhl and P. Iverson (1995) for example, argue that the boundaries between perceptual categories are not lost for life even after puberty. In fact, extensive training can potentially improve perceptual discrimination of foreign language contrasts. Several studies examining the impact of formal instruction on L2 pronunciation have not provided solid results for pronunciation teachers because they do not specify instructional variables (Elliott, 1995; Flege, Munro, et al., 1995a; Flege, Yeni-Komshian, et al., 1999; Thompson, 1991).

Most, if not all, of the above studies[8] lack an explicit theoretical framework. Similarly, Thomson and Derwing (2014) argue that most L2 pronunciation teaching studies do not make use of control groups or comparison groups. Some studies who claim to have used control groups when instead, they used comparison groups. Finally, Settinieri (2008) pointed out that L2 learners generally prefer training that involves individual attention, focus on segmental and suprasegmental features, language learning awareness raising, and authentic meaningful tasks alongside visual support. Piske (2017) argues for five essential elements for successful L2 attainment, that is a) continuous intensive exposure to the L2, b) a large amount of authentic and (almost) native-like input, c) the opportunity to use the target L2 language by actively speaking (and writing) as frequently as possible in versatile contexts, d) skill-targeted instruction that supports structural differences between the L1 and L2 especially in heterogeneous classrooms, and finally e) motivating learning environments.

2.4 Literacy and orthography in L2 speech learning

In the models discussed in section 2.5, L2 speech learning was approached using a lens of perception, articulatory constraints, and production. Research has shown a difference between modalities of input relevant in making such comparisons. Rosenblum (2008) argues that speech perception is multimodal, that is learners not only rely on auditory cues, but also visual cues such as the written text – in other words orthography. In L1 acquisition, input is typically restricted to auditory mode and potentially visual mode such as facial cues, whereas in adulthood, input is bimodal involving simultaneous auditory and visual exposure in the form of orthography and

---

[8] An exception is Couper (2011), who utilises a cognitive phonology framework.

potential facial cues (particularly those relevant to visible articulation such as lip-rounding). Evidence for this robust demonstration is the illusion of the McGurk effect, whereby an auditory /ba/ presented alongside a synchronised mouth movement depicting “ga” is perceived by the listener as /da/, a different syllable altogether, known as the *McGurk percept* (McGurk and MacDonald, 1976). Most adult L2 learners, unlike L1 learners, have literacy skills (reading and writing) when they are first introduced to the target language. There has been an increase in research investigating the role of orthographic input on L2 learners’ lexical and phonological development (Bassetti, 2008; Escudero and Wanrooij, 2010). Several studies have also demonstrated the effect of native orthography on L2 pronunciation in the absence of orthographic information (Escudero, Hayes-Harb, and Mitterer, 2008; Weber and Cutler, 2004). Bassetti (2008) argues that written text serves as a visual representation of language.

Orthographic input (Young-Scholten, 2002) can have two types of effects. It can facilitate or inhibit the L2 learner’s perception, production (Bassetti, 2008; Showalter and Hayes-Harb, 2015), and word recognition (Cutler, 2015). Evidence of positive influence of L2 orthographic input comes from various research findings (Escudero, Hayes-Harb, et al., 2008; Escudero and Wanrooij, 2010). For example, Steele (2005) compared the effect of auditory vs. auditory and orthographic input on the pronunciation of the French uvular fricative by Mandarin beginner learners of French. He argues that the auditory only condition perceived the consonant as aspiration, whereas the auditory and orthography condition saw that the test items contain two consonants and thus pronouncing it as such. L2 orthographic input has also been claimed to facilitate establishing phonological contrasts which are not present in the target language (Bassetti, 2017).

Negative effect of orthographic input, on the other hand, leads to non-target-like pronunciations. Bassetti (2008) identifies several types of orthography-induced non-target-like pronunciations. One common type, she identifies is spelling pronunciations. In this type, L2 learners pronounce silent letters such as /l/ in ‘walk’ and /b/ in ‘thumb’ (Bassetti and Atkinson, 2015; Browning, 2004). Other types include phoneme additions such as vowel epenthesis in consonant clusters which violate L1 syllable structure (Young-Scholten, 1998; Young-Scholten et al., 1999). The latter case is especially interesting. When faced with complex consonant clusters, adults and children resort to different types of repair mechanisms. Children acquiring languages with complex consonant clusters have a tendency to omit consonants (Weinberger, 1987) whilst adults acquiring L2 languages with complex consonant clusters resort to vowel epenthesis (Young-Scholten, 1998) in what Weinberger (1987) argues to be the result of *recoverability*, that is retaining all consonants in a cluster. Young-Scholten et al. (1999) explain that vowel epenthesis evident in adult L2 phonology is the result of orthographic input. Their findings show that when adult L2 learners are presented with orthographic input alongside auditory input, they resorted to epenthesis in producing L2 complex consonant clusters. However, when presented with auditory input only, their productions reflected those of L1 child acquisition, that is cluster simplifications. Another type of orthography-induced non-target-like pronunciation Bassetti identifies is phoneme omission. Bassetti (2007) investigated the production of certain Chinese diphthongs and triphthongs by Italian learners of L2 Chinese. Her results show that learners produce vowels successfully when they are represented in pinyin spelling. However, they tend to omit the same vowels when they are not present in the pinyin spelling. A similar finding was observed in Korean learners of L2 English (Lee, 2004), where orthography interacts with the learners' L1 phonology. Phoneme substitution is yet another type of orthography-induced non-target-like pronunciation. According to Bassetti, Sokolović-Perović, Mairano, and Cerni (2018: 578),

> substitutions are often associated with incongruences between grapheme-phoneme correspondences in the first and second language, whereby the two languages map the same grapheme (single letter, digraph or trigraph) onto different phonemes.

A study by Vokic (2011) revealed that Spanish learners of L2 American English produce intervocalic [t] and [d] which are otherwise produced as flaps by native American English speakers. She argues that the reason for this is the orthographic representation of flaps in written text. Another study by Zampini (1994) showed that English learners of L2 Spanish sometimes substitute the Spanish phoneme [v] for /b/ when it is represented by the letter 'v' in writing even though the Spanish phoneme inventory does not include /v/. Zampini attributed this to incongruence in grapheme to phoneme correspondence. Similar orthography-induced substitutions were found in other language backgrounds varying in congruence in grapheme to phoneme correspondence such as L1 Italian learners of L2 English (Kenworthy, 1987; Speck, 2002), L1 French learners of L2 Italian (Costamagna, 2000), L1 Italian learners of L2 English (Piske, Flege, MacKay, and Meador, 2002), L1 Italian learners of L2 Chinese (Pinyin) (Bassetti, 2006), L1 English learners of L2 German (Young-Scholten, 2002). (In)congruence in grapheme to phoneme correspondence between the learner's L1 and L2 languages has been found to have a facilitating or hindering effect on performance. Escudero, Simon, and Mulak (2013) concluded that exposure to orthographic input in training has a facilitating impact on performance on minimal pairs with congruent orthography, but has a negative impact on performance on minimal pairs with incongruent orthography.

The facilitating impact of orthographic input can occur at any stage of L2 acquisition leading to an observable difference between preliterate children's L1 and/or L2 phonological acquisition and adults L2 phonological acquisition (Bassetti, 2008). For adult L2 learners, orthographic input enhances warped perception warranting learners the production of sounds they cannot perceive accurately. However, for preliterate children acquiring either their L1 or L2, perceiving sounds and sound contrasts accurately is a precursor for accurate production (Bassetti, 2008).

2.5 Theories of cross-linguistic influence

The findings of a great many earlier studies conducted from the seventies and onward indicate that the learners' L1 has considerable influence on their acquisition of an L2 phonology. Where studies are of perception, they indicate that foreign sounds are perceived through the ears of L1 phoneme categories whether it is the case for inexperienced listeners or second language learners (Abramson and Lisker, 1970; L. Williams, 1977; Werker and Tees, 1984). This prompted researchers to propose mechanisms that account for mapping non-native sounds to existing L1 phoneme categories. Models agree that the L1 system affect how an individual perceives and potentially produces non-native sounds.

Accounts for sources of difficulty faced by the L2 learner comprises relative markedness, L1 knowledge, perception, and functional constraints, that is perceptual salience and frequency (Colantoni and Steele, 2008). Relative markedness deals with implicational universals and typically investigates phonological processes and developmental sequences relevant to the phonological aspects of the present thesis. The second source of difficulty, perception, has been the basis for explaining cross-linguistic influence by most L2 speech learning models. Learners fail to produce target language (TL) sounds because they are interpreted inaccurately. Non-TL-like phenomena are especially vivid during the early stages of L1 transfer. In this chapter, models predicting segmental acquisition difficulties will be discussed in the light of various studies. Models of L2 speech learning include the Perceptual Assimilation Model (Best, 1994; 1995 and Best and Tyler, 2007), the Second Language Perception Model (Escudero and Boersma, 2004), the Speech Learning Model (Flege, 1995) and the Magnet Effect (Kuhl, 1991; Kuhl, K. Williams, et al., 1992). A hybrid model (Colantoni and Steele, 2008) incorporating the various aspects of speech learning models is dealt with especially in accounting for the role of feedback and delayed production. In the next section, practical issues relating to the terminology used in each model, the different stages of acquisition each models builds on are tackled. Last but not least, predictions for the current thesis in relation to the models will be made. The following models explain how this process takes place in the learners' perceptual system.

2.5.1 *Magnet Effect and Native Language Magnet Model*

This model proposed in Kuhl (1991), Kuhl, K. Williams, et al. (1992), and Kuhl and P. Iverson (1995) posits that a child is born with a universal auditory ability that allows him/her to distinguish differences in all speech sounds. By the end of the first six months of an infant's life and with linguistic experience from ambient auditory input, this ability gradually decreases and becomes language-specific. This ability continues to diminish as humans reach adulthood. Linguistic experience changes the mechanism which underlies the perception of speech (Kuhl and P. Iverson, 1995). The infant's perception becomes attuned to sound contrasts in his/her native language and ignores non-native sound contrasts. As a result, proto-type categories are created during this time of an infant's life. A proto-type is considered the best exemplar of a native sound category (Kuhl and P. Iverson, 1995: 123). Once proto-types are established, each serves as a perceptual magnet (Kuhl, 1991). Kuhl and P. Iverson (1995: 121-2) explain, that

> the magnet effect shows that exposure to a particular language results in a distortion of the perceived distances between stimuli; in a sense, language experience warps the acoustic space underlying phonetic perception.

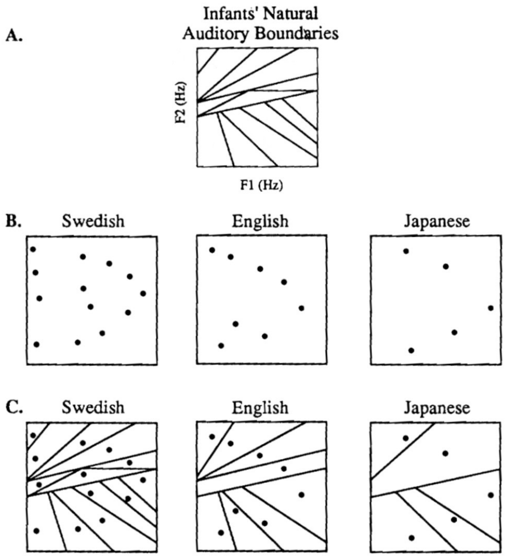


Figure 2.1 The three phases of phonetic representation in an infant's first year of life adopted from Kuhl and P. Iverson (1995: 140)

The perceptual magnet effect and its results are integrated into the Native Language Magnet model (NLM) (Kuhl and P. Iverson, 1995). Ambient language input eventually affects the child's perception and production of human speech (Kuhl, 1994). Figure 2.1 includes three phases depicting changes in phonetic representation in an infant's first year of life. The first phase (A) is where the infant has universal boundaries, the second phase (B) depicts cross-linguistic differences in the three languages shown underlying perceptual representation as a by-product of different linguistic experiences by the end of six months. The last phase (C) illustrates the loss of some (universal) phonetic boundaries due to the perceptual magnet effect.

Kuhl and P. Iverson describe this as warping the perceptual space surrounding native sound categories, doing so by means of two opposing mechanisms. One mechanism is attracting any sound that is acoustically close to a proto-type, reducing the perceived distance between the proto-type and stimuli within its vicinity and eventually assimilating them into a single category. Thus when the L2 learner is faced with an L2 sound that is acoustically similar and/or close to a native sound proto-type, his/her discriminability between the two is diminished.

The other mechanism is maximising the perceived distance around the edges of a phonetic boundary, that is not in the vicinity of the proto-type and making it as dissimilar as possible (Kuhl and P. Iverson, 1995: 141). The acoustic signal embedded in an L2 sound contrast is processed through native language cue-weighting resulting in incorrect perceptual representations of the L2 sound contrast as well as longer processing time (P. Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, and Siebert, 2003).

Assuming Arabic voiceless plosives have a short VOT and RP voiced plosives also have a short VOT (see Chapter 3), it is predicted, based on the NLM model and the perceptual magnet effect, that the RP voiced plosives will be warped by and attracted to the native Arabic prototype for voiceless plosives especially when the second language learners in the present study are first exposed to the target language. Additionally, since Arabic does not utilise long VOT the way RP English does for voiceless plosives, the model predicts that long VOT in the L2 will potentially be lengthened even further as a result of the second mechanism proposed above.

Studies with an NLM framework have been carried out on vowel categories (Kuhl, K. Williams, et al., 1992; P. Iverson and Kuhl, 1995), on voicing in velars /k/ and /g/ (Davis and Kuhl, 1992; 1993; 1994), voicing in bilabial plosives in Hebrew and Arabic (Segal, Hejli-Assi, and Kishon-Rabin, 2016) and on place of articulation contrast in plosives /k/ and /p/ with an observed perceptual magnet effect in three to four-month-old infants (Miller and Eimas, 1996) and for both adults (P. Iverson and Kuhl, 1995) and infants (Kuhl, K. Williams, et al., 1992).

In their model, Kuhl and colleagues explain the effect of the perceptual magnet with reference to non-prototypic members of the category. Consider the following:

> The magnet effect causes other non-prototypic members of the category to be perceived as more similar to the category prototype than to each other, even though the actual physical differences between the stimuli are equal (Kuhl, K. Williams, et al., 1992: 606-7),

Kuhl, K. Williams, et al. (1992: 608) go on to say that the magnet effect of the prototype can reveal the reason older children and adults lack the ability to discriminate sounds of a foreign language when they approximate a prototype in their L1. An example of this is /r/ and /l/ for native Japanese-speaking listeners. Findings by Kuhl (1991) showed that in speech perception, phonetic prototypes act as perceptual magnets for other speech sounds within the category.

However, a criterion for what qualifies as a non-prototypic member of the same category, other sounds in the category, or resemblance of sounds to a native prototype was not clearly specified, especially considering the idea that what counts as a non-prototypic member of a category in one language is not necessarily non-prototypic in another. If we were to put the model to the test, it seems that the only way to decide on the range of acoustic values for either a prototype or what qualifies as a non-prototype for any given language is to carry out a perception task using native speakers of that language.

An important indication Kuhl, K. Williams, et al. (1992: 608) make is that phonetic prototypes are not the result of the appearance of contrastive phonology nor word learning. Instead, they argue that it is 'an underlying cognitive capacity and proclivity to store in memory biologically important stimuli and from the ability to represent information in the form of a prototype'. They also suggest that these native prototypes help infants establish categories for speech sounds before they start acquiring lexical meanings by the end of the first twelve months (Kuhl and P. Iverson, 1995; Kuhl, K. Williams, et al., 1992).

This model is based on first language experience and its effect on how naïve individuals respond to foreign stimulus. No claims are made regarding simultaneous bilingualism or developmental aspects of L2 experience. Kuhl and P. Iverson (1995) suggest that extensive training can potentially improve perceptual discrimination of foreign language contrasts as the boundaries between categories are not lost for life, even with adults learners basing their suggestion on studies carried out by Flege (1995), Logan et al. (1991), and MacKain, Best, and Strange (1981). They also indicate that the perceptual magnet effect contributes to difficulties in learning an L2 at a phonological level as indicated by Flege's Speech Learning Model,

> The difficulty posed by a given foreign-language unit will depend on its proximity to a native language magnet; the nearer it is to the magnet,

> the more it will be assimilated to the native language category, making it indistinguishable from the native language sound (Kuhl and P. Iverson, 1995: 143).

One point should be noted here is that the model makes assumptions for the initial exposure to foreign input analogous with the TL input at the first contact in the training process. It does not state how training can affect this and if it does, it does not state at which rate. If the learners do show signs of discriminable categories, it can be assumed that learning took place and that the training was effective. We return to this point further below.

2.5.2 *The Perceptual Assimilation Model*

PAM (Best, 1994; 1995) also makes predictions for difficulty in cross-linguistic perception. The model posits that unfamiliar non-native speech sounds are perceptually assimilated to the most articulatory-similar sound category in the native language by naïve listeners. The non-native phonetic information naïve listeners perceive depends on their native linguistic experience and language universal perception sensitivities (Best and Tyler, 2007: 17). However, the model adds that perception of speech is continuously refined even for the native language throughout the lifespan.

The model follows a direct-realist approach which bases its assumptions on articulatory gestural properties, defined in relation to a combination of audio-visual cues[9] (Best, 1995: 181-2), vocal organs, place and degree of constriction. A phoneme is considered a constellation of these articulatory variants, which serves a linguistically-contrastive function. PAM maintains that non-native phoneme discrimination declines when phonetic contrasts involve the same/similar articulatory gestures as in the native language. This articulatory similarity can be manifested in spatial-proximity of constriction, place and/or the articulators. This decline in perception results in non-native sounds being assimilated to the learner's native sound categories. The degree of assimilation determines how well a learner perceives these non-native contrasts. A non-native speech sound will be heard as either a good exemplar of a native category or a poor exemplar and thus eventually categorised. Alternatively, it will be heard as different from any native sound category and thus

---

[9] It seems from her explanation that she meant lip-rounding as an example.

uncategorised. A rather unusual case is when the listener considers the non-native sound a non-speech sound, in which case it cannot be assimilated to any native category and, therefore non-assimilated. It predicts relative difficulty in perception by comparing L1 to non-native sound contrast in that a non-native contrast assimilated to a single L1 category, called single-category assimilation, is predicted to be harder to acquire because it requires the learner to establish a whole new category. A TL contrast that is assimilated to two L1 categories, called two-category assimilation, is predicted to be easier to acquire because all the learner is required to do is adjust the boundaries of the pre-existing L1 category. In case of single category assimilation, one member of the contrast might be a better exemplar of the L1 phoneme than the other, a case referred to as category goodness. Discrimination is predicted to be intermediate between single category and two category.

One study which validates the prediction that two category contrasts are easier (near-ceiling accuracy) to perceptually discriminate – even when both members are not present in the native phonemic inventory – is that by Best, McRoberts, and Goodell (2001) . They tested English listeners with consonantal contrasts in Tigrinya and Zulu to test the model's predictions of accuracy of discriminating non-native contrasts and assimilating them to native categories. The non-native Zulu contrast (the lateral fricatives /ɬɛ/-/ɮɛ/) were tested alongside two native English contrasts (/sɛ/-/zɛ/ and /ʃɛ/-/ʒɛ/), which were chosen on the basis that they shared the same spelling in English as well as involving the same active articulators. Participants showed a higher discrimination accuracy for two category assimilation (lateral fricatives) followed by category goodness (velar stops /ka/-/k'/) and the least discrimination accuracy for single category assimilation (bilabial stops /bu/-/ɓu/), which confirms PAM's predictions. The researchers also conclude that not all non-native contrasts pose the same difficulty. Rather, it depends on the assimilation scenario.

The model also predicts a few more scenarios. The uncategorised-uncategorised scenario is a case where each member of the TL contrast can be assimilated to approximately the same degree to two or more native sounds. The model does not make predictions regarding the degree of assimilation being either poor or good but it could depend on the phonetic proximity to native categories. Additionally, the categorised-uncategorised scenario is a case where one sound in the

TL phonemic contrast can be assimilated to only one native sound and the other sound in the contrast can be assimilated to two or more native sounds. Finally, the non-assimilable scenario is a case where the members of a non-native contrast do not resemble any native sound and cannot be assimilated to any native phoneme category.

A study which tested all assimilation scenarios of non-native vowels is by Tyler, Best, Faber, and Levitt (2014). They tested thirteen L1 American-English listeners for assimilation in the form of categorisation and rating goodness of fit patterns of six non-native vowel contrasts (from three different language groups, vis-à-vis Romance – French, Germanic – Norwegian, and Tai - Thai). Results supported PAM assimilation types even for vowels, although large individual differences have been observed. Most of the contrasts were assimilated as two category, uncategorised-categorised, or uncategorised-uncategorised, with the first two scenarios, discrimination being at ceiling. single category assimilation scenarios were observed for Thai /i/-/y/ vowel contrasts only.

Even though PAM was originally developed for cross-linguistic perception of a foreign language by naïve listeners, it was later developed to account for developmental stages in L2 learning using the same principles, PAM-L2 (Best and Tyler, 2007).

PAM extension to PAM-L2 reasons that phonology plays a pivotal role in the development of L2 perception as learners develop an L2 system or as Best and Tyler (2007) call it an IL system in a way that is inaccessible for naïve listeners perceiving non-native speech sound contrasts. The extended model explains that the main source of difficulty in accurately perceiving L2 contrasts is due to L2 learners having a shared phonological space for L1 and L2 categories (Best and Tyler, 2007: 26). The model argues for a common phonological space for L1 and L2 and that learning occurs at both the phonetic and phonological level. It links the phonetic level to the phonological categories through the concept of goodness of fit as pointed out earlier. The levels of phonetics and phonology dynamically interact between L1 and TL and this interaction may impact trajectories for learning to perceive TL sounds and distinguish them from native speech categories and also how this can mutually have an impact on native sound categories/contrasts (Best and Tyler, 2007). In a TL contrast if only one member is assimilated to an L1 category into a common IL category, it is modified slightly to subsume phonetic detail of both L1 and TL categories. Because the TL category is perceived as a good exemplar of the L1 category, not a great deal of modification will take place. In this case, the phonetic details of the IL category will most largely resemble those of the L1. However, the phonetic detail of TL category should not be grossly different from that of the L1 for this to occur. On the other hand, if the TL member is phonologically comparable to that of an L1 category but is phonetically different enough, the difference will be noticed and a new category will be formed. However, as is the case for the previously discussed model (NLM and the Magnet Effect), the lack of clear and precise criterion for measuring similar and different, makes it challenging to accurately predict assimilation scenarios assumed by the model.

PAM-L2 includes a postulation that neither of the above models have regarded. It argues that the perceptual events in question depend on the listener's perceptual objectives or levels of attentional focus. This in turn varies from requiring attention at the gestural level to either the phonetic or phonological level. L2 perceptual learning does not depend solely on the phonetic level, rather, on all three levels contingent to the context of goals.

2.5.3 *The Speech Learning Model*

SLM (Flege, 1995) deals with single segments but has opposing predictions whilst at the same time accounting for extra-linguistic factors such as age of exposure to TL correlating with L1 proficiency at the time of TL exposure, amount of L1 use and L2 proficiency correlating with amount of input. Similar sounds are harder to acquire and different ones are easier. If two sounds from TL and L1 are perceptually similar enough, they are equated perceptually. 'Equivalence classification is a basic cognitive mechanism which permits humans to perceive constant categories in the face of inherent sensory variability found in the many physical exemplars category' (Flege, 1987d: 50). Equivalence classification blocks the creation of a new TL category (Flege, 1987a: 94). In this case, a merged category that equates the L1 and TL categories is used to process both sounds in a common phonological space and affects the production of both L1 and TL sounds. Predictions of SLM are borne out in a series of studies investigating both consonants (MacKay et al., 2001) and vowels (Bohn and Flege, 1992; Flege, 1987d; Flege, Navarra, Sebastián-Gallés, and Soto-Faraco, 2003).

Evidence for a merged category is found for both voiceless stops /p t k/ (Flege and Hillenbrand, 1984; Flege, 1987c; Flege, 1988) and voiced /b d g/ (MacKay et al., 2001) where a VOT category does not resemble that of either monolingual group. In their first study, data were based on actual measurements that showed longer L1 VOT values compared to L1 monolingual productions and shorter L2 VOT values compared to monolingual TL productions. In their second study, data were of two types, production and perception. Production data were examined acoustically. VOT was classed into two categories, i.e. lead-lag and short-lag. Data were compared in terms of frequency of pre-voicing versus frequency of short lag productions. The study also considered variation in terms of age of exposure and L1 use to explain amount of phonetic learning amongst the participants. Kang and Guion (2006) also investigated stop acquisition of early and late Korean-English adult bilinguals with an onset of learning of 3;8 years and 21;4 years respectively by measuring VOT, f0 and amplitude differences between the first two harmonics compared to English and Korean monolinguals. Results showed that late bilinguals did not resemble English monolinguals in any of the properties in all stops but were different from the Korean monolinguals in VOT of fortis stops. Early bilinguals did not differ from monolinguals from either language. They suggest independent systems for early bilinguals and a merged system for late adults. Here, the L1 properties of the two groups and the TL properties were not a variable as suggested by PAM and L2LP, yet developmental paths varied as a result of age of exposure. Findings by Khattab (2002b) regarding independent systems showed that VOT values of her 5 to 10-year-old bilingual participants did not match either monolingual group despite being exposed to both Arabic and English simultaneously. Lead-lag Arabic VOTs were replaced with short-lag ones. She argues that a similar pattern was exhibited by Arabic child monolinguals and was thus attributed to developmental rather than cross-linguistic influence.

Similarly, SA Lee and G. Iverson (2012) measured VOT and vowel-onset f0 in word-initial stops of Korean-English bilinguals in two age groups, five and ten-year-old children with an age of exposure ranging between 1;6-2;10 and 2-4;6 respectively to determine whether bilinguals have a shared or independent phonological systems. They found that 5-year-olds distinguished stops based on VOT only whereas 10-year-olds distinguished them on the basis of the two correlates measured. They interpreted this as 5-year-olds having partially shared stop systems and that systems continue to evolve during development period. Here, it can be argued the five-year-olds have not reached phonetic maturity at this age. Even though sound categorisation begins in infancy, it may take up to the age of 12 to reach adult-like form (Hazan and Barrett, 2000). Also, the amount of input as measured by length of residence varied in favour of the ten-year-old group, which suggests that amount of L2 input is an intervening factor.

2.5.4 *Second Language Perception Model*

Unlike PAM, in L2LP (Escudero and Boersma, 2004) predictions are designed for vowels and based on acoustic comparisons between L1 and TL especially that similarities and differences in vowels are harder to discern in terms of articulatory gestures and due to the effect of coarticulation with the preceding consonant. It proposes that native speakers are optimal perceivers as they have a perception grammar, which is used in parsing continuous auditory input and mapping it to perceptual representations of phonemes and allophones. The extent of variability in the L1's acoustic values of a given perceptual representation influences (and can predict how) their native perception grammar and how a TL category is perceived.

Common Grounds for PAM and L2LP

PAM/PAM-L2 and L2LP both acknowledge the role of L1 transfer in L2 segmental learning. They focus on differences in phonemic contrasts instead of individual segments. They both predict relative difficulty in perception by comparing L1 to TL sound contrast in that a TL contrast assimilated to a single L1 category (single-category assimilation single category (PAM); new scenario (L2LP)) is harder to acquire because it requires the learner to establish a whole new category. A TL contrast that is assimilated to two L1 categories (two-category assimilation two category (PAM); similar scenario (L2LP)) is easier to acquire because all the learner is required to do is adjust the boundaries of the pre-existing L1 category. (Consider the diagram in figure 2.2 designed by the author).

In the case of single category assimilation, one member of the contrast might be a better exemplar of the L1 phoneme than the other (category goodness). Discrimination

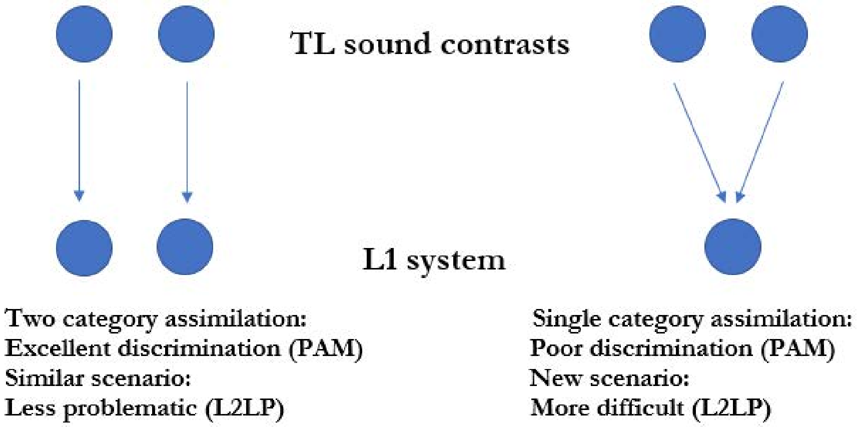


Figure 2.2 Mapping TL contrasts to native phonemes

is predicted to be intermediate between single category and two category.

The predictions for naïve listeners in PAM and L2LP resemble the initial state in L2 perception. The ones relevant to the study are: In a TL contrast if only one member is assimilated to an L1 category into a common interlanguage category, it is modified slightly to subsume phonetic detail of both L1 and TL categories. Because the TL category is perceived as a good exemplar of the L1 category, not a great deal of modification will take place. In this case, the phonetic details of the IL category will most largely resemble those of the L1. However, the phonetic detail of TL category should not be grossly different from that of the L1 for this to occur. On the other hand, if the TL member is phonologically comparable to that of an L1 category but is phonetically different enough, the difference will be noticed and a new category will be formed.

PAM-L2 also predicts that audio input is not the only cue in perception. In the case of single category assimilation (a hypothetical example would be TL /p b/ assimilating to L1 /b/), learners can rely on orthographic cues to tease apart a distinction that cannot be made perceptually. Lexical minimal pairs may provide cues to the phonemic distinction for the two sound contrasts (Bundgaard-Nielsen, Best, and Tyler, 2011). In the present study, assuming learners are not able to discern differences perceptually say between /p/ and /b/, orthography supported with meaning contrasts in the form of pictures in minimal pairs such as ‘bad’ and ‘pad’ should provide cues other than the auditory percept and thus helps motivate learners to create new categories for them.

A study comparing the perception of Southern British English vowels by Salento

Italian and Peruvian Spanish listeners who have a five-vowel inventory (Escudero, Sisinni, and Grimaldi, 2014: 1577) but varying in their vowels' acoustic implementation, shows that the listeners perceived (by means of categorising stimulus vowels) Southern British English vowels differently (Escudero, Sisinni, et al., 2014). Similarly, a study comparing the perception of two British varieties, Scottish and Southern English varying in the acoustic values for the /i/-/ɪ/ contrast by L1 Spanish learners, found that learners followed different developmental patterns depending on the TL variety. The Spanish-Scottish English group distinguished the TL contrast based on temporal differences which are not used in their L1. The Southern English /i/ is longer and higher than /ɪ/. However, the Spanish-Southern English group where the TL contrasts varied in F1 frequencies, a cue which correlates with fronting perceived the contrast as a similar one in Spanish /i/-/e/. The contrast was mapped to a single native category /i/. The different acoustic implementation between the contrasts in the two target varieties predicted different learning paths for speakers of a single L1 background (Escudero and Boersma, 2004).

Even though the two models make similar predictions, it must be noted that there are significant methodological differences between the two. The PAM relies on articulation which is more or less categorical and is typical of consonants. L2LP, on the other hand, measures differences and similarities using raw acoustic values and restricts its territory to vowels. Nonetheless, its tenets should in principle be applicable to consonants and their acoustic correlates.

2.5.5 *A Hybrid Schemata of Segmental Acquisition*

A comprehensive theory that explains perception and production difficulties has not yet been put forward. We assume not 11 years later either. However, Colantoni and Steele (2008: 522-3) have incorporated aspects from principles of L2 speech learning, aspects of perception models and principles from phonetic production research into a hybrid schemata of segmental acquisition.

The schemata in figure 2.3 demonstrates that the TL input to which the learners are exposed triggers perception, where the learner compares the segmental properties of TL to L1 acoustically and phonotactically. Less proficient learners rely more on the acoustic signal and less on the distributional patterns. Based on this comparison, the segment is then classified as old (already existing in the L1 inventory), similar to one or two existing L1 segment(s), or novel (cannot be associated with any existing L1 segment).

This stage of processing is based on the SLM by Flege (1995) and Flege, Navarra, et al. (2003). With more exposure, the learner creates a cognitive representation of the segment in his/her IL. The mental representation will then be used for planning articulatory gestures, which are position-sensitive. The more experienced the learner is, the more information s/he carries on differences in allophonic properties as a result of being exposed to more native input. Even if the learner has target-like representations and accurate motor planning, target-like productions cannot be guaranteed. This is because articulation will be shaped by L1 articulatory patterns as well as universal articulatory constraints such as aerodynamic (air stream-related) and elasto-inertial (related to articulators) constraints. These result in the produced output. The output will be either more or less accented and will feed back into the learner's perception alongside other native input. This allows the learner to continue to modify their productions and the mental category continues to evolve accordingly. If feedback from the learner's own output is perceived as being no different from their input, the categories become fossilised and thus will no longer evolve. In this case, it is said to have reached an end state.

Additionally, Colantoni and Steele (2008) postulate that variability in perception and production of speech may be affected by linguistic factors such as prosodic and phonological context, position in an utterance (Demuth and Song, 2011) and the differences in acoustic/ articulatory properties of sounds across languages. This is the case not only for L2 acquisition but also that for L1. In American (including African-American) and British English L1 acquisition studies on variable processes affecting coda clusters for example, such as coda deletion and coda cluster reduction in adults are found to be influenced by contextual variation in some dialects (Docherty, Foulkes, Tillotson, and Watt, 2006; Foulkes, Docherty, and Watt, 2005; Stockman, 2006). An explanation for this can be derived from Lindblom's (1990; 1996) Hyper and Hypo theory (H&H), which hypothesises that speakers adjust their (L2) speech precision based on prosodic dominance (Diehl and Lindblom, 2004). H&H theory predicts that speakers tend to hyper-articulate in stronger phonetic contexts, for

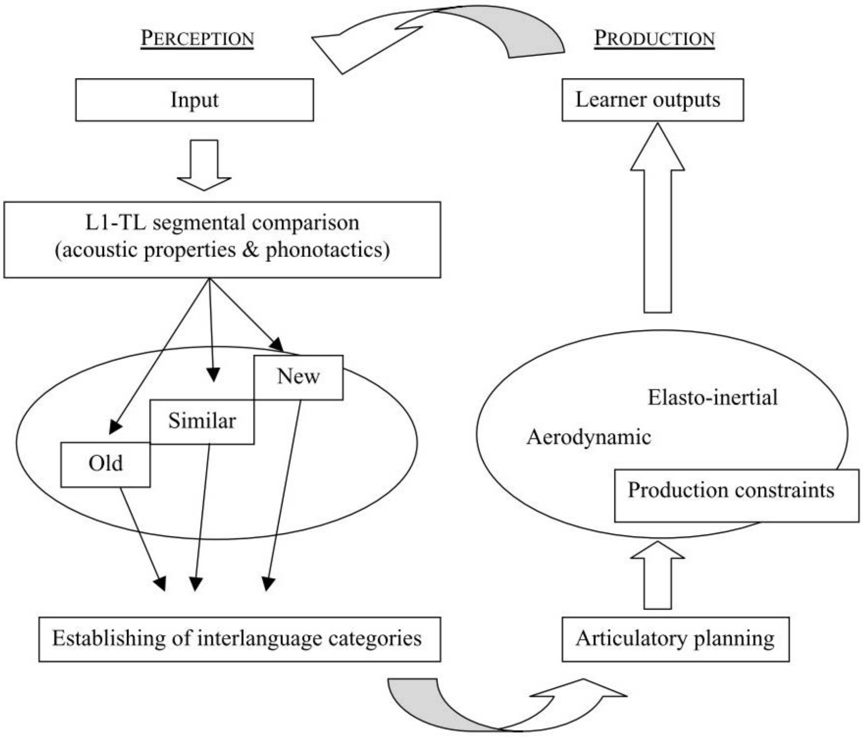


Figure 2.3 Colantoni and Steele's (2008) diagram of perception-production model of L2 segmental acquisition

example onsets, and hypo-articulate in weaker phonetic contexts such as codas.

Feedback has also been shown to impact first language category acquisition (M. Goldstein, Schwade, and Bornstein, 2009). In the case of delayed production, there will essentially be no output as this experimental group will not be required to practice production. It will, thus, not be affected by the learner's output, which feeds back into learner's perception. Krashen (1982) argues that the pressure to perform results in premature use of L2. Even though Krashen did not specify this for L2 pronunciation, it seems to be compatible with the role of the learner's own production in his/her perception.

2.6 Practical issues

One of the issues in examining theoretical assumptions is that crucial aspects underpinning these models are not well-defined. The concept of similar and different is not explicitly defined nor directly and/or reliably measurable. Major (2001: 39) indicates that criteria usually used for discerning similar-non-similar sounds include perceptual, acoustic, articulatory, native and non-native intuitions, listener judgements and occasionally orthography.

Acoustically speaking, variation is an inherent nature of human speech even within an individual (Lindblom, 1990; Lindblom, Brownlee, Davis, and Moon, 1992). Articulatory similarity can also be misleading since sounds sharing the same organ or degree of constriction can also vary as a result of coarticulation especially with vowels (Lindblom, 1963; Lindblom, 1983). Psycholinguistically, gross differences are more easily perceived as a result of perceptual salience, whereas minimal differences (similarities) are less often perceived. In psychoacoustics, the notion of just noticeable difference defined as the minimal perceivable change in an acoustic stimulus is a reliable criterion for measuring discernible differences. Sounds differing by an amount smaller than the just noticeable difference are heard as similar (Ghitza and J. Goldstein, 1982).

Models do not consistently account for other potential variables. Most studies designed with the above models investigate naturalistic acquisition differently from laboratory training. Stimuli from various speakers in various contexts are argued to be essential for maintaining long term representations in perception (P. Iverson, Hazan, and Bannister, 2005) and production (Bradlow, Akahane-Yamada, et al., 1999).

In addition, some models focus on the initial state and others focus on experienced learners and they do not consistently account for development of IL over time. SLM states that with experience, the learner can reach native-like norms. The PAM predictions for naïve listeners resembles that of the initial state in L2 acquisition.

2.7 Summary

This chapter provided a review of some prominent factors affecting L2 speech learning. It presented the role of implicational markedness constraints and how Sonority Dispersion affects the acquisition of L2 consonant clusters. It also examined the literature on L2 pronunciation teaching and the impact of literacy and orthography on L2 speech learning. The chapter also provided a review of some models proposed to explain mechanisms involved in speech learning. These models provide various explanations based on perception, production and the role of feedback. The literature on L2 speech learning raises several issues that require addressing. These include:

1. What is the role of delayed production in L2 speech learning?
2. What is the role of native vs. non-native input in child L2 learners?
3. What is the role of markedness and sonority in L2 speech learning?
4. What is the role of phonotactic context in L2 speech learning?

Chapter 3: Libyan Arabic and British English

3.1 Introduction

It is important to understand phonological and phonetic differences between English and Arabic if we are to measure learning based on some form of pronunciation training. Arabic is in the Afro-Asiatic Semitic family and English is in the Indo-European Germanic family. The two languages have different phoneme inventories and different aspects of phonological structure. Amongst their similarities is that Arabic has numerous varieties and so does English.

Misrata Libyan Arabic is chosen for several reasons: 1) familiarity with the variety's overall phonological structure given that it is the researcher's native language; 2) convenience for data collection since Misrata is the researcher's hometown; 3) the differences between Arabic and English phonology provide an ideal ground for investigating language acquisition; 4) to provide insight into L2 learning by Arabic child learners. Standard Southern British English was chosen as the target language because the English language materials used in the training software, the *Digital Literacy Instructor* are recordings made by two speakers of Standard Southern British English (Overal, 2014).

The following chapter gives an overview of the phonological and some relevant phonetic differences between the native language of the participants, that is, Libyan Arabic (henceforth LA) as spoken in the city of Misrata /misˤrata/ in section 3.2. A similar overview is provided for the target language, that is SSBE English in section 3.3. The subsequent section, 3.4 provides a very brief overview of the difference in writing systems between Arabic and English. The chapter concludes by presenting a brief summary of the differences between the two languages.

3.2 Characteristics of Libyan Arabic

The Libyan Arabic variety descends, as do all Arabic varieties from the variety of Quraish. It belongs to the Maghrebi group (Pereira, 2007; Versteegh, 2014). Throughout its history, Libyan Arabic has had contact with other languages including the native languages of Amazigh (Berber), Twareq (Tmasheq) and Tabu (being the native inhabitants[10] of Libya prior to the Arab conquests), Turkish (due to the colonisation by the Ottoman Empire (1551 - 1711) and the Karamnli Dynasty (1711 - 1835)), and Italian (due to the Italian occupation (1912 - 1947)), all of which have affected the local descendant in one way or another (Pereira, 2008) making it distinguishable from other descendants of Arabic varieties.

There are a few sub-varieties of the Libyan Arabic. The variety under scrutiny is that spoken in Misrata, Libya's third-largest city. It is inhabited by more than 380,000 people (World Popoulation Review, 2017), which creates further sociolinguistic variation within it. Sociolinguistic variation in Misrata is understudied. One of the few studies is an unpublished paper by Habara (2017). Two examples of sociophonological variation will be pointed out. In some areas in the city, for example Zammoura, Jhanat, Kerzaz, Tummina, Ramla, Skeirat, Magasba, and Ghiran, the pharyngealised dental plosive /dˤ/ is realised as a pharyngealised dental fricative [zˤ]. The latter is assumed to be the non-prestigious variant and is usually stigmatised (Habara, 2017). Consider the examples in table 3.1.

Table 3.1 /dˤ/-/zˤ/ variation

/dˤ/	non-prestigious variant /zˤ/	gloss
/dˤei/	/zˤei/	light
/ʔaɾədˤ/	/ʔaɾəzˤ/	land
/dˤuːg/	/zˤuːg/	taste (v.)
/bəjadˤ/	/bəjazˤ/	whiteness
/beːdˤa/	/beːzˤa/	white (f.)
/dˤəɾuːs/	/zˤəɾuːs/	molars

The distribution of this variant is evident in older speakers, in spontaneous speech (as opposed to read text) and is prominent in illiterate speakers of the community (Habara, 2017). This sound-shift, albeit its limited prevalence in Libyan Arabic of

---

[10] See Swei'y (2009) for a full record of the Arabisation of old Libyan language.

Misrata, is believed to result from contact with Berber (Habara, 2017). The Berber phonological system is characterised by a tendency towards spirantisation and generally a heavy use of fricatives and affricates (Chaker and Mettouchi, 2006).

Another example of socio-phonological variation is found in the vowel in the feminine plural bound object pronoun, which varies by geographical location within the city. In central parts of Misrata, the vowel is high front /-hin/ e.g. /diːɾ-**hin**/ ‘make them’, whereas in the rural part Qasr Ahmad, the vowel is low front /-han/ e.g. /diːɾ-**han**/. This is a personal observation, which has not been formally studied in this variety. Sociophonological variation in Libyan Arabic in general and Misrata Libyan Arabic in particular has caused some disagreement amongst researchers describing the phonemic inventory of the language. We return to this point below.

There are three main regions in Libya: Tripolitania (the North West region), Fezzan (the South West region), and Cirenaica or Barqa (the East region). See figure 3.1. Misrata falls under the Tripolitan region, which makes it a mixture of Bedouin (for example, the substitution of [g] for /q/) and urban phonological features (the loss of interdental fricatives /θ, ð, ðˤ/, which merged with [t, d, dˤ] respectively) albeit the prevalence of Bedouin features (Pereira, 2007). The koine is believed to have resulted

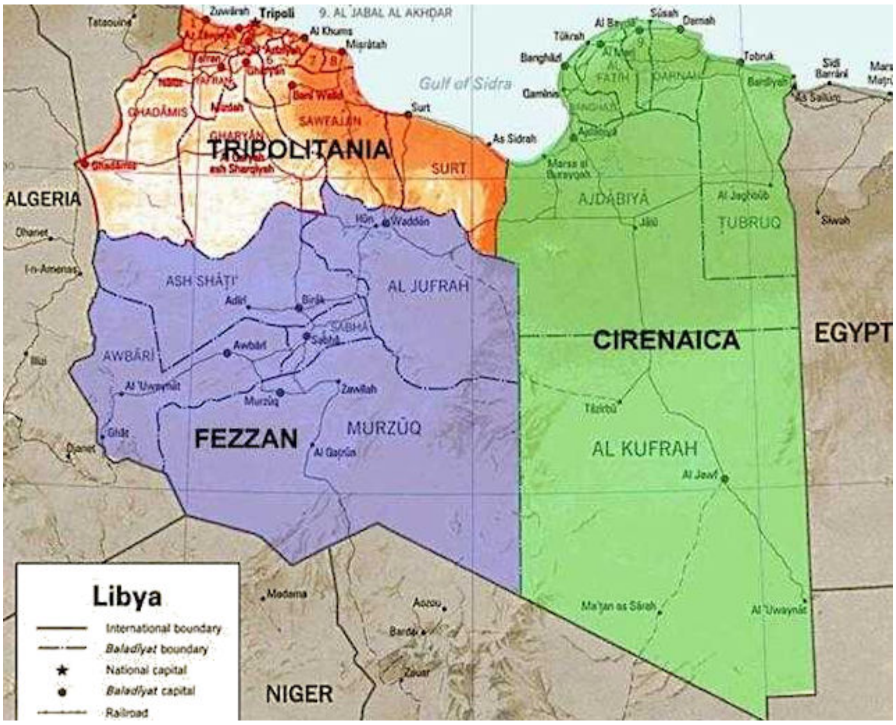


Figure 3.1 The three main regions in Libya: TRIPOLITANIA, FEZZAN, CIRENAICA

from the two waves of Arabicisation in North Africa, the first occurring in the seventh century (pre-Hilalian – sedentary/urban) and the second in the 11th century (Hilalian – Bedouin). The contact with Berber, has further affected the variety. For example, a Libyan speaker from the Tripolitanian region might use ‘hanna’ /ħanna/ from Berber or ‘gadda’ /ʒadda/ from Arabic to refer to grandmother. Moreover, the co-occurrence of a high variety [H] – Classical Arabic ‘al-fusˤħa’ – side by side with a low variety [L] – Libyan colloquial speech – each having a distinct role makes it diglossic (Ferguson, 1959). [H] is used in formal situations, for example in schools and broadcasting, whereas [L] is used at home, with family and friends and generally in less formal situations. Children are typically exposed to [H] in schools when starting their formal education (around the age of six) albeit exposure from TV cartoons and programmes prior to that is inevitable.

3.2.1 *Libyan Arabic Consonants*

Surveying the literature on Libyan Arabic, there is disagreement on the size of consonant inventory. In this section, we argue for the case that Libyan Arabic spoken in Misrata has 25 consonant phonemes including /w/. These are shown in table 3.2.

Table 3.2 Misrata Libyan Arabic consonant inventory

	B-Lab	Lab-Den	Den-Alv	Pal-Alv	Pal	Vel	Uvu	Pha	Glo
Plosive	b		t d			k g			ʔ
(emphatic)			tˤ dˤ						
Nasal	m		n						
Trill/Tap			ɾ						
Fricative		f	s z	ʃ ʒ			x ɣ	ħ ʕ	h
(emphatic)			sˤ						
Approximant			l		j				

B-Lab=bilabial, Lab=labial, Den=dental, Pal=palatal, Vel=velar, Uvu=uvular, Pha=pharyngeal, Glo-glottal.

Shitaw (2014) claims there are 27, Elramli (2012) and Laradi (1983) claim 28, Elgadi (1986), Fantazi (2003), and Garib (2014) claim 30, whilst Elfitoury (1976) claims 34 consonants.

Discrepancies in size stem from disagreement on the following sounds: /ʤ, θ, ð, ðˤ, v, ɫ, zˤ/. Several reasons can explain this discrepancy. First, /ʤ/ is part of Modern Standard Arabic and whilst it is maintained in some modern day varieties of Arabic, for example Kuwaiti Arabic (Holes, 2007), it is lost in Libyan Arabic and is realised as a fricative /ʒ/ only .

As for /θ, ð, ðˤ/, the sub-varieties of the Eastern region (Cirenaica) are Bedouin, thus the Bedouin feature, that is the interdental sibilants both plain /θ, ð/and pharyngealised /ðˤ/ are retained (Owens, 1984) despite earlier reports of their loss (Panetta, 1943). In the Western region (Tripolitania) however, the loss of these sibilants and their merger with the dental plosives /t, d, dˤ/ respectively is considered the only remaining feature of the pre-Hilalian urban variety (Pereira, 2007). Based on personal experience, these phones are used sporadically in formal settings in the sub-varieties of Tripolitan Arabic. Their use is restricted to Quranic recitation or a stylistic choice, which depends on register (Pereira, 2007), for example /θawɾa/ 'revolution', /jaðhab/ 'he goes', and /ðˤaɾf/ 'circumstance'.

The case of /v/ is somewhat different. /v/ can only be found in loan words from either Italian/foreign languages or technology- or media-related words such as the mobile phone application Viber, /vaɪbəɾ/ a popular communication tool for Libyans. In Italian loanwords, /v/ may be replaced by the either one of the native /b, f, w/. Alternatively, it can be maintained as /v/. Consider the examples from Abdu (1988) and their pronunciation in Libyan in table 3.3.

Table 3.3 Lexical borrowings from Italian into Libyan

phonological context	Italian	Libyan	gloss
word-initially	vapore	/babuːɾ/	'steamship'
	valvola	/valvala/	'valve'
	velo	/veːlːu/ - /beːlːu/	'wedding dress'
	valigia	/faliːʒa/	'suitcase'
	varichina	/waɾakiːna/	'bleach' (n.)
word-medially	avocatto	/ʔabukaːtːu/	'lawyer'
	lavandino	/lawandiːnu - /lavandiːnu/	'washing basin'
	modifica	/mudiːfka/ /mudiːvka/	'modification'

Velarised [ɫ] in Arabic is considered an allophonic variant of the plain /l/ (Embarki, 2013). It is not contrastive, thus cannot be part of the phonemic inventory. The case of the pharyngealised dental fricative /zˤ/ is not straightforward. It can be a sociolinguistic variant of /dˤ/ amongst a certain population of speakers within the city as mentioned earlier. Additionally, it alternates with either /dˤ/ for example /zˤarəf/ - /dˤarəf/, ‘envelope’ or /sˤ/ for example /lazˤga/ - /lasˤga/, ‘adhesive’. In some words, it is always realised as /zˤ/ for example /bːzˤːabːtˤ/ ‘exactly’. This is restricted to a minimal set of lexical items and is believed to be due to the direct contact with Berber (Swe’iy, 2013).

An additional phone is argued to be a stylistic variant. All of the aforementioned studies proposing the phonemic inventory of Libyan Arabic considered /q/ to be phonemic. However, it is argued that this sound is an allophonic variant of the phoneme /g/ since although in contrastive distribution, they do not form minimal pairs. For example, [gal] and [qal] ‘he said’ have the same meaning. This sound is fused with the voiced plosive /g/ in both urban and rural varieties across the country (Pereira, 2007). However, the sound is retained in direct borrowings from Standard Arabic, especially legal and religious terminology, for example, /bɪtˤaːqa/ ‘card’, /qurʕaːn/ ‘Qur’an’, /qɪsm/ ‘section’ or ‘department’, and in formal settings. These include public speech, recitation from the holy Quran, and language of instruction. /q/ can also be considered a sociolinguistic variant of /g/. For example, the proper name Tariq /tˤaːrəq/ is consistently produced as such by educated individuals only regardless of the formality of settings. Moreover, according to Pereira (2007: 84), alternating between /g/ and /q/ is a stylistic choice, which depends on the register.

Having excluded the controversial phones, we argue that Libyan Arabic spoken in Misrata has 24 consonant phonemes as shown earlier in table 3.2.

Plosives

The variety has eight main plosives; six plain /b, t, d, k, g, ʕ/ and two emphatic /tˤ, dˤ/. The phonemic distinction between the emphatic fricatives /dˤ/ and /ðˤ/ in MSA disappeared and the two sounds were merged into a single emphatic in Libyan Arabic, that is /dˤ/. The voiceless counterpart /p/ of the plosive /b/ does not exist in the variety. /p/ in Italian loan words into Libyan Arabic are consistently replaced by native /b/ irrespective of their phonological context. For example, words such as pacco, ‘box’ and ospedale ‘hospital’ are realised as /baːkːu/ and /sbetaːr/ respectively (Abdu, 1988). However, the voicing contrast between voiceless and voiced cognate plosives in other places of articulation such as coronal, and velar is utilised by the language. The plosive

/ʔ/ rarely occurs word-finally, though it is maintained word-initially and sometimes word-medially.

Despite, the similarity in labels used to denote plosives across languages, the phonetic implementation of voicing contrast varies by language. Languages utilise different VOT values to mark voicing contrast in word-initial plosives (Lisker and Abramson, 1964). Phonologically speaking, three categories of VOT are used, voicing lead, short lag, and long lag (aspirated) (Cho and Ladefoged, 1999).

Phonetic characteristics of Arabic plosives

A number of studies investigated VOT in various Arabic varieties with varying numbers of speakers. For example, Bukshaisha (1985) examined two male Qatari speakers, M. Ahmed (1984) investigated Sudanese, Al-Ani (1970) and Khattab, F. Al-Tamimi, and Heselwood (2006) examined Jordanian speakers, Al-Ani (1970), Giannini and Pettorino (1982), Heselwood (1996), and Odisho (1973) investigated Iraqi speakers, Heselwood (1996), Rifaat (2003), and Shaheen (1979) examined Egyptian speakers, and Flege and Port (1981) examined Saudi Arabians. Findings show that the implementation of voicing as far as VOT is concerned varies across Arabic varieties. Whilst some varieties utilise a negative VOT for voiced plosives and short-lag VOT for voiceless plosives (for example Lebanese), other varieties utilise a negative VOT for voiced plosives and aspirated VOT for voiceless plosives (for example, Saudi Arabian, Egyptian and Iraqi).

For Libyan Arabic adult speakers in various vowel contexts Kriba (2009: 212-3) describes VOT of /t/ as spoken in Zliten, a neighbouring city to Misrata both of which fall into the realm of Tripolitanian Libyan Arabic (henceforth TLA). He states that the mean VOT for /t/ is slightly aspirated in most vocalic contexts /iː/ = 51 ms (highest mean value exhibiting the strongest aspiration among vowel contexts), /ɪ/ = 35 ms, /eː/ = 35 ms, /oː/ = 32 ms, /u/ = 18 ms, /ɛ/ = 30 ms and /æː/ = 30 ms. He describes /t/ as aspirated based on Laver (1994) indication that audible aspiration ranges between 25 ms and 30 ms.

Garib (2014) also conducted a study, which involves measuring VOT values for syllable initial plosives in Libyan Arabic mono-syllabic words. He gathered data from 15 native speakers of Libyan Arabic aged 20-36. His results indicate that the phonologically voiced plosives in Libyan Arabic have negative VOT. /b/ has an average of -25 ms, /d/ -52 ms, and /g/ -38 ms. This means that Libyan Arabic falls under the same category as most other varieties of Arabic. However, Garib (2014: 26) claims that 'on average, these sounds are produced with voicing during the stop closure.' referring to Flege and Port (1981: 129) data that showed 'continuous glottal pulsing through the stop closure interval' for voiced plosives. However, he does not make a direct link to his own data nor a visual illustration of such finding.

On the other hand, phonologically voiceless plosives /t k/ have positive VOT. /t/ has an average of 50 ms and /k/ 51 ms. Garib (2014) results for VOT duration of /t/ are inconsistent with findings from Kriba (2009). Garib only included two vowel contexts for /t/, these are /ɪ æ/, for which the results are 50 ms collectively. The results for the matching vowel contexts in Kriba's study are 35 ms and 30 ms respectively. This could be due to variation in segmentation criteria or most likely having participants from two different regions in the country. Having measured the average duration for VOT across all vowel contexts also contributes to the lack of consistency since vowel context has been reported to affect VOT duration (Klatt, 1975). Garib does not provide a detailed criterion for his segmentation procedure. He does not mention the region of the participants' residence although it is clear from the test items that they belong to the eastern part of the country.

For Yemeni Arabic, Al-Nuzaili (1993) demonstrated that VOT of a Yemeni speaker ranges between -120:-15 ms for /b/, 15:55 ms for /t/, -130:-40 ms for /d/, 0:80 ms for /k/, and -130:-10 ms for /g/ depending on the vowel context. His results indicate that for Yemeni Arabic the voiced plosives fall into the lead voicing category, whereas /t/ and /k/ fall into long lag voicing category with the exception of /t/ in a low-front vowel context.

For Egyptian Arabic, Shaheen (1979: 87) states that plosives are characterised by a binary voicing distinction. Word-initially, /t, tˤ, k/ are completely voiceless, whereas /d, dˤ, g/ are completely voiced. Word-finally, /d, dˤ, g/ vary between completely voiced, partially voiced or voiceless. Inter-vocalically, voicing continues between the preceding vowel into the hold phase irrespective of the plosive's voicing class. Shaheen (1979) relied on temporal measurements of aspiration and argues that presence versus absence of aspiration is what separates voiceless from voiced stops respectively. VOT was argued to be difficult to segment. In voiced plosives, bursts often blend into the resonance of neighbouring vowels (resonance commences immediately after the hold phase) and in voiceless plosives, bursts often blend into the aspiration (aspiration commences immediately after the hold phase). Additionally, the multiplicity of bursts contributed to this difficulty suggesting that choice between bursts can be arbitrary and thus unreliable. The mean duration of aspiration varies by place of articulation. It was 30 ms for dental and post-dental stops and 45 ms for velar stops. Rifaat (2003) adds that voiced and voiceless plosives consistently have lead voicing and short lag respectively and neither of which ever overlap. His data show that stress and place of articulation are the only factors that influence VOT: VOT is longer for velars than dentals. Gender, emphasis, and length of the following vowel are not determining factors in VOT duration in his data.

For Iraqi Arabic, Al-Ani (1970) indicates that word-initially, VOT for the voiced plosives /b/ and /d/ is negative and ranges between -60:-110 ms and -80:-100 ms respectively, whilst for the voiceless plosives /t/, /k/ and /q/, VOT is positive and ranges between 40:60 ms, 60:80 ms, and 30:40 ms respectively depending on the following vowel length (longer VOT preceding longer vowels and vice versa). He argues that /t/ and /k/ are aspirated (as indicated by frictional noise following the release) syllable-initially and finally, whilst /q/ is unaspirated. Word-finally, /b/ and /d/ can be voiced or unvoiced, released or unreleased (though for /d/ mostly unreleased), whereas /t/, /k/ and /q/ can either be released or unreleased (though for /t/ mostly released). The place of articulation for /t, d/ he states, is dental.

In Jordanian Arabic, the case is not straightforward. VOT is always positive: short for voiced plosives but longer for voiceless counterparts. This also varies by vowel context. Mitleb (2001) reports that preceding the short vowel /i/, VOT durations for /t, d, k, g/ are 37 ms, 10 ms, 39 ms and 15 ms respectively. Preceding the long vowel /iː/, they are respectively 64 ms, 32 ms, 60 ms, and 20 ms. This indicates that whilst VOT serves to distinguish voiced from voiceless cognates, when followed by a short vowel, values of both voiced and voiceless plosives fall under the short lag category. When followed by a long vowel, voiced and voiceless stops fall under short lag and long lag categories respectively. Additionally, VOT in Jordanian does not seem to vary by place of articulation as it is the case in other languages/Arabic varieties.

Studies on Saudi Arabic VOT reveal that the variety employs a binary VOT distinction between phonologically voiced and voiceless stops, that is lead voicing and short lag respectively. VOT measurements reported in Flege and Port (1981) for word-initial /b, d, g, t, k/ are respectively -85 ms, -82 ms, -75 ms, 37 ms, and 52 ms. There were instances where /t, k/ exhibited voicing lead. However generally, they were considered slightly aspirated. The ranges for VOT in /t, k/ were respectively 20:65 and 30:85 ms, thus falling partly in the long-lag range. Alghamdi (1990) reported similar findings (for the Ghamdi variety of Saudi Arabic) of read materials embedded in a carrier phrase. VOT durations observed for /b, d, g, t, k/ are respectively -72 ms, -71 ms, -69 ms, 32 ms, and 42 ms arguing that /t, k/ are slightly aspirated. For Lebanese Arabic, Khattab (2002a) reports VOT patterns for Lebanese Arabic speakers that exhibit a binary VOT distinction, that is lead voicing and short-lag for voiced versus voiceless plosives respectively. VOT durations for /b, d, t, k/ are respectively -55 ms, -63 ms, 28 ms, 31 ms. /g/ was not included in the analysis as it is not utilised in Lebanese Arabic except in loan words.

For studies involving MSA, Yeni-Komshian, Caramazza, and Preston (1977) investigated VOT patterns in MSA read speech by Lebanese talkers. They found that results fall under a binary VOT pattern, that is lead voicing VOT for voiced plosives /b, d/ – with an average of respectively -65 ms and -56:6 ms – and short-lag VOT for the voiceless cognates /t, k/ – with an average of respectively 25 ms and 28 ms. These values varied by the vowel contexts /a, u, i/: in /b/ -80, -75, and -40 ms respectively: in /d/ -60, -70, and -40 ms respectively: in /t/ 20, 25, and 30 ms respectively: and finally, in /k/ 25, 30, and 30 ms respectively. A study by Jesry (1996) on the production of MSA read materials by Syrian speakers revealed similar averages of VOT for /b, d, t, k/ in words embedded in carrier phrases. These are -68:7 ms, -66:8 ms, 27:8 ms, and 32 ms respectively.

Fricatives

The variety has 11 main fricatives /f, s, z, sˤ, ʃ, ʒ, x, ɣ, ħ, ʕ, h/. For example, a word such as /ðˤalaːm/ in Classical Arabic meaning ‘darkness’ is only pronounced as such in formal speech/recitation of holy Quran. Otherwise, it would be /dˤəlaːm/ in informal speech or /zˤəlaːm/ by a small subset of the city's population making it a sociophonemic variant not a main phoneme. This subset of Misrati residents are believed to be affected by Tamazight contact in those areas at one point in Libyan history (Habara, 2017).

Liquids, approximants and nasals

The variety has two variants of the rhotic consonant, the tap [ɾ] and the trill [rː]. The latter occurs in geminate contexts and arguably in word-final positions and the former in all other contexts (Abumdas, 1985; Muftah, 2001). Issa (2016) argues for an additional rhotic variant, that is the approximant for both geminate and non-geminate contexts. However, the variety she describes is Tripolitan and although Misrati Libyan Arabic linguistically falls under that region, no such phonetic observations have been confirmed in the literature. Misrati Libyan Arabic also has three approximants, the lateral /l/, /w/[11], and /j/ as well as the two nasals /n/ and /m/.

Emphatics and gutturals

As can be seen above, the variety also exhibits a range of emphatics /tˤ, dˤ, sˤ, zˤ/ and gutturals /x, ɣ, ħ, ʕ, h/. Emphatics are rare across languages (Ladefoged and Maddieson, 1996; Mitchell, 1993). They have a secondary articulation and are realised in pharyngeal and uvular locations in the oral tract (McCarthy, 1989). The tongue root in this secondary articulation is retracted towards the back wall of the pharynx, hence *pharyngealised* (Kenstowicz, 1994). Whilst some studies associate emphatics' realisation with retraction of the epiglottis, raising of the larynx, tense voice quality and protrusion of lips (e.g. F. Al-Tamimi and Heselwood, 2011), J. Al-Tamimi (2017) demonstrates that pharyngealisation in Arabic involves a retraction of tongue root accompanied by a simultaneous back and down gesture resulting in a constricted glottis. Rhotics are also considered emphatics in certain phonetic contexts that is when adjacent to the /a/ vowel (Shahin, 1996).[12]

---

[11] Note that /w/ is not included in 3.2 because it is a labio-velar and thus could not be fitted in the chart.

[12] Geminates were not considered in the description of Libyan Arabic spoken in Misrata.

3.2.2 *Vowels*

Monophthongs

Studies on Libyan Arabic vowel inventory vary in their findings with regards to number of vowels and the issue of allophonic variation. The earliest studies found in this regard is that by Griffini (1913: xxiv), who states that the total number is 15 vowels. Another by Panetta (1940: 9) states there are nine vowels but in a later study (1943: 2, 16), he argues for eight only. Abumdas (1985: 41) acknowledges ten vowels. He claims that the emphatic variants of Libyan Arabic are phonemes in their own rights and not merely allophones. Botagga (1991) provides a description of vowels spoken in Sebha, a city in the southern region. He adds the vowel /ʌ/, which can only be heard in words containing emphatics and is, thus, considered by A. Ahmed (2008: 85-6) to be an allophone of /a/. Botagga (1991: 70) nevertheless, also claims that /æ/ and /ʌ/ are allophones of /ɑ/. Aurayieth (1982: 23), on the other hand, deals with vowels spoken in the eastern region. Elramli (2012) provides a description of vowels spoken in Misrata. They all agree that Libyan Arabic has three short vowels /a/, /i/ and /u/ and three long counterparts /aː/, /iː/, /uː/ with the addition of the two long vowels /oː/ and /eː/ that are the transformed version of the Classical Arabic /aw/ and /aj/ in word-internal position (Owens, 1984: 10). However, an instrumental study carried out by A. Ahmed (2008) demonstrates that the short vowels are not only different in quantity, that is less than half as short as the long vowels, but also different in quality. He recognises the following vowels to constitute the Libyan Arabic vowel system, /iː/, /ɪ/, /uː/, /ʊ/, /eː/, /oː/, /æ/, and /ə/ pointing out that contrary to the commonly held view, the short vowels are rather centralised (A. Ahmed, 2008: 202) not only in Libyan Arabic but also other varieties of Arabic (Ghazali, 1977; J. Al-Tamimi, Carré, and Marsico, 2004; J. Al-Tamimi, Carré, et al., 2004). In addition, the long vowels are also argued to be realised with tenser tongue compared to the short vowels (Muftah, 2001: 82). The following is an illustration of Libyan Arabic vowel plots taken from A. Ahmed (2008).

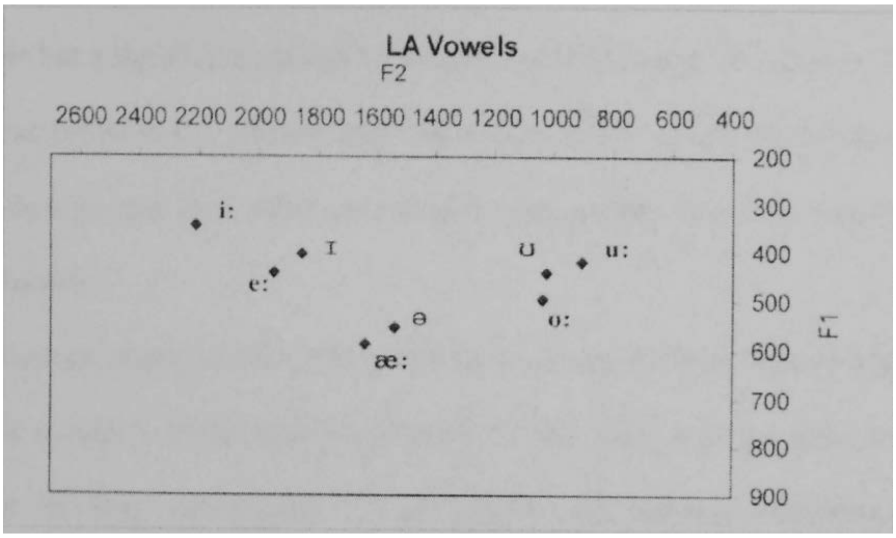


Figure 3.2 Libyan Arabic vowel inventory adopted from A. Ahmed (2008)

Diphthongs

In Misrati Libyan Arabic , there are two diphthongs, /aw/ and /ej/, which can only occur word-finally (Elramli, 2012: 20). An example of each is, /saw/, ‘harm’ and /ʃej/, ‘a thing’ respectively. Word-medially, these two diphthongs have gone reduction from Standard Arabic into the vowels /oː/ and /eː/ respectively (Elramli, 2012: 20). Examples are:

	Standard Arabic	Libyan Arabic	gloss
3.1	/jawm/	/joːm/	‘day’
	/bejt/	/beːt/	‘house’

In formal speech and recitation of the Holy Quran, these diphthongs are maintained.

Based on this distribution of [au] and [ai] word-finally and [oː] and [eː] elsewhere in the word, and given that all instances of the latter are respectively derived from the former in Classical Arabic (Owens, 1984: 10) not only in Libyan but in most Arabic varieties (Abu-Mansour, 1992: 49), it can be assumed that each pair, that is [aw] and [oː], on one hand, and [aj] and [eː], on the other, are allophones rather than independent phonemes. Moreover, this feature, Pereira (2007: 85) states, is found in nomadic varieties across North Africa.

3.2.3 *Structure*

One aspect that the study seeks to investigate is the acquisition of English clusters of two consonants. Libyan Arabic syllable structure is more restricted than that of English. While English allows up to three consonants in an onset and up to four in a coda (Roach, 2009), Arabic does not. Not only is it restricted, it also varies from one variety to another. For this reason, Libyan Arabic learners of English as a foreign language are expected to face difficulties in this area.

Clusters, in the literature on Arabic, are the result of morphemic concatenation, syncope or across word boundaries (Abu-Mansour, 1992; Aquil, 2013; Elramli, 2012; Hamdi, Ghazali, and Barkat-Defradas, 2005; J. Watson, 2002). The latter is considered more of a consonant sequence than a cluster given its members belong to different syllables (Pulgram, 1965: 76). In MSA, however, a syllable cannot start with a vowel or have a complex onset. The literature on Libyan Arabic syllable structure and phonotactics is less consistent. According to Laradi (1983: 25), in TLA, an onset may have up to three consonants, whilst the coda may have up to two consonants, and/or be a geminate. She also proposes that onsets and codas are optional and speculates that permitting complex onsets may be due to contact with Berber. Syllable types based on Laradi include those in table 3.4.

Table 3.4 Syllable types in Laradi (1983)

syllable	example	gloss
CV	/zaʕ.ma/	‘maybe’
CVC	/ʃaɾ.ba/	‘soup’
CVCC	/bint/	‘a girl’
	/sadd/	‘dam’
CCV	/tla.tiːn/	‘thirty’
CCVC	/ʃbaħ.nah/	‘we saw him’
CCVCC	/ʕɾaft/	‘I knew’
	/ʕɾabb/	‘Arabs’
CCCVC	/nftaħ/	‘it opened’
VːC	/aːʃ/	‘how’
CVː	/laː/	‘no’
CVːCC	/ʃaːdd/	‘exception’
CCVː	/ɣtˁaː/	‘cover’ (n.)
CCVːC	/ktaːb/	‘a book’
CCCVːC	/stɾaːħ/	‘he rested’

Elgadi (1986: 56-7 in Shitaw, 2014) shares the same views but adds two additional syllable types as shown in 3.2.

3.2 CV:C /baːb/ ‘door’
CCCV /nkwe/ ‘to be cauterised’

Although Abumdas (1985) does not list the syllable type CVVC explicitly, it appears as the second syllable in one of the examples /tla.tiːn/ ‘thirty’.

Additionally, Abumdas (1985: 89) adds three more types in the examples shown in 3.3.

3.3 V /a.be/ ‘he agreed’
Vː /uː.guf/ ‘stand up’ (v.)
VC /aswad/ ‘black’

Shitaw (2014: 27) argues that these can be preceded by a glottal stop. Nonetheless, no study has formally (phonetically) investigated this dispute either in words in isolation nor continuous speech. It must be mentioned here that the variety of Libyan Arabic Abumdas discusses is that spoken in Zliten, whereas all the others are Tripolitan. It is assumed that they – including Misrata Libyan Arabic – fall under the umbrella of Tripolitan Libyan Arabic. However, generally speaking, there are inevitable phonological, if not also phonetic, differences.

According to Elramli (2012) for Misrata Libyan Arabic and Shitaw (2014) for Tripolitanian Libyan Arabic, Libyan Arabic does not allow more than two consonants in each position, unlike Laradi’s (1983) account above which shows that onsets allow up to three consonants. The following overview of syllable types in Misrata Libyan Arabic is adapted from Elramli (2012: 23). He outlines ten Misrata Libyan Arabic syllable types in table 3.5.

Table 3.5 Syllable types in Elramli (2012). The boldfaced syllable types are also shared with Laradi (1983)

syllable	example	gloss
CV	/bɨ.ɾad/	‘got cold’
CVV	/laa/	‘no’
CVC	/guɾ.ma/	‘gossip’
CVCC	/bint/	‘a girl’
CCVV	/tnaa.du/	‘you call’
CVVC	/guul/	‘say’
CCV	/ħal.lqa/	‘shaved it’
CCVVC	/blaad/	‘country’
CCVC	/gɾib/	‘water bags’
CCVCC	/kɾumb/	‘cauliflower’

In Misrata Libyan Arabic, onsets can comprise of one to two consonants and may not violate the Sonority Sequencing Principle (SSP). This principle commands that in a sequence of consonants in a syllable, the most sonorous one is the closest to the nucleus. Onset clusters that violate SSP are typically the result of a historical process of syncope (Elramli, 2012: 22), whereby a high short vowel is elided in weak syllables (McCarthy, 2007). For example /ktaab/ ‘book’ and /ɾgaad/ ‘sleeping’ are historically derived from /kɪtaab/ and /ɾuqaad/ respectively.

Epenthesis is a common phonological process in complex onsets and codas. It is one of the repair strategies for impermissible clusters or a sequence of consonants. In Arabic its conditioning is heterogeneous across and within varieties. For instance, in San’ani and Cairene, it is argued to be triggered by either the sequence of three consonants in concatenation or violation of the Sonority Sequencing Principle (J. Watson, 2002). An additional assumption for Cairene is that it is the result of a violation to syllable well-formedness as well-formed syllables are maximally two morae (Aquil, 2013).

For coda clusters in LA, Abumdas (1985: 86) argues that the first member can only be a continuant, such as, /l/, /r/ or /n/, whereas Elramli (2012: 23) claims that SSP does not apply. He supports his argument using examples form MSA such as ‘e.g. /ʔism/ ‘name/noun’, /maɦɾ/ ‘dowry’ /baɾq/ ‘lightning’, /ħaɾb/ ‘war”. While this is true for MSA in pre-pausal position, it is not how these words are realised in the local variety, at least not in pre-pausal position.

Like Elramli (2012), Abu-Mansour (1992) argues that epenthesis in CVCC syllables in Makkan Arabic is the result of segmental conditioning of the cluster. This is to say that CVCC is already assigned to well-formed syllables in Makkan Arabic and that epenthesis only takes place when the cluster violates SSP. While this holds true for Makkan Arabic, the case is not as straightforward as Elramli suggests for Libyan Arabic. In this variety, CVCC is also considered a well-formed syllable /bint/ ‘girl’. Nonetheless, epenthesis occurs even when the segments do not violate SSP. Consider the examples in table 3.6.

Table 3.6 Epenthesis not violating SSP in Misrata Libyan Arabic

word	gloss	derivations	gloss
/basˤɨq/	‘spitting’	/basˤqa/	‘a spit’
/manəħ/	‘granting’	/mɪnħa/ - /munħa/	‘grant’
/tˤaɾəf/	‘piece’	/tˤaɾfeːn/	‘two pieces’
/sabɨt/	‘Saturday’	/sabteːn/	‘two Saturdays’
/ʃaməs/	‘sun’	/ʃamsi/	‘solar’

However, there are cases where epenthesis does not take place. This occurs in clusters where the first member is either a lateral or a homorganic nasal. Consider the examples in 3.7.

Table 3.7 Examples of (non)permissible coda clusters

underlying form	surface form	gloss
/bint/	[bint]	‘girl/daughter’
/band/	[band]	‘item’
/kanz/	[kanɨz]	‘treasure’
/hind/	[hind]	‘Hind (a girl’s name)’
/ʕunf/	[ʕuɱf]	‘violence’
/bank/	[baŋk]	‘bank’
/danb/	[damb]	‘fault’
/krunb/	[krumb]	‘cauliflower’

In Tripolitan Libyan Arabic plosive clusters, be it the result of syncope or morphemic concatenation, Shitaw (2014: 41) found that in onsets, the first member can be realised with or without an audible release, or a short central ‘inter-consonantal interval’ heard as a schwa. He identifies this vocalic interval as an excrescent vowel given it is voiceless between voiceless plosives in a cluster. In pre-pausal coda clusters comprising two plosives, he indicates that the vocalic interval is heard as a full schwa-like sound that is consistently voiced, thus regarded epenthetic.

3.3 Characteristics of British English

3.3.1 *Consonants*

Table 3.8 British English SSBE consonant inventory

	B-Lab	Lab-Den	Den	Alv	Pal-Alv	Pal	Vel	Glo
Plosive	p b			t d			k g	ʔ*
Nasal	m			n			ŋ	
Fricative		f v	θ ð	s z	ʃ ʒ			h
Affricate					ʧʤ			
Approximant					ɹ	j		
Lateral				l				

B-Lab=bilabial, Lab-Den=labio-dental, Den=dental, Alv=alveolar, Pal-Alv=palato-alveolar, Ret-flex=retroflex, Pal=palatal, Vel=velar, Glo-glottal.

Table 3.8 shows that English has six plosives, two bilabial /p b/, two alveolar /t d/, and two velar /k g/. The glottal stop [ʔ] is a variant of /t/ syllable-finally when followed by a non-syllabic consonant (Kortmann and Upton, 2008: 249). /p t k/ are consistently voiceless, unlike /b d g/, which are either fully voiced, partly voiced or unvoiced depending on their position in a syllable (Roach, 2009). The present study is primarily concerned with syllable/word-initial plosives so description is limited to this position. According to Roach (2009: 26) and Cruttenden (2008: 158), plosives are formed through a sequence of articulatory events, namely closing phase, compression phase, release phase, and post-release phase. For syllable initial plosives, the closing phase is silent for all plosives. The compression phase is characterised as silent for phonologically voiceless plosives but not for the voiced counterparts /b d g/. For the latter, there can be little voicing shortly preceding the release (Roach, 2009). Additionally, there could be more voicing throughout the compression phase in slow, careful speech. In rapid speech, there is no voicing in the compression phase. The release phase for /p t k/ is also characterised with burst noise followed by aspiration in the post-release phase, after which comes voicing for the following vowel. The release phase for /b d g/ is characterised with a rather weak plosion, where voicing starts either simultaneously or shortly after. It is the aspiration that primarily distinguishes phonologically voiceless from voiced plosives in English rather than voicing per se. For this reason it has been suggested that a more accurate description is to label /p t k/ as fortis and /b d g/ as lenis instead (Roach, 2009) though this aspect is controversial. This variety of English also has nine fricatives, two labiodental /f v/, two dental /θ ð/, two alveolar /s z/, two post-alveolar /ʃ ʒ/ and one glottal /h/, only two affricates /ʧ ʤ/, three nasals /m n ŋ/, the lateral approximant /l/, the post-alveolar approximant /r/, and the glides /w/[13] and /j/.

Dental fricatives

The dental fricatives in English are produced when air flows through a narrow constriction formed between the tongue tip and sides as well as the inner surface and upper side of the teeth (Gimson, 1970; Jones, 1956; Ladefoged and Maddieson, 1996; Roach, 1990). The shape of the tongue flat rather than groove compared to that in /s, z/ is what yields a rather lower frequency frication noise (Cruttenden, 2014). It is worth noting that in most cases, they are realised with stopping, that is there are usually realised with a burst followed by a fricative.

They pose pronunciation difficulties when followed by /s, z/ and in such contexts (clusters) tend to be elided by native speakers.

The lateral approximant

The lateral approximant /l/ is mostly voiced and alveolar in place. According to Cruttenden (2014), it has a range of allophones depending on the phonological context. He describes that it is:

1. clear [l] before either vowels or /j/ word-initially (whether singleton or in a cluster), word-medially, or word-finally if it is followed by a vowel or /j/.
2. clear and fully devoiced [l̥] when preceded by voiceless plosives in accented syllables.
3. clear and partially devoiced [l̥] in:
    - unaccented syllables/across syllable boundaries;

---

[13] Note that /w/ is not included in 3.8 because it is a labio-velar and thus could not be fitted in the chart.

    - and when preceded by voiceless fricatives.

4. dark [ɫ] everywhere else:

    - word-finally when preceded by a vowel;
    - between a vowel and consonant;
    - syllabic [ɫ̩], which can be partially devoiced when preceded by a voiceless consonant.

Interestingly, in contexts where /l/ is final and preceded by a bilabial consonant, speakers of General British in general and Estuary English in particular tend to use [ʊ] solely for [ɫ] (Cruttenden, 2014). However, this vocalisation process is avoided in contexts of other preceding consonants and claimed to be characteristic of child speech after alveolar plosives in words such as *little, console*, and *paddle* according to Cruttenden (2014). Other accounts (Scobbie and Wrench, 2003; Trudgill, 1984; Tollfree, 1999; Hardcastle and Barry, 1985) assert that the tongue-alveolar ridge contact for /l/ in final position is weakened or lost in adult speakers as well. The velarisation of [ɫ] usually affects the preceding front vowel by retracting and lowering slightly its articulation (Cruttenden, 2014). When preceded by [iː], a central glide can be heard between the vowel and the lateral. In the diphthongs /eɪ, aɪ, ɔɪ/, [ɪ] becomes either shorter or elided. When /uː/precedes [ɫ], it tends to be monophthonged and closer to [u] than /uː/ in words like *tool, pool*, and *cool* (Cruttenden, 2014: 219).

The alveolar approximant is formed, as its label indicates, when the tip of the tongue is in close approximation with the alveolar area but no contact is made, contrary to Arabic /ɾ/ where a contact between the tongue and dental-alveolar region is necessary. In environments where /r/ follows /p t k/, it is rather voiceless and fricative in nature (Roach, 2009: 5). One additional feature that accompanies this articulation in English is a slight lip-rounding. The position of this sound is limited to contexts preceding vowels because SSBE English is a non-rhotic accent. It is never realised after vowels unless it is intervocalic.

3.3.2 *Vowels*

SSBE has six relatively lax vowels /ɪ, e, æ, ʌ, ɒ, ʊ/ and five rather long vowels /iː, aː, ɜː, ɑː, ɔː, uː/ and the diphthongs /ɪə, eə, ʊə, eɪ, aɪ, ɔɪ, əʊ, aʊ/. The vowels /i/ (e.g. /ˈhæ.pi/ 'happy') and /ə/ (e.g. /ˈkʌ.lə/) are limited to unstressed syllables and are thus considered allophones.

3.3.3 *Structure*

The essential part of a syllable is the nucleus, which is mostly a vowel. It may comprise of a nucleus only (V), in which case it is called a minimum syllable. It can be preceded by one consonant (CV), which is the onset, or followed by one consonant (VC), which is the coda or preceded and followed by one consonant on each side, for example:

Table 3.9 Epenthesis not violating SSP in Misrata Libyan Arabic

V		CV		VC		CVC	
are	/ɑː/	far	/fɑː/	odd	/ɔd/	can	/kæn/
err	/ɜː/	sore	/sɔː/	as	/æz/	hill	/hɪl/
or	/ɔː/	her	/hɜː/	if	/ɪf/	could	/kʊd/
oh	/əʊ/	bye	/baɪ/	egg	/ɛg/	girl	/gɜːl/

In English in singleton onsets, the onset can be any consonant although /ŋ/ and /ʒ/ are rare (Roach, 2009: 57).

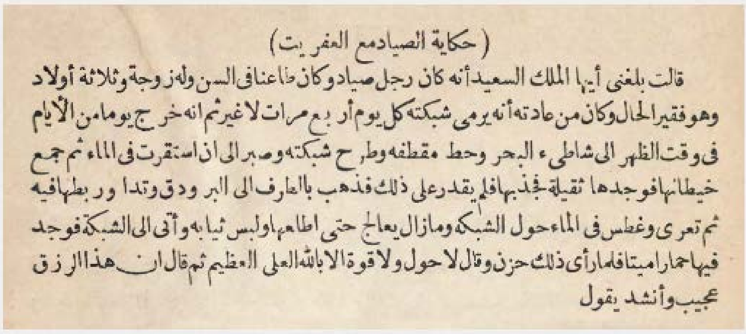
The English syllable is more complex than that in Arabic allowing up to three consonants in onsets and up to four in codas (CCC)V(CCCC). There are however phonotactic restrictions as to the type of consonant allowed in each slot for such complex clusters (see for example, Roach, 2009).

3.4 Orthographic systems of Arabic and English

One of the differences between Arabic and English is their written systems. Whilst Arabic has a consonantal writing system representing consonants, that of English represents phonemes as letters and words as a space-separated string of letters

(Bassetti, 2008). In such a system, Arabic shows a relatively highly regular grapheme-to-phoneme correspondence, whereas English has a relatively less regular grapheme-to-phoneme correspondence. Arabic is right-to-left writing system. By comparison, English is left-to-right. Arabic tends to be a vowel-less writing system, except in cases where there is a long vowel, or in early grades in school when children start to learn to read and write. Script-wise, Arabic utilises Arabic script, and English utilises Roman letters (c.f. figure 3.3).

Figure 3.3 Excerpt illustrating Arabic cursive writing system



(حكاية الصياد مع العفريت)
قالت بلغنى أيها الملك السعيد أنه كان رجل صياد وكان طاعنا فى السن وله زوجة وثلاثة أولاد
وهو فقير الحال وكان من عادته أنه يرمى شبكته كل يوم أربع مرات لا غير ثم انه خرج يوما من الأيام
فى وقت الظهر الى شاطىء البحر وحط مقطفه وطرح شبكته وصبر الى ان استقرت فى الماء ثم جمع
خيطانها فوجدها ثقيلة فجذبها فلم يقدر على ذلك فذهب بالطرف الى البر ودق وتدا وربطها فيه
ثم تعرى وغطس فى الماء حول الشبكة ومازال يعالج حتى اطلعها ولبس ثيابه وأتى الى الشبكة فوجد
فيها حمارا ميتا فلما رأى ذلك حزن وقال لا حول ولا قوة الا بالله العلى العظيم ثم قال ان هذا الرزق
عجيب وأنشد يقول

— An Arabic edition of *One Thousand and One Nights*, showing the beginning of the story "The Fisherman and the Jinni"

Arabic letters are relatively more fluid compared to English. Their shape depends on their position in a word. Examples illustrating context-dependent letter shape and connectors (top figure in figure 3.4) vs non-connectors (bottom figure in figure 3.4).

Figure 3.4 Different letter shapes in Arabic

Furthermore, Arabic, unlike English, does not utilise case distinction. Table 3.10 below presents a summary comparison of Arabic and English orthographic systems.

Table 3.10 Summary comparison of Arabic and English orthographic systems

Arabic	English
higher grapheme-to-phoneme correspondence	lower grapheme-to-phoneme correspondence
right-to-left system	left-to-right system
diacritics for short vowels	letters for all vowels
Arabic script	Roman alphabet
cursive style	optionally cursive
no case distinction	upper/lower-case letters
context-dependent shape of letters	context-independent shape of letters

3.5 Summary

This chapter presented the phonological background of the participants' first language and that of the target language. Two points are worth reiteration. First, Arabic children in Misrata – from where the participants of the present study are sampled – had minimal exposure to MSA. Second, the testing materials, as we will describe in Chapter 5, were recorded in British English SSBE. Table 3.11 below provides a summary of the main differences between the phonology of each language.

Table 3.11 Summary comparison of Arabic and English phonology

Arabic	English
- has emphatics and gutterals - does not have; /p, ŋ, v, θ, ð, ʧ, ʤ/	- does not have emphatics or gutturals
/r/-sound - flap everywhere (except) - trill when geminated	/r/-sound - approximant word-initially/-medially - not realised post-vocalically
/b, d, g/: lead voicing	/b, d, g/: short lag
/t, k/: short lag	/p, t, k/ long lag and aspirated in singleton onsets
/t/ dental	/t/ alveolar
/l/: - clear word-initially - clear word-finally - only velarised in ‘Allah’	/l/ : - clear word-initially - velarised word-finally
- diphthongs restricted: - [aw] and [aj] word-finally	- wider range of dipthongs - /ɪə, eə, ʊə, eɪ, aɪ, ɔɪ, əʊ, aʊ/
- 8 vowels: - 3 high-front; - 1 low-front; - 1 central; - 3 high-back	- 6 lax vowels: - /ɪ, e, æ, ʌ, ɒ, ʊ/; - 5 long vowels: /iː, aː, ɜː, ɑː, ɔː, uː/
- less complex syllable structure: - CCVCC	- more complex syllable structure: - CCCVCCCC
- Onset compulsory	- Onset optional

Chapter 4: Children's development of phonology in English and Arabic

4.1 Introduction

Factors that are proposed to influence language learning/acquisition include implicational markedness (Eckman, 1991; Eckman, 2008), input frequency effects (Beckman and Edwards, 2000; Bernhardt and Stemberger, 2017; Edwards, Beckman, and Munson, 2004; Levelt, Schiller, and Levelt, 2000; Munson, Edwards, Beckman, Cohn, Fougeron, and Huffman, 2011; Roark and Demuth, 2000; Storkel, 2004; Zamuner, 2003; Zamuner, Gerken, and Hammond, 2004), and functional load (Stokes and Surendran, 2005). They also include so-called developmental factors that were based on perceptual, cognitive and biological (poor oro-motor skills) limitations (Dodd, Holm, Hua, and Crosbie, 2003; Lindblom, 1992; MacNeilage, 1980). In addition, demographic factors such as gender (usually in favour of females) (Dodd, Holm, Hua, et al., 2003), language experience (Kuhl, K. Williams, et al., 1992), position within siblings, and expectations from caregivers (Dodd, Holm, Crosbie, and Zhu, 2013) can also influence language acquisition. The question can be asked of whether the error patterns observed are typical of the target language development or reflective of L1 transfer. The present study examined the development of L2 English phonology by Libyan Arabic seven-year-olds, and it is even more important to consider the developmental processes and error patterns a monolingual child demonstrates in acquiring English as a native language.

In addition, looking at the L2 learners' L1 phonological development can be useful. J. Paradis (2008) stresses the importance of comparing child L2 learners to age-matched native speakers of the target language. In addition, it is also necessary to consider the stages in phonological acquisition of the learners' first language, Arabic. This is to distinguish any possible processes that are typical of their first language acquisition from those of the target child language acquisition.

In the following sections, the literature on English child language acquisition followed by Arabic child language acquisition will be explored in terms of age of acquisition of consonants and consonant clusters. In addition, children's developmental errors and processes will be discussed and compared for each language.

4.2 Children's acquisition of English phonology

For English acquisition, once children acquire fifty words, usually around the age of two and as Ingram (1999) calls it *the word spurt*, they start to display consistent error patterns (N. Smith, 1973). These patterns are generally substitutions and can be categorised into (velar) fronting, backing, stopping, gliding of liquids, affrication, deaffrication, vocalisation, and voicing (Dodd, Holm, Crosbie, et al., 2013). Despite the consistency of patterns observed, there are individual differences across children.

Several studies have been carried out to explore the age of acquisition of the consonantal inventory in children with various age groups for American English (Dyson, 1988; Smit, Hand, Freilinger, Bernthal, and Bird, 1990; Stoel-Gammon, 1987), for Australian English (McIntosh and Dodd, 2008), for British English (Dodd, Holm, Hua, et al., 2003; Dodd, Holm, Crosbie, et al., 2013).

Mcleod and Crowe (2018) provided a review of 15 studies (within a larger review of 27 languages) on the acquisition of consonants in different varieties of English. Six of these studies are General American (Arlt and Goodban, 1976; B. Pearson, Velleman, Bryant, and Charko, 2009; Poole, 1934; Prather, Hedrick, and Kern, 1975; Temp1in, 1957; Wellman, Case, Mengert, and Bradbury, 1931), three Australian English (Chirlian and Sharpley, 1982; Kilminster and Laird, 1978; McIntosh and Dodd, 2008), one African American (B. Pearson et al., 2009), one Midwestern American (Smit et al., 1990), one British (Dodd, Holm, Hua, et al., 2003), one Cape Town (Mowrer and Burger, 1991), one Irish (Monaghan, 2014), and one Malaysian (Phoon, 2010). Unlike the former cited studies, where participants were monolinguals, participants from the Malaysian study were multilingual. Children's ages in these studies collectively ranged between 1;11 to

12;11. The number of participants collectively was 7369 ranging between 60 to 1756 participants. Mcleod and Crowe (2018) categorised consonants into early, middle, and late. Early consonants are those mastered (90-100% correct) between the ages of two and four years. Their results showed that early consonants were /p, b, m, d, n, h, t, k, g, w, ŋ, f, j/ in this order. Middle consonants are identified as those mastered between the ages of four and five. These were /l, ʤ, ʧ, s, v, ʃ, z/ also in this order. Late consonants were those mastered between the ages of five and seven. These were identified as /ɹ, ʒ, ð, θ/ in this order.

The British English study by Dodd, Holm, Hua, et al. (2003) investigated the acquisition of consonant inventories and developmental errors of 684 British English children aged between 3;0 and 6;11 and Dodd, Holm, Crosbie, et al. (2013) added 32 children aged 2;0-2;11 (from Newcastle upon Tyne) to supplement data on developmental patterns. They too considered a 90%-threshold for acquisition. Their sampling population included children from London (14.9%), North East (16.7%), North West (7.3%), South East (16.2%), South West (14.6%), Midlands (12.5%), Wales (7.3%), and Scotland (10.4%). Their results show that the plosives /p, b, t, d, k, g/, the nasals /m, n, ŋ/, the fricatives /f, v, s, z, h/, and the approximants /w, j/ and word/syllable-initial /l-/ were mastered by the age of 3;0 to 3;5. The voiceless affricate /ʧ/ was mastered by the age of 3;6-3;11. Together, these are considered early consonants in British English. The fricative /ʒ/ and the voiced affricate /ʤ/ were mastered by the age of 4;0 to 4;5. The voiced alveo-palatal fricative and voiced affricate are considered middle consonants.

The fricative /ʃ/ was mastered by the age of 5;0 to 5;5. The rhotic approximant /ɹ/ was mastered by the age of 6;0 to 6;5. However, according to Shriberg (1993) native English children do not fully master the English rhotic approximant production until the age of eight. Furthermore, rhotic approximants seem to be one of the most problematic sounds for native English children (Shriberg and Kwiatkowski, 1994; Smit et al., 1990). Finally, both the interdental fricatives /θ, ð/ were mastered after the age of seven. The age of acquisition of the three final categories, the alveo-palatal fricative /ʃ/, the rhotic approximant and the interdental fricatives, classifies them as late consonants according to Mcleod and Crowe (2018), who consider consonants acquired after the age of five to be late. Discrepancies between the various dialects of English on the one hand and British

English on the other, in the early consonants were demonstrated in the acquisition of /l/. This could be due to the fact that /l/ varies by prosodic context and allophony across those dialects. Dodd, Holm, Crosbie, et al.'s study does not provide information on the acquisition of final /-l/ because although they include it in their stimulus, its occurrence ratio compared with word-initial /l-/ was 1:4 in singletons (p.641) and 2:5 in clusters. /l/ in most other varieties compared in the review by Mcleod and Crowe (2018) velarised and word-finally dark /-l/ tends to undergo a process of vowelisation (Grunwell, 1985; Hodson and Paden, 1981; Stoel-Gammon and Dunn, 1985). In addition, the voiced fricatives /v, z/ and /ʒ/ are acquired relatively earlier in British English in comparison to data collated from the collection of dialects, a part of which is British English. The only consonant acquired relatively later in British English was /ʃ/.

According to Bankson and Bernthal (1998), error types that affect individual sounds are substitution processes. These are identified as (velar) fronting, backing, stopping, gliding of liquids, affrication, deaffrication, vocalisation, and voicing (Bankson and Bernthal, 1998; Dodd, 1995; Stoel-Gammon and Dunn, 1985). Phonological rules determine the context in which these processes occur (Holm, 1998). Developmental patterns reported in English child acquisition include reduplication, harmony, final consonant deletion, cluster reduction, fronting (including velar fronting and depalatalisation), fricative stopping, gliding (mostly liquids), and voicing (Grunwell, 1985; Stoel-Gammon and Dunn, 1985). Whilst gliding of liquids persists until the age of 5;11 (Dodd, Holm, Hua, et al., 2003; Dodd, Holm, Crosbie, et al., 2013), reduplication, harmony, voicing, and final consonant deletion disappear/are limited after the age of three (Grunwell, 1985; Stoel-Gammon and Dunn, 1985).

Stopping declines considerably after the age of 3;6 (Dodd, Holm, Hua, et al., 2003; Dodd, Holm, Crosbie, et al., 2013; Grunwell, 1985; Stoel-Gammon and Dunn, 1985). The next process to decrease is weak syllable deletion and fronting (around the age of four) (Dodd, Holm, Hua, et al., 2003; Dodd, Holm, Crosbie, et al., 2013). Although, velar plosive fronting was absent after the age of four (Dodd, Holm, Hua, et al., 2003), over 10% of the sample continued to front /ŋ/ to [n] in the word 'fishing' until the age of five. In addition, Th-fronting /θ/ to [t], was reported to appear after three (Grunwell, 1985; Stoel-Gammon and Dunn, 1985). Deaffrication is also one of the long-lasting processes in British English and starts to disappear around the age of five (Dodd, Holm, Hua, et al., 2003; Dodd, Holm, Crosbie, et al., 2013). Grunwell (1985) and Stoel-Gammon and Dunn (1985) report an additional developmental error not reported by Dodd, Holm, Hua, et al. (2003) or Dodd, Holm, Crosbie, et al. (2013) and that is liquid vocalisation in codas.

4.2.1 *Consonant clusters*

In terms of clusters and related developmental patterns (Greenlee, 1974; Kirk, 2008; McLeod, van Doorn, and Reed, 2001b; McLeod, van Doorn, and Reed, 2001a), it is argued that they are difficult to learn (Kirk, 2008) and that they are one of the long lasting aspects in acquisition (McLeod et al., 2001b).

The production of clusters in English-speaking children emerges around the age of two, when children can produce some cluster types correctly (Lleó and Prinz, 1996; McLeod et al., 2001a; M. Watson and Scukanec, 1997). They are typically mastered after the age of three albeit for some children it can extend to around eight to nine years of age.

Children first produce consonant clusters that do not match adult targets (M. Watson and Scukanec, 1997). It has also been found that there are asymmetries in acquisition between word-initial and word-final clusters (Kirk and Demuth, 2003; Kirk and Demuth, 2005; Macken and Barton, 1977). For example, the range of cluster types word-finally is greater than that in word-initially by the age of 2;9 (McLeod et al., 2001b). Moreover, bi-consonantal clusters are produced and mastered before tri-consonantal counterparts (Smit et al., 1990). This has been attributed to added difficulty resulting from increased phonotactic complexity as well as inherent difficulties lying in elements of tri-consonantal clusters (McLeod et al., 2001a). It has also been observed that clusters containing plosives are acquired before those containing fricatives (e.g. Templin, 1957). By the age of four, three out of four children master plosive-liquid clusters (Templin, 1957). By the same age they cannot produce any fricative-liquid cluster (Powell and Elbert, 1984: cited in McLeod et al., 2001a). A number of scholars (Dyson and Paden, 1983; Greenlee, 1974; McLeod et al., 2001b; Smit, 1993; M. Watson and Scukanec, 1997) demonstrate that children go through a sequence of processes (developmental stages) that are interrelated and seem to be a prerequisite for mastery of consonant clusters. As the diagram in figure 4.1 indicates, a child may exhibit two stages at once but for different cluster types and in some cases may return to a stage for revision and refinement (Dyson and Paden, 1983; Smit et al., 1990; Temp1in, 1957). Cluster deletion is not reported for British English-speaking children in Dodd, Holm, Hua, et al. (2003) or Dodd, Holm, Crosbie, et al. (2013). Greenlee (1974) proposes that it is the first stage in cluster acquisition albeit rare. Consider the diagram illustrating stages of cluster acquisition inspired by the study of Greenlee (1974) in figure 4.1.

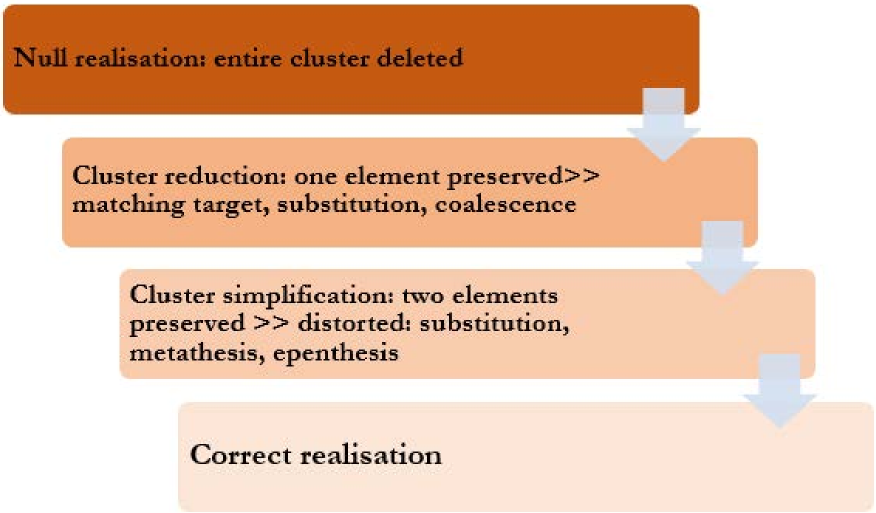


Figure 4.1 Developmental stages of consonant cluster acquisition

Cluster reduction is the second stage and the longest lasting (Shriberg and Kwiatkowski, 1980) as it may last for a few months, if not years. It is the most common developmental process in consonant cluster acquisition especially in younger children (Preisser, Hodson, and Paden, 1988; Roberts, Burchinal, and Footo, 1990; M. Watson and Scukanec, 1997). It is characterised by the reduction of a bi- or tri-consonantal cluster to one element. Several studies have investigated cluster reduction in various ages (Dodd, Holm, Hua, et al., 2003; Dodd, Holm, Crosbie, et al., 2013; Haelsig and Madison, 1986; Preisser et al., 1988; Roberts et al., 1990; M. Watson and Scukanec, 1997) and in male and female children (McCormack and Knighton, 1996). These studies unanimously reveal that cluster reduction negatively correlates with age and with the subsequent stage in cluster acquisition, that is cluster simplification (M. Watson and Scukanec, 1997). However, there were discrepancies in the rate of reduction per age. Thus, the findings of each study will be reported separately. The highest reduction rate (93%) was reported for children aged between 1;6 and 1;9 by Preisser et al., who examined 60 children. This rate reduced to 76% in children aged between 1;10 and 2;1 and later to 51% in children aged between 2;2 and 2;5. By comparison, M. Watson and Scukanec reported considerably lower figures in their sample – 46% and 48% – for children aged 2;0 and 2;3 respectively. For older children – 2;6, 2;9, and 3;0 – reduction rates continue to decrease to reach 34%, 25%, and 17% respectively. Interestingly, children aged 2;6 in Preisser et al.'s study report two sets of figures in their population sample – 39% and 59%. The first rate comes close to that reported for the same age group by M. Watson and Scukanec (1997). This is the rate of cluster reduction in female children. The rate for boys is 50% more than that for girls. This indicates that the girls' stage of consonant cluster acquisition at this age (2;6) exceeds that of age-matched boys. Roberts et al. reported higher figures in their sample of 145 children; 68% at the age of 2;6, 42% ate the age of 3;0, 25% at the age of 3;6. Haelsig and Madison's findings exhibit slightly lower figures; 30% at the age of 3;0, 18% at the age of 3;6. The latter two studies investigated developmental patterns in relatively older children and their findings suggest that cluster reduction becomes limited (<10%) after the age of five. The latter findings are corroborated in the two studies by Dodd, Holm, Hua, et al., 2003 and 2013 for British children. They conclude that the process becomes limited after the age of four to tri-consonantal clusters, which in turn becomes limited after the age of five. Attempts have been made to account for the influence of sonority in predicting which element in the cluster is preserved and which is deleted by Gnanadesikan (2004), D. Ohala (1999), and Pater and Barlow (2003). In onset clusters, the most common reduction pattern is one in which the more sonorous consonant of the adult target form is deleted and the least sonorous is preserved (D. Ohala, 1999).

Clusters can be reduced in two main ways; 1) deletion of one element, whilst the other is maintained, that is matching an adult form; 2) deletion of one element, whilst the remaining element does not match an adult form (Kirk, 2008). The latter case is referred to as coalescence. It occurs when a cluster is reduced to one consonant that is neither of the target elements but may share phonological attributes from each member (Pater and Barlow, 2003). It is typical around the age between 2;0 to 3;0 (Dyson and Paden, 1983).

Kirk (2008) also explored the acquisition of clusters by English-speaking children aged 1;5 to 2;7. The cluster types she examined were onset and coda clusters. Onset clusters include /s/ + stop, /s/ + nasal, consonant + glide, obstruent + /l/, and obstruent + /r/. Coda clusters she examined include nasal + /z/, stop + /s, z/, nasal + stop, and /s/ + stop. Word-final clusters that had liquids were not included in the study since the American dialect studied was a) non-rhotic and b) children at this age struggled with producing post-vocalic /l/ (usually glided) either in singletons or clusters correctly. The patterns she observed are summarised in table 4.1. Reduction was the most observable pattern at this age with 43% of the tokens. Predictable substitution is that based on the child's production of singleton consonants. This type of developmental pattern is found in 15% of the tokens examined in her study. Unpredictable substitutions are those which cannot be based on a child's production of singleton consonants. This type contributed to 6% of the tokens in the study (6% in onset clusters and 7% in coda clusters) which form a third of all substitutions in clusters. Her study shows that in onset clusters, children preferred labial place of articulation, whereas in coda clusters, their preference was for coronals.[14] C. Paradis and Prunet (1991) attribute this preference to the unmarked nature of coronals, especially that many languages restrict codas in general to coronals. In English, the majority if not all English coda clusters contain at least one coronal, with he exception of /ŋk, mp/ which share place of articulation Kirk (2008). Kirk (2008) categorised unpredictable substitution errors by place, manner and those involving both place and manner of articulation. Almost 70% of these unpredictable substitutions were motivated by a shared place (54%) or manner (12.5%) of articulation or both (2.5%), that is assimilation. 11% of these unpredictable substitutions resulted in dissimilation. Substitution errors involving change in manner of articulation are less frequent than the former (12.5%). 2.5% of substitutions involved assimilation of manner, 6.5% occurred in fricative + nasal, where nasality is lost to orality. substitution errors involving dissimilation of manner through fortition of fricatives, whereas 3.5% of substitutions involved members becoming *more* similar through lenition of stops. Fortition substitutions are argued to be the result of a preference towards a large sonority distance between the members (Gierut, 1999). 3.5% of substitutions involved

---

[14] Coronals in Kirk (2008) include alveolars /t, d, n, s, z, (l, r)/, and palatals /j, ʃ, ʧ, ʤ/.

confusions between /l, ɹ/ or between liquids and glides. 8.5% of unpredictable substitutions involved change in both manner and place. Of these, 2.5% involved assimilation of place with lenition and 2.5% with fortition. The remainder (3.5%) involved dissimilation in place and manner. Finally, 4% of unpredictable substitutions were difficult to categorise, for example /sta/ as [tsa], and /sməʊk/ as [tsoʊk].

Table 4.1 Developmental patterns in bi-consonantal clusters in English-speaking children in Kirk (2008)

Reduction	Correct	Predictable substitution	Unpredictable substitution	Deletion	Metathesis	Consonant insertion	Non-Schwa epenthesis
43%	34%	15%	6%	<1%	<1%	<1%	<1%

Kirk (2008) shows that epenthesis occurs mainly in word-initial clusters, where the second element is a sonorant. This process is rather controversial since adults seem to exhibit a similar pattern in colloquial speech especially when speaking emphatically (Smit, 1993). It is argued that the inserted schwa is a transition from the first consonant into the sonorant and does not represent a true epenthetic vowel. However, cases in which a vowel other than a schwa is inserted accounted for less than 1%. Other less frequent processes in English child acquisition are metathesis and consonant insertion (Kirk, 2008).

It is worth-mentioning however, Scobbie (1998) argues that whilst the presence of a segment/s may not be perceptible to the transcriber, the child may exhibit covert contrast indicating his/her acquisition/perception of the underlying structure and producing a contrast in what sounds to the adult listener as a homophone. In this respect, McLeod, van Doorn, and Reed (1998) compared the production of word final /sk/ and /st/ with word-final /-k/ and /-t/. They found evidence of compensatory (vowel) lengthening. This implies that a child may have acquired the structure CCV or VCC, for example and that due to production constraints, he/she may not produce a target structure fully and successfully.

As for vowels, according to Irwin and Wong (1983), they are mastered at the age of three with 100% correct rate but regress slightly at the age of four in English-speaking children. The whole phonemic inventory reaches adult-like maturity by the age of eight.

4.3 Children's acquisition of Arabic phonology

In this section, I have grouped published studies by variety of Arabic and then compared acquisition across varieties. Studies describing child acquisition of Arabic found in the literature concern Arabic spoken in Jordan (Amayreh and Dyson, 1998; Amayreh and Dyson, 2000; Amayreh, 2003), in Egypt (Ammar and Morsi, 2006; M. Saleh, Shoeib, Hegazi, and Ali, 2007), in Kuwait (Ayyad, 2011; Ayyad, Bernhardt, and Stemberger, 2016; Alqattan, 2015), and in Qatar (Al-Buainain, Shain, Al-Timimy, and Khattab, 2013).

4.3.1 *Jordanian Arabic*

Amayreh and Dyson carried out a series of normative studies on the acquisition of Educated Spoken Arabic consonants by Jordanian children (Amayreh and Dyson, 1998; Amayreh and Dyson, 2000; Amayreh, 2003; Dyson and Amayreh, 2002). Educated Spoken Arabic is a diglossia and a slightly less formal version of Modern Standard Arabic as it is syntactically less demanding (Zughoul, 1980). This variety is usually first introduced in schools when children are six years old.

The first in the series was that by Amayreh and Dyson (1998), which investigated 180 children aged between one year and two months and six years and four months. Their results showed that early consonants acquired prior to the age of 4;0 were /m, t, k, f, n, w, b, d, l/ with a 90% correct for acquisition criterion with the added consonants /q, ʔ, h, j/ when applying a 75% correct acquisition criterion. Middle consonants acquired between 4;0 and 5;0 were /s, ʃ, ħ/. Late consonants acquired after 5;0 were /r, j/. Some of the sounds that had not been acquired by the age of 6;4 include /θ, ð, z, ʤ/.

A follow up study (Amayreh and Dyson, 2000) investigated younger children aged between one and two revealed the appearance of earlier consonants /ʃ, ħ, ʕ/ in addition to the early consonants found in Amayreh and Dyson (1998) using a 75% correct threshold for acquisition but with the exclusion of the consonants /k, f/ for this age group.

Amayreh (2003) later investigated these sounds, amongst others that are considered late, in a follow up study. He grouped them into two categories. The first category is those sounds that are late due to lack of input from Educated Spoken Arabic and that have local variants in Jordanian colloquial speech. These are /dˤ, q, ð, θ, ðˤ, ʤ/. The second category is those sounds that are late due to inherent difficulties. These are /tˤ, dˤ, q, ðˤ, θ, ð, z, sˤ, ʕ/. Difficulties as he describes are due to them being typologically more marked. Results showed that within the younger age group, the only sound mastered (90% correct) was /ʕ/. An addition to this sound in the older age group was /z/. The rest of the difficult sounds had a correct rate below 90%.

Developmental processes reported for the non-mastered sounds include de-emphasis /tˤ, dˤ, sˤ/ – [t, d, s] respectively, devoicing, fronting /q/ – [k, g, ʔ], stopping /ð, θ, ðˤ/ – [d, t, dˤ] (respectively), and deaffrication /ʤ/ – [ʒ]. /θ/ was realised correctly 67% by children aged between 6;6 and 7;4 and 72% by children aged between 7;6 and 8;4.

Developmental errors included substitution by [t] 26% in each age group. He considered this substitution acceptable since it is the typical colloquial variant of the target consonant /θ/. The remaining 7% from the former age group substituted the target sound with [f], [ʔ], and [s] or deleted it, whereas the older group exhibited such substitutions and deletion at 3% only. The voiced affricate /ʤ/ was produced accurately 46% of the time by the younger age group, and 41% by the older age group. The most common substitution was [ʒ] (deaffrication) with 45% in the younger age group and 51% in the older age group. This was also considered an acceptable colloquial variant. Substitution with [z] was observed in 5% of each group. Other less frequent processes included stopping [d], deaffrication [ʃ], [ð] and deletion, all of which occurring at 5% or less.

Amayreh’s results also showed a higher accuracy rate for stops and affricates in onsets, whereas fricatives were either more accurate in codas or of equal accuracy.

4.3.2 *Egyptian Arabic*

In Egyptian Arabic, the phonemic inventory consists of 27 consonants including the emphatics /tˤ, dˤ, sˤ, zˤ/. /q/ and /ʒ/ had a relatively lower frequency in the dialect compared with other varieties of Arabic (Ammar and Morsi, 2006). The vowel inventory consists of eight vowels – three short vowels /i, a, u/ and five long ones /iː, aː, oː, uː, eː/ for Cairene Arabic (Ammar and Morsi, 2006). Additional vowels and diphthongs identified in the literature (M. Omar, 1973) include four short vowels /e, æ, ä, o/ and three long vowels /eː, äː, oː/ in Egyptian Arabic spoken in Sheikh Mubarak village. Six diphthongs are utilised in this variety, four in medial and final position and two in medial position only. Consider examples in table 4.2.

Table 4.2 Diphthongs in Egyptian Arabic

	diphthong	realisation	gloss
medial/ final	/æj/	/ʃæj/	‘tea’
	/aj/	/raj/	‘irrigation’
	/aw/	/ʕawza/	‘she wants’
	/iw/	/jiwʕid/	‘he promises’
medial only	/ij/	/mijja/	‘hundred’
	/uj/	/bujuːt/	‘houses’

Syllable shapes included CV, CVC, CVV, CVVC, and CVVCC (Ammar and Morsi, 2006). Studies exploring the acquisition of Egyptian Arabic found in the literature include Ammar and Morsi (2006), M. Omar (1973), and M. Saleh et al. (2007). The earliest consonants acquired in Egyptian Arabic comprised /b, t, d, m, n, ɣ, ħ, ʕ, j/ with a 50% accuracy rate threshold for acquisition reported in M. Saleh et al. (2007). They examined Egyptian children aged between one and two years and six months. Given the limited age range in their study and the low threshold assigned for acquisition, their study will not be dealt with any further. Early consonants in the Ammar and Morsi (2006) study were /b, t, d, k, g, q, ʔ, m, n, r, f, θ, s, ʃ, x, ħ, ʕ, h, j, w, l, ʤ, tˤ, sˤ/. They applied an acquisition criterion of 75% accuracy rate. Geminates are also reported to be acquired around the age of 3;6 (M. Omar, 1973). M. Omar (1973) demonstrates that /ʔ/ is initially acquired in word-initial position only. Word-medially, it started appearing after the age of 3;6 and it becomes phonologically stabilised around the age of four. Moreover, /r/ and /q/ in her study are reported to be late consonants acquired around the age of 5;0 and 6;6 respectively. Different vowels on the other hand are acquired at different ages. Short vowels were acquired before long vowels. Amongst the first to be acquired are /a, i, u/ *the fundamental triangle* (Jakobson, 1968) around the age of 18 months, /æ/ around two, and /e, o/ around two years and three months. Long vowels and diphthongs are acquired around the age of three.

Developmental patterns cited in Egyptian Arabic include final consonant deletion, velar fronting, voicing, de-emphasis (limited after 3;6), depalatalisation, /r/

lateralisation (persists until the age of five) (M. Omar, 1973). It is worth-noting that the sample sizes investigated by the above studies on Egyptian Arabic and the acquisition criteria do not allow for systematic comparison.

4.3.3 *Kuwaiti Arabic*

Kuwaiti Arabic consists of 29 consonants: nine plosives /b, t, tˤ, d, dˤ, k, g, q, ʔ/, two nasals /m, n/, 12 fricatives /f, s, sˤ, z, θ, ð, ðˤ, ʃ, χ, ʁ, ħ, h/, two affricates /ʧ, ʤ/, two liquids /l, r/, two glides /j, w/ and six vowels: three short and three long /i, u, a, iː, uː, aː/ (Ayyad, 2011). Kuwaiti Arabic also has the aspects of emphatics and gemination. Ayyad et al. (2016) and Alqattan (2015) studied the acquisition of Kuwaiti Arabic. Alqattan's study examined 70 children between the ages of one year and four months and three years and four months. Her results showed that early consonants included /p, b, t, d, k, g, ʔ, m, n, f, s, w, l, ɫ/ when applying the 90% accuracy threshold. A 75% criterion yields the additional consonants /r, z, ʃ, x, ħ, ʔ, h, j, ʤ, ʧ, tˤ, sˤ/. Consonants that were not acquired (less than 50%) by this age included /ŋ, v, θ, ð, ʒ, dˤ, zˤ/. The velar nasal is an allophone of [n]. The most frequent developmental errors reported in her study were de-emphasis (64%), deaffrication (32%), /r/ lateralization (28%) and stopping (11%). Less frequent processes include devoicing (3%), fronting (2%), and gliding (1%). Ayyad (2011) and Ayyad et al. (2016) studied 80 slightly older children aged between three years and ten months and five years and two months. The findings showed that early consonants included /b, t, d, k, g, ʔ, m, n, f, ðˤ, h, ʧ, ħ, rː, w, j/. Middle consonants comprised /tˤ, q, ʃ, χ, ʁ, l/. Children mastered many consonants including pharyngeal and uvulars. Consonants under development included coronal fricatives and affricates, the trill /r/, and some emphatics. Common developmental patterns reported in their study included de-emphasis, the neutralisation of coronal versus grooved[15] contrast in alveolar and inter-dental fricatives. These were the most frequent errors. The next most common error type was /r/ lateralisation, and stopping. Full and partial devoicing was also reported. Findings showed a decrease in error rates in the older age group. De-emphasis, deaffrication and stopping of fricatives are explained in terms of the low input frequency (Alqattan, 2015) or articulatory complexity (Dyson, 1988). Emphatics have the lowest frequency of occurrence in onset position in Kuwaiti Arabic (Alqattan, 2015).

---

[15] Their study was based on a phonological feature framework.

An exception to this is /ðˤ/. Stopping on the other hand, was attributed to articulatory complexity (Alqattan, 2015) as the frequency of input did not support it.

The acquisition criteria in the above studies was not consistent to allow for cross-dialect comparisons. However, studies using 75% correct acquisition criteria demonstrate that early consonants in Arabic dialects comprise stops, nasals and most fricatives. In Kuwaiti Arabic, a higher accuracy rate was observed in pharyngeals, uvulars and uvularised consonants of the four-year-olds compared with their Egyptian and Jordanian age-matched peers (Amayreh and Dyson, 1998; Ammar and Morsi, 2006; Ayyad et al., 2016). However, Jordanian children showed earlier mastery of coronal stops (Amayreh and Dyson, 1998). The earlier acquisition of /ħ/ was attributed to its higher functional load (represented in the higher morphological value) in Egyptian Arabic (Alqattan, 2015: 230) compared with the other varieties reviewed here. The prefix /ħ/ is a bound morpheme utilised in Egytian Arabic to mark future tense (Alqattan, 2015) – e.g. /ħa.naːm/ 'I will sleep'.

4.3.4 *Consonant clusters*

In Arabic, cluster reduction was shown to be at its peak (36%) in the speech of Kuwaiti children aged between 2;0 and 2;3 (Alqattan, 2015). This figure reduces to 16% in children aged between 2;8 and 2;11 and by the time they are aged between 3;4 and 3;7, it diminishes to a further 9%. A further examination of the frequency of cluster reduction across syllable positions revealed that it is most frequent in word-final position 43% of all reduction errors, with /-lb/ accounting for 56% of word-final cluster reductions, whereas word-medially 23% of all reduction errors with /-dɾ/ accounting for 55% of all reductions in this position. Word-initially on the other hand, exhibited 34% of all reduction errors with /dl-/, /ʃl-/, and /bɾ-/ accounting for 15%, 12% and 10% of all reductions in this position.[16]

The final stage before a cluster type is mastered, also referred to as cluster simplification, is a stage where the number of the elements in a cluster matches that of an adult but may exhibit change (in manner or place for example). This *change*,

---

[16] See Alqattan (2015: 196) for further details on error percentages per different substitution types within and across cluster types.

according to Greenlee is manifested in substitution. However, according to McLeod et al. (2001b) it involves any non-adult like production. The most frequent type of cluster simplification reported in Grunwell (1987) is liquid gliding – /l/, /r/ >[j], [w]. Kirk (2008) and Ingram (1976) confirm that a child may exhibit two stages at once but for different cluster types. Data from M. Watson and Scukanec (1997) indicates that cluster reduction and simplification co-occur in their sample of children aged between two and three. Whilst reduction decreases over time, simplification seems to negatively correlate with reduction until the age of 2;6, when it peaks (~46%) before it starts to decline – 33% and 30.8% at the age of 2;9 and 3;0 respectively. In another related vein, McLeod et al., 2001a carried out a typological study, in which they examined the productions of final consonant clusters in children aged two. Their goal was to determine whether the first, second, or both elements are likely to be substituted. Their findings demonstrate that the second element of a word-final cluster is more likely to be substituted than the first.

In Jordanian Arabic, very low percentages of syllable deletion, coda deletion and cluster simplification were reported in children aged between two and four years (Dyson and Amayreh, 2000).

For syllables, cluster deletion and simplification was evident in combinations that comprised at least one of the late consonants. Vowel epenthesis is also reported even in older children (aged 5;0) in certain combinations (M. Omar, 1973).

In Kuwaiti Arabic vowel epenthesis in clusters occurred 8% of the time in the productions of children aged between 2;8 and 2;11 and declined to 2% in children aged between 3;0 and 3;4 (Alqattan, 2015). In word-final clusters, the second most affected cluster type is /-lb/, which accounts for 18% of all cluster epenthesis productions.

4.4 Phonetic acquisition of plosives

The studies cited above deal with phonological acquisition (of phonemes in the case of segments) without investigating the fine-grained phonetic detail of segments or combinations of segments. Demuth and Song (2011) argue that there is a possibility of covert contrast that cannot be accessed using auditory-based transcription alone. They explain that,

> ... if the transcription indicates that the child produced no coda consonant on the word dog, it is impossible to know if the child's representation was really CV, or if there might have been vowel lengthening, indicating that the child has some knowledge of the "missing" coda consonant (Demuth and Song, 2011: 398).

Scobbie, Gibbon, William, Hardcastle, and Fletcher (2000) argue that there is a lag of approximately seven months between the time a child produces a *phonetic* (acoustic/articulatory) contrast and the time it is perceived auditorily by adults. The normative studies above tend to focus on phonemic rather than *phonetic* acquisition.

As discussed in Chapter 3, despite the similarity in the denotation of plosives across languages, the implementation of voicing contrast varies by language (Abramson and Whalen, 2017; Cho, Whalen, and Docherty, 2019; Lisker and Abramson, 1964). Such variation can be problematic for the L2 learner especially when the short-lag VOT category is assigned to different phonemic voicing categories in the native and target language. Both Libyan Arabic and English have a two way distinction in plosives. However, whilst English utilises a short-lag and long-lag voicing distinction between cognate plosives, Arabic utilises lead voicing and short-lag for the same voicing distinction. Consider figure 4.2 adapted from Khattab (2002b: 118) illustrating a representation of the VOT continuum which reflects the overlap between English and Arabic plosives.

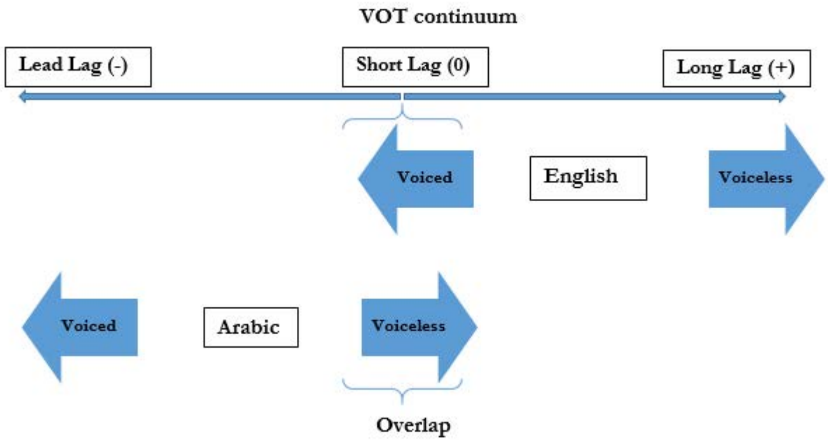


Figure 4.2 Voicing overlap between Arabic and English

A number of studies examined the acquisition of VOT in English-speaking children and report they acquire voicing contrast as early as 17 months old (Macken and Barton, 1977). In the time before that (around the age of six months) they exhibit 'uniform distributions along the VOT continuum' (Kewley-Port and Preston, 1974: 125). Both voiced and voiceless plosives are produced with a short lag VOT pattern. However, it could take a further 11 months for children's productions to reach a point, where adults perceive such contrast. A progress spurt in the voicing contrast is noticed around the age of two in American English (Macken and Barton, 1977) and British English (Foulkes, Docherty, and Watt, 1999). It could take months if not years to master adult-like articulatory skills to produce the voicing contrast in word-initial plosives (Macken and Barton, 1977; Whiteside, Dobbin, and Henry, 2003).

Studies on British English (Whiteside, Dobbin, et al., 2003) demonstrated that British monolingual children (from Sheffield) exhibit variability in their production of VOT word-initially that could last up to the age of 11;10. This is argued to be a precondition for continued refinement of oro-motor skills. There are also reports of gender differences in VOT durations in British children. Whiteside, Henry, and Dobbin (2004) examined VOT durations (for all plosives preceding high front and a low back vowels) in children aged between 5;8 and 13;2. Their findings showed that females demonstrate longer VOT durations than males. The difference heightens around the age of 13;2 and in voiceless plosives preceding a high front vowel /i/.

In Arabic-speaking children, it must be borne in mind that VOT patterns vary according to variety (see Chapter 3). For Arabic children, Khattab (2002b) examined VOT productions of three monolingual Lebanese Arabic children as part of the sample of her study. These children were aged five, seven and ten. Her results indicate that the five- and seven-year-olds' VOT patterns did not resemble Lebanese adult-like norms. The ten-year-old participant was the only speaker, compared to the five- and seven-year-old children, who had acquired full voicing lead for all three voiced plosives measured in the study /b, d, g/ and all the voiceless plosives /p, t, k/ and appeared to gradually progress towards adult values. Moreover, he generally had the most consistent patterns in VOT values. The five- and seven-year old participants' VOT values were more variable and had signs of incomplete acquisition of negative VOT marked by either short VOT for some of the voiced plosives or a very small negative VOT value. Consider 4.3 below as taken from Khattab (2002b: 101).

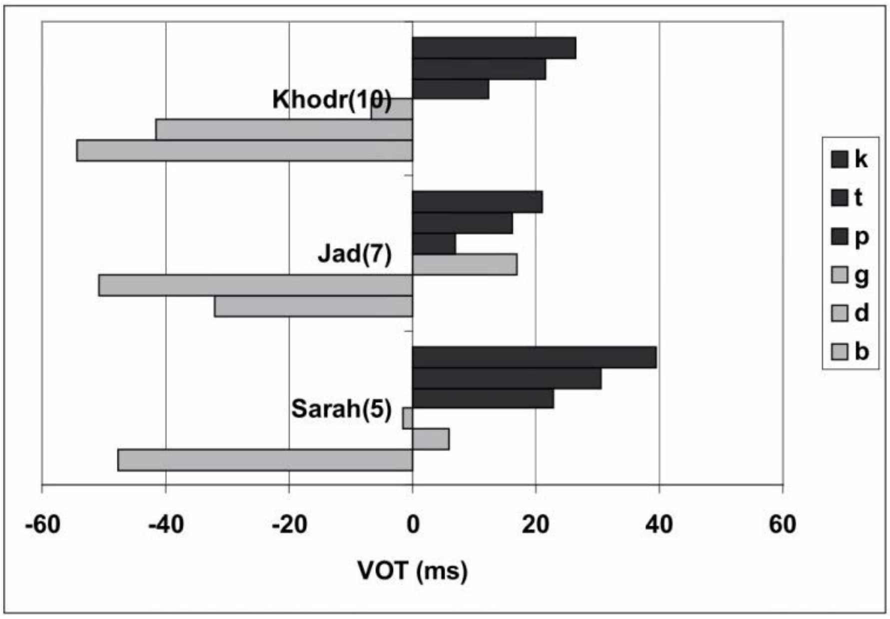


Figure 4.3 Mean VOT values for Lebanese Arabic child speakers adopted from Khattab (2002b)

However, the acquisition of VOT is not straightforward. Typologically, voiceless unaspirated plosives are acquired before either voiceless aspirated and prevoiced plosives (Jakobson, 1968). It is demonstrated that children acquiring languages of lead versus short-lag VOT (e.g. Arabic, French, Spanish) master adult-like patterns later than those acquiring languages using short versus long-lag VOT (e.g. RP English) (Allen, 1985; Khattab, 2002b). Macken and Barton (1980a) and Macken and Barton (1980b) observed differences in the production of word-initial plosives between Spanish-speaking and English-speaking children. Moreover, mastering adult-like acoustic-phonetic cues for long-lag VOT may extend to preadolescence (Whiteside, Dobbin, et al., 2003).

Various explanations have been put forward to account for such relatively protracted development of lead-lag VOT. One is perceptual (acoustic saliency) (Davis, 1995), one relates specifically to Mexican-Spanish and that is frequency of input, and one is articulatory (oro-motor coordination) (MacLeod, 2016).

Prevoiced VOT is argued to be acoustically less salient compared with short-lag and long-lag VOT (Davis, 1995). This has been examined and proven to be the case in infants aged between six and twelve months (Aslin, Pisoni, Hennessy, and Perey, 1981). However, children with L1 backgrounds that implement lead voicing versus short-lag, resort to strategies other than VOT to exhibit voicing contrast in plosives. This contrast is manifested in spirantisation for Spanish-speaking children (Macken and Barton, 1980b) and continuous voicing in French (Allen, 1985). This indicates that despite the reduced perceptual saliency, children can perceive the contrast but cannot produce adult-like long-lag VOT. Macken and Barton (1980b: 456) suggest that the late development of lead-lag in Spanish /b, g/ can be due to either the 'spirantization rule or extreme asymmetries in the distribution of stop phonemes'. Spirants in Mexican Spanish are the allophonic variants of voiced stops.

The articulatory account, on the other hand, posits that the acquisition of lead-lag and long-lag VOT is physically more demanding than short-lag (MacLeod, 2016). Difficulty lies in the timing coordination between laryngeal mechanisms – the initiation of vocal cord vibration – and supra-laryngeal mechanisms – the release of the plosive. In the production of short-lag VOT, these two mechanisms are either simultaneous or one follows shortly after. This coordination is compatible with younger children's motor speech skills (Green, Moore, Higashikawa, and Steeve, 2000). Lead-lag VOT, nonetheless, requires the adduction of vocal cords and pressure build up in oral cavity prior to the release (Kewley-Port and Preston, 1974).

4.5 Summary

This chapter provided an overview of the literature on English child language acquisition and Arabic child acquisition in terms of age of acquisition of consonants and vowels and consonant clusters. It discussed children's developmental errors and processes and compared these for each language. The following summary in table 4.3 provides typical developmental processes in British-English children.

Table 4.3 Phonological error patterns used by British-English children in Dodd, Holm, Crosbie, and Zhu (2013) and McIntosh and Dodd (2008)

Age group	Gliding	De-affrication	Cluster reduction	Fronting	Weak syllable deletion	Stopping	Voicing	Coda deletion
2;0–2;5	■	■	■	■	■	■	■	■
2;6–2;11	■	■	■	■	■	■	■	■
3;0–3;5	■	■	■	■	■	■		
3;6–3;11	■	■	■	■	■			
4;0–4;5	■	■	*					
4;6–4;11	■	■	*					

* In tri-consonantal clusters exclusively.

Table 4.4 provides a summary of typical developmental processes in Kuwaiti-Arabic, Egyptian-Arabic, and Jordanian-Arabic children. From the tables, several differences in developmental processes between English and Arabic children can be observed. These are summarised as follows:

- *(Bi)consonantal cluster reduction* or *simplification* cease at roughly similar times in both languages with slight inter-variety variations in Arabic.
- *Gliding of Liquids* persists longer in British-English children (4;6-4;11) than Arabic-speaking children (1;4-1;7).
- *(Velar-)fronting* persists longer in British-English children (3;6-3;11) than Arabic children (3;0-3;5).
- *Weak syllable deletion* persists longer in British-English children (3;6-3;11) than in Arabic children (4;0-4;5).
- *Deaffrication* persists longer in Arabic children[17] (up to 8;4 (Amayreh, 2003)) than British-English children (4;6-4;11).
- *Stopping* persists longer in Arabic-speaking children (7;6-8;4) than British-English children (3;0-3;5).

---

[17] Only those whose varieties they speak utilise affricates. In other Arabic varieties, which do not employ affricates, such developmental pattern is not applicable.

- *Voicing* errors persist longer in Arabic-speaking children (3;6-3;11) than British-English children (2;6-2;11).
- *Coda deletion* persists longer in Arabic-speaking children (3;6-3;11) than British-English children (2;6-2;11).
- Arabic children exhibit patterns which are not part of English children's phonological development, such as *de-emphasis*, *lateralisation*, *dentalisation*, *spirantisation*, *metathesis*[18], *stridency deletion*.
- In English:
    - clusters containing plosives are acquired before those with fricatives.
    - coalescence[19] and assimilation are forms of cluster reduction and simplification respectively.
    - In final consonant clusters, children may resort to compensatory lengthening of vowels.
    - post-vocalic /l/ is usually glided/vocalised in singleton or clusters.
    - epenthesis occurs especially in word-initial clusters, mainly if the second element is a sonorant.

---

[18] Metathesis and consonant insertion are reported in English phonological acquisition albeit they are less frequent (less than 1%) (Kirk, 2008).

[19] See page 73.

Table 4.4 Phonological error patterns used by Arabic-speaking children in Alqattan (2015) and Ayyad (2011) for Kuwaiti, Ammar and Morsi (2006) for Egyptian, and Amayreh (2003) and Dyson and Amayreh (2000) for Jordanian

Age group	De-affrication	Stopping: including stridency susbtitution	Lateral-isation	Dental-isation	Epenthesis	Weak syllable deletion	De-emphasis	Cluster simplification	Spirantisation	Voicing	Metathesis	Coda Deletion	Fronting	Stridency deletion	Gliding
1;4–1;7		Kuwaiti	Kuwaiti				Kuwaiti					Kuwaiti			Kuwaiti
1;8–1;11		Kuwaiti	Kuwaiti				Kuwaiti					Kuwaiti			
2;0–2;5		Jordanian Kuwaiti	Jordanian Kuwaiti			Jordanian	Jordanian Kuwaiti	Kuwaiti	Jordanian	Jordanian		Kuwaiti	Jordanian	Jordanian	
2;6–2;11		Jordanian Kuwaiti	Jordanian Kuwaiti			Jordanian	Jordanian Kuwaiti	Kuwaiti	Jordanian	Jordanian Kuwaiti			Jordanian	Jordanian	
3;0–3;5		Kuwaiti	Jordanian Kuwaiti	Jordanian	Kuwaiti	Egyptian Kuwaiti	Jordanian Kuwaiti	Egyptian Kuwaiti	Kuwaiti	Jordanian Egyptian	Kuwaiti	Kuwaiti	Jordanian		
3;6–3;11		Kuwaiti	Jordanian Kuwaiti	Jordanian	Kuwaiti	Kuwaiti	Jordanian Kuwaiti	Egyptian	Kuwaiti	Jordanian Egyptian	Kuwaiti	Kuwaiti			
4;0–4;5		Kuwaiti	Jordanian Kuwaiti	Kuwaiti	Kuwaiti	Kuwaiti		Jordanian Kuwaiti							
4;6–4;11		Kuwaiti	Kuwaiti	Kuwaiti	Kuwaiti										
6;6–7;4	Jordanian*	Jordanian*					Jordanian*								
7;6–8;4	Jordanian*	Jordanian*													

* These are particularly late because they are introduced later, that is in Educated Spoken Arabic only.

Chapter 5: Methodology

5.1 Introduction

In Chapter 3, we have seen that Arabic deploys a short lag VOT for phonologically voiceless plosives such as /t k/, whereas British English RP deploys a short lag VOT for phonologically voiced plosives such as /b d g/ making the two categories acoustically close in their VOT implementation. We have also seen the differences between the phonemic inventories of Arabic and English. For example the rhotic in Arabic can be trilled in geminated contexts and a flap elsewhere, whereas the English rhotic is an approximant and is classified as a non-rhotic variety. We have also seen differences in phonotactic constraints between English and Arabic, whereby the former allows more complex clusters and the latter is less complex. Additionally, we have seen how the voicing implementation as demonstrated by VOT implementation varies between the two languages. These two language backgrounds thus provide a rich ground for investigating L2 speech learning. We have also shown in Chapter 2, how L2 speech learning can be affected by markedness constraints and that not all L2 sounds are equally difficult. Furthermore, we have seen how literacy and orthography can influence L2 speech learning either positively or negatively and how various training studies vary in their aims and implementations albeit there is a general lacking of theoretical frameworks underpinning these studies. It was also shown how native language experience shapes the perception of *foreign* sounds and sound contrasts. The NLM's assumptions, for example, are based on initial exposure to foreign input, that is the first contact with TL input. It does not specify how extended exposure or phonetic training could affect perception of various TL sounds and if it does, it does not specify at which rate. Kuhl and P. Iverson (1995) suggest that discrimination abilities are not lost for life and that they can possibly be improved by extensive training. Research findings on adults, for example Japanese speakers learning English /r/ and /l/ contrast (Bradlow, Akahane-Yamada, et al., 1999; Logan et al., 1991), Mandarin speakers learning Canadian-English vowels (Thomson, 2011), American-English speakers learning Mandarin tones (Y. Wang, Jongman, and Sereno, 2003) show improvements in perception with extensive training. Kuhl and P. Iverson (1995: 142) argue, that

> change occurs at a higher level, one that involves memory and/or attention.... exposure to a given language results in the development of a speech representational system that alters the underlying perceptual system... reducing the prominence of certain distinctions when compared to the language-general initial state.

Difficulties in L2 learning Kuhl and P. Iverson (1995) argue, occurs at the phonological level. Similar to the assumptions by the SLM's equivalence classification, difficulty lies in the proximity of the target sound to the native magnet. The closer the target to the native magnet, the more likely it is to become assimilated to the native category making them indistinguishable. If this holds true, a TL sound, for instance British English RP /d/ with a short VOT and an Arabic /t/ also with a short VOT are acoustically close and would be indistinguishable. The Arabic prototype category for VOT is expected to perceptually attract TL VOT for /t/ and assimilate it to the native Arabic prototype. In this case, the L2 output for /d/ VOT, based on the NLM and the SLM, is predicted to be identical/homogeneous with their Arabic L1 VOT values for /t/.

According to the PAM-L2, if one member in a TL sound contrast is *phonologically* comparable to that of an L1 category but is *phonetically* different enough, the difference will be noticed and a new category will be formed. However, as is the case for the NLM and the Magnet Effect, the lack of clear and precise criterion for measuring similar and different, makes it challenging to accurately predict assimilation scenarios assumed by the model. For example, the Libyan Arabic native phonemes /r/ and /ɾ/ and the English phoneme /ɹ/ share a few articulatory characteristics but they are not identical in their entirety of articulatory gestural constellation. The PAM-L2 includes a postulation that neither of the NLM or the SLM models have regarded. It argues that the perceptual events in question depend on the listener's perceptual objectives or levels of attentional focus. This in turn varies from requiring attention at the gestural level to either the phonetic or phonological level. L2 perceptual learning does not depend solely on the phonetic level, rather, on all three levels contingent to the context of goals. The PAM-L2 also predicts that audio input is not the only cue in perception. According to Samuel (2011), individuals use a combination of auditory information from speech and the visual information such as lexical priming to assist speech comprehension. Rosenblum (2008) echoes the multimodal nature of speech perception indicating that learners also rely on visual cues such as the written text. In the case of single category assimilation for instance, learners can rely on orthographic cues to tease apart a distinction that cannot be made perceptually. Lexical minimal pairs may provide cues to the phonemic distinction for the two sound contrasts (Bundgaard-Nielsen et al., 2011).

On the other hand, the Hybrid model (Colantoni and Steele, 2008) proposes that less proficient learners rely more on the acoustic signal and less on the distributional patterns in speech perception. This suggests that difficulties in L2 speech learning are further position-sensitive. It is argued that novel articulations in L2 acquisition are mastered word-initially before those word-finally (Wenk, 1979; Wenk, 1983). The Hybrid model further argues that feedback has been shown to impact first language category acquisition (M. Goldstein et al., 2009). When L2 learners practise production, their outputs allow them to tune their productions further to match the target input. Krashen (1982) however, argues that the pressure to perform results in premature use of L2. Even though Krashen did not specify this for L2 pronunciation, it seems to be compatible with the role of the learner's own production in his/her perception.

However, neither of the studies on L2 speech have further investigated the role of delayed production by comparing it to immediate production practice – which allows learners the opportunity of getting feedback, that is, to compare their outputs to the modelled input – in beginner child learners of English from a different L1 background, that is Arabic. The present study, thus, asks whether and, if so, L2 pronunciation in seven-year-old children varies as a function of 1) presentation method, that is perception-only-practice vs. perception-and-production practice, and 2) input type, that is native vs. Arabic-accented. This leads to the following research questions.

RQ1: Which training method will result in more target-like (and less L1 like) pronunciations of problematic sounds Libyan learners typically display learning English, including:

1. Affricates
2. CC Coda clusters
3. Diphthongs
4. Plosives
5. Voiceless dental fricatives
6. Rhotic approximants

RQ2: What developmental processes do beginner Arabic child learners of L2 English exhibit in each condition?

RQ3: Which training method will result in most learned words?

If the learners do show signs of discriminable categories, it can be assumed that learning took place and that the training was effective. In this case, the source of input is typically versatile including a range of male and female speakers, from adults and other children. However, in the current study, the input from the training programme is restricted to two adult SSBE speakers – one male and one female, input from other participants and the learner's own productions (in the case of the Listen and Speak condition), which feeds right back into their perception. In the case of the delayed production (Listen Only) condition, the only source of input is the programme's two (one male and one female) adult speakers.

The following sections provide an overview of the participant conditions, the training programme, training materials, the stimulus and data elicitation procedures.

5.2 Training

In this section, the computer-assisted pronunciation training programme used in the study is described in detail. Also, the procedure followed in the training course is detailed.

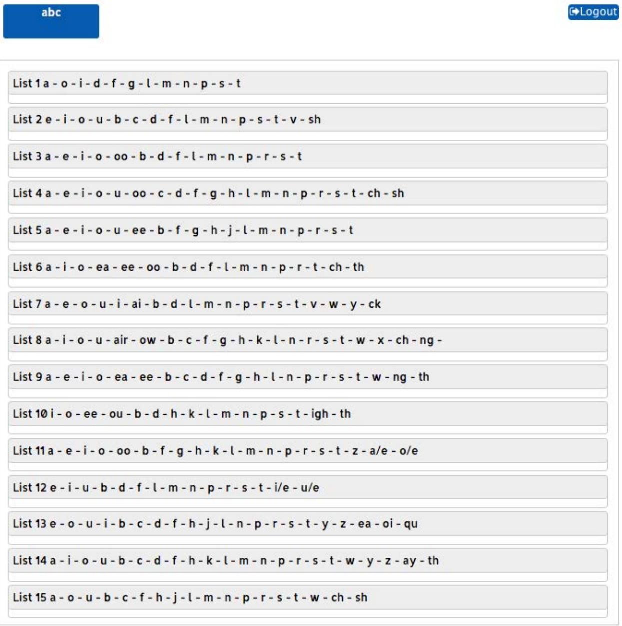
5.2.1 *The programme*

The programme designated for this investigation was the Digital Literacy Instructor (DigLin for short). DigLin was developed in 2013. It was designed to teach adult L2 beginners grapheme-phoneme correspondences with a set of vocabulary in Dutch, English, Finnish, and German (Cucchiarini, Dawidowicz, Filimban, Tammelin-Laine, Craats, and Strik, 2015). The choice of DigLin was based on the availability of an integrated Automatic Speech Recognition (ASR) function, which provides automated and instant feedback. The advantages of using this in pronunciation teaching have been asserted by Levis (2008: 184), who states that through ASR, learners gain individual feedback, something that a traditional teacher has practically little time to provide consistently and objectively. Cucchiarini, Neri, and Strik (2009) show that ASR feedback, when tuned according to specific pedagogical goals, can contribute to significant improvement in pronunciation compared to training without it. They studied a group of adult immigrants from heterogenous groups of L1 backgrounds learning Dutch in the Netherlands. All participants attended regular classes of a beginner course and self-study sessions in the language lab. They were additionally assigned to groups of differing conditions; a) using CAPT with ASR feedback on segmental quality; b) using CAPT without ASR feedback (they could still record their productions and compare them to example utterances relying on their own discrimination skills; c) no additional training to regular lessons. In a pre-test, post-test design, analysis was based on global segmental quality. Native speaker judges rated their segment pronunciation on a 10-point scale. The rating was followed by auditory analyses of a subset of the data by means of annotations of specific segmental errors. They concluded that the group that had CAPT training with integrated ASR feedback yielded the largest (and statistically significant) improvement (z = -2.827, p = .002, one-tailed) in comparison to the CAPT group without ASR feedback. Percentage of errors of target phonemes in the pre-test and post-test for the with ASR and without ASR feedback is 20% -13% and 12% - 9% respectively. This means that ASR feedback is both adequate and effective in improving pronunciation errors after a few hours over a month.

It was thought that this area - training *with* ASR feedback vs. training *without*

ASR feedback - might be worth investigating with child learners. However, when the training commenced, ASR did not work owing to security and firewall restrictions in Libya, where the training took place.

The programme has been designed for non-literate adult beginner immigrants in the Netherlands. The programme does not assume any previous knowledge, written or otherwise, of the target language. On this basis, it was deemed to be suited for children with no prior instruction or knowledge of English (orthography).[20] It comprised 15 word lists ascending in grapheme to phoneme complexity. DigLin utilises native speaker input from two Southern British English speakers, one male and one female.[21] Not only does DigLin teach pronunciation (via grapheme-phoneme correspondences), but also the written form and meaning through associated pictures. Thus, it teaches spelling, associates spelling and speech to picture-meanings, and allows the learner to work independently if they choose. See figure 5.1.



List 1 a - o - i - d - f - g - l - m - n - p - s - t

List 2 e - i - o - u - b - c - d - f - l - m - n - p - s - t - v - sh

List 3 a - e - i - o - oo - b - d - f - l - m - n - p - r - s - t

List 4 a - e - i - o - u - oo - c - d - f - g - h - l - m - n - p - r - s - t - ch - sh

List 5 a - e - i - o - u - ee - b - f - g - h - j - l - m - n - p - r - s - t

List 6 a - i - o - ea - ee - oo - b - d - f - l - m - n - p - r - t - ch - th

List 7 a - e - o - u - i - ai - b - d - l - m - n - p - r - s - t - v - w - y - ck

List 8 a - i - o - u - air - ow - b - c - f - g - h - k - l - n - r - s - t - w - x - ch - ng -

List 9 a - e - i - o - ea - ee - b - c - d - f - g - h - l - n - p - r - s - t - w - ng - th

List 10 i - o - ee - ou - b - d - h - k - l - m - n - p - s - t - igh - th

List 11 a - e - i - o - oo - b - f - g - h - k - l - m - n - p - r - s - t - z - a/e - o/e

List 12 e - i - u - b - d - f - l - m - n - p - r - s - t - i/e - u/e

List 13 e - o - u - i - b - c - d - f - h - j - l - n - p - r - s - t - y - z - ea - oi - qu

List 14 a - i - o - u - b - c - d - f - h - k - l - m - n - p - r - s - t - w - y - z - ay - th

List 15 a - o - u - b - c - f - h - j - l - m - n - p - r - s - t - w - ch - sh

Figure 5.1 The segments (as graphemes) in the word lists in the Digital Literacy Instructor

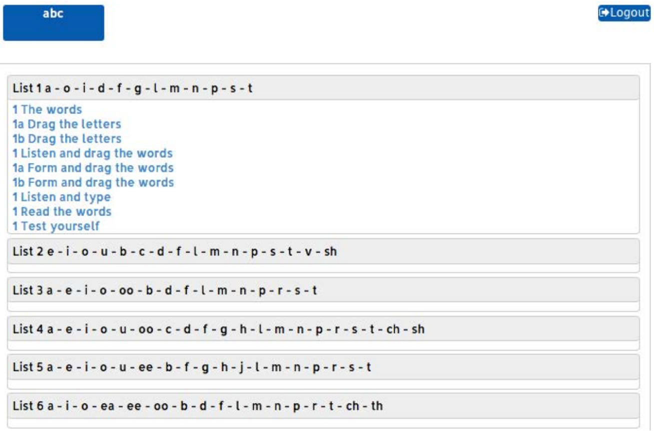
Each word list contains 20 monosyllabic words rendering 300 words in total. Within each word list, there were five different task types in addition to the words in that list,

---

[20] Developers of the programme have encouraged exploiting their software in other training situations (Cucchiarini, Dawidowicz, et al., 2015).

[21] The DigLin software was meant to be RP, but the speakers were Southern Standard British English – not strictly RP – speakers.

which have the pronunciation and a matching picture of all 20 words in that list as well as the test task ‘test yourself’. Tasks include drag the letters (a) and (b), listen and drag the words, form and drag the words (a) and (b), listen and type and read the words. See figure 5.2.



List 1 a - o - i - d - f - g - l - m - n - p - s - t
1 The words
1a Drag the letters
1b Drag the letters
1 Listen and drag the words
1a Form and drag the words
1b Form and drag the words
1 Listen and type
1 Read the words
1 Test yourself
List 2 e - i - o - u - b - c - d - f - l - m - n - p - s - t - v - sh
List 3 a - e - i - o - oo - b - d - f - l - m - n - p - r - s - t
List 4 a - e - i - o - u - oo - c - d - f - g - h - l - m - n - p - r - s - t - ch - sh
List 5 a - e - i - o - u - ee - b - f - g - h - j - l - m - n - p - r - s - t
List 6 a - i - o - ea - ee - oo - b - d - f - l - m - n - p - r - t - ch - th

Figure 5.2 Tasks within each word list

The first tab in the task list is the words. This task familiarises the participant with the target words in that list. The user can listen to the words and access the associated picture. Data elicitation, as can be seen later, is based on the ability to name the picture and read the word. Thus, all participants had to spend time on the words tab to become familiar with the target words. As shown in figure 5.3, next to each word, the user sees two circles: a bigger one and a smaller one. Clicking on the bigger circle plays the word and clicking on the smaller circle shows a picture of the word. At the bottom of the screen on which these are displayed, there are letters, which can play the corresponding sound for the user’s reference. See figure 5.3.

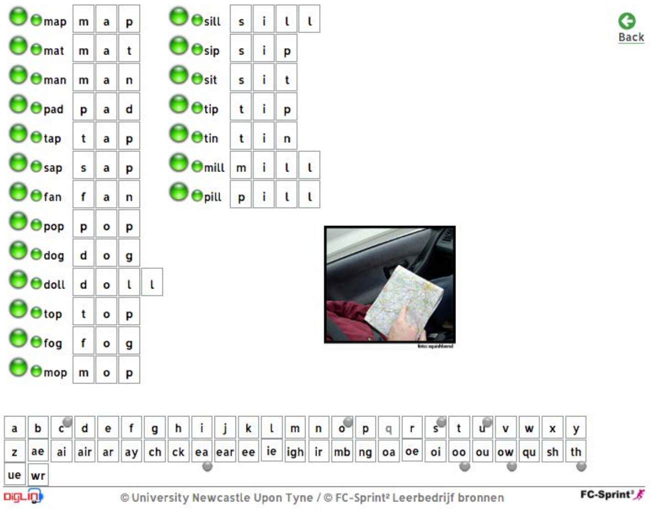


Figure 5.3 The words

In the following section, each task that was used in the present study is described. The allocation rationale of each task to the appropriate experimental condition is explained. The first experimental condition using the CAPT programme for listening and speaking practice is labelled Listen and Speak. The second experimental condition, the delayed production condition, using the CAPT programme for listening only is labelled Listen Only. The programme has 300 words but a subset of only 75 words (see table 5.2) was selected, based on two criteria: targeting problematic sounds typically found in Arabic adult L2 English (e.g. Al-Saidat, 2010; Avery and Ehrlich, 1992; J. Watson, 2002) and the number of words to be learned which was reasonable enough to fit into a three-week training programme for children.

5.2.2 *Procedure*

The training (learning of the words) took place in a language centre in a quiet room supplied with computers. However, the room capacity was enough for sixteen computers only. This meant that the training had to be rerun for additional participants. In addition, due to unexpected withdrawal of some participants, training had to be done two additional times excluding the original. That was to ensure the number of participants in each condition reached or approximated the intended number of participants.

Two assistant researchers supervised the experimental conditions. They spoke Libyan Arabic to the children. The children worked together in the groups to which they were assigned. The assistants were given 15 word lists for the children to learn using DigLin, one list per day. Each word list had 15 words, so that by the end of each week, participants had practised all of the 75 words selected from DigLin. They then reviewed the words the second week and reviewed them again in the final week. The order of the list of words was kept the same. In other words, the procedure that was carried out in week one, was repeated again in weeks two and three of the training. For the Listen and Speak condition, each word was allocated a three-minute listening practice and a minute of speaking practice.[22] For the Listen Only condition, each word was allocated an additional listening minute instead of the minute speaking practice. This was to eliminate the impact of differences in time allocated for each word between the Listen and Speak and Listen Only conditions. This yielded a total of four minutes of listening practice but were not required to produce the words. The assistant researchers were instructed to guide the children to work on specific words at a specific time as a whole group to reduce individual differences in practice time per word. On the first training day, the assistant demonstrated what each button on the word list does. For example, she told them that during the listening practice, they can listen to the whole word by clicking on the bigger circle or individual sounds by clicking on the relevant boxes next to the word. She also demonstrated how clicking on the smaller circle shows a picture depicting the meaning of that word whilst encouraging them to learn the words and their meanings. The choice of words/pictures may have not been child-targeted at times. For example, the depiction of the word pairs coin/pound, breast/chest, and broth/dish made them easily confused. Additionally, depiction of the words thin, broth, dish, and tool, for example, was not intuitive (see figure 6.1). Having said this, the participants still liked the programme and were very eager to use it whenever they can. They were, nevertheless, restless towards the second half of each session. That is because children generally have a limited attention span not because of the programme per se. Each training day, the assistant wrote a list of the target words for that day on the whiteboard. She then explained to the children that they were to spend some time practising each word on the list. She told them that they would start the listening practice by order of appearance on the board when she tells them. A

---

[22] The speaking practice inherently involved the participants listening to their own output. This allowed the participants of this group to reflect on their output and potentially modify their speech production throughout the training period as suggested by the Hybrid model in Colantoni and Steele (2008).

short amount of time is given to the participants to make sure they all located the target word successfully. The children were instructed to stop listening and start the speaking practice when they heard the timer set by the assistant. As for the Listen Only condition, when the participants heard the timer, it meant that they should start looking for the next word on the list on the whiteboard. Again, the timer did not start until the assistant made sure all participants had located the target word successfully.

The participants were children and sessions of sixty minutes were rather long for such an age. For this reason, a ten-minute break was allowed in the middle of each session: 30 minutes of practice, ten minutes' break and then another 30 minutes of practice. During the ten-minute break, children were entertained by a variety of activities. These included drawing, colouring in, stories and songs.

For the Traditional condition, teaching took place in two groups, one in the summer and one in November of the year during which data were collected. The summer group started off with 16 and ended up with only six participants. In recruiting the teacher, even though she was instructed to be 'traditional' in every sense of the word,[23] she was a Masters candidate in TESOL. It turned out to be the case that three of her six remaining participants performed exceptionally. For this reason and due to the small number of participants in the training, a second training in November was carried out. The latter training was carried out by a typical 'traditional' Libyan teacher of English. It included 20 participants. The teacher was kept ignorant of the training of experimental conditions using CAPT. This was to avoid the influence of such information on her teaching. The teacher spoke Libyan-accented English and was instructed to provide a three-week course to the children and include the test words and the same pictures as the experimental conditions within her teaching. These were sent to her in the form of a PowerPoint presentation. Each slide consisted of a word in both written and picture format within the same slide. The teacher was free to include, for example, grammar tasks or alphabet drills, for example, as are typical of traditional classroom teaching. She was instructed not to use any electronic devices in her teaching as traditional classrooms never use them. Teaching was meant to ideally take place in a different building from the experimental conditions. Therefore, children were taught in a primary school that they

---

[23] For more on what traditional teaching and the traditional curriculum typically involves, see Chapter 1.

attended. This was to avoid having participants from experimental conditions exposed to activities those in the Traditional condition did.

5.3 Participants

The participants were 58 Libyan children aged between 6;11 and 8;0 (+2 months in the delayed post-test). All participants were reported[24] to have had no prior instruction in English before the training whatsoever. Second language researchers (Flege, 1987b) emphasise the significance of having homogeneous participant groups owing to the inherent variability of speech in general and L2 speech learning in particular and all were born and had been raised in Libya. Prior to the training, parents and/or legal guardians filled in a background questionnaire that involved exclusionary variables such as prior instruction in English (potentially in private language schools) or systematic exposure to English either on television or the web. Any participant who had had such exposure was excluded from the training.[25] Also, questions relating to place of birth, places lived, parents' nationality and languages spoken at home were added. None of the children were reported to have any known hearing or cognitive problems. Because gender could be a potential variable, both boys and girls were included as equally as possible in each condition.

The participants were then divided into two groups. The Listen and Speak condition consisted of 20 participants: nine males and 11 females. The Listen Only condition consisted of 18 participants: 10 males and eight females. The Traditional condition consisted of 20 participants: 11 males and nine females. Once parents provided consent for the participation of their children, the list of volunteers were divided randomly into two conditions. For the practical reasons noted above, the Traditional teaching participants were drawn from the school where the Traditional teacher worked. They were all pupils from the same class attending grade 2 in that school. As noted above, the teaching took place in the local primary school the participants attended. Teaching sessions for the Traditional condition took place in a different location to avoid participants meeting and exposing details of their session to

---

[24] As indicated by their parents when signing consent forms and background questionnaires.

[25] It was learned later however, that some participants had siblings who have had formal instruction in English.

each other. This might have affected decisions of the teacher in the Traditional condition. Training sessions for experimental conditions took place in a language centre. A fully equipped classroom with computers was used.

The participants included three sets of twins, one set in the Listen and Speak condition, two in the Listen Only condition and one set divided between the Listen and Speak and Listen Only conditions.

Table 5.1 Study participants

Post-test	males	females
Listen and Speak	9	11
Listen Only	10	8
Traditional	11	9
Delayed post-test	**males**	**females**
Listen and Speak	7	7
Listen Only	9	8
Traditional	9	9

5.4 Design

A pre-test and post-test design can determine the effect of a Training and various conditions. The productions of the target words could be compared before and after the training and across the three conditions immediately after the training. However, there is no prior instruction in English before the training. English is introduced in the national curriculum at year five when pupils were between nine and ten years of age. In this case a pre-test was not possible. There is a remote likelihood that some pupils would have been admitted to English language summer courses. To eliminate this possibility, it was confirmed that none of the participating children had had any instruction in English either by asking their parents or by filling in a background questionnaire.

In addition to the post-test, a delayed post-test was carried out nine to ten weeks after the training. The same tasks were used to elicit production from participants across the three conditions and the two times of the tests.[26]

---

[26] The tasks used for data elicitation were mentioned in section 5.6.

5.5 Stimuli

Table 5.2 Test items

rhotic	dental fricative/	coda cluster		diphthongs		affricate
rain	thigh	/-ft/	left	/aɪ/	pile	chair
read	thin	/-ld/	cold		pine	chest
red	thread	/-lk/	milk		thigh	chin
rice	three	/-lt/	belt		rice	chip
rip	thumb		quilt	/aʊ/	mount	chop
rob	bath		salt		mouth	march
roll	broth	/-mp/	jump		pound	jam
root	moth	/-nd/	pound		round	jar
rose	mouth		sound		sound	jeans
round	path	/-nt/	paint	/eɪ/	paint	jet
wrap	tooth		mount		rain	jog
wreck			point		tail	jug
wrist		/-nz/	jeans		waist	jump
wrong		/-st/	breast	/ɒʊ	cold	
beard			list		rose	
chair			vest	/eə/	chair	
fear					hair	
hair				/ɪə/	beard	
hear					fear	
jar					hear	
march				/ɔɪ/	boil	
					coin	
					point	

In the present study, stimuli of problematic sounds were divided as shown in table 5.2. These items were grouped on the basis that the target English and the L2 English counterpart as produced by Libyan adult speakers are not the same phoneme. English /ɹ/ is typically produced /ɾ/, /θ/ as /t/, /ʧ/ and /ʤ/ as /ʃ/ and /ʒ/ respectively. As with diphthongs, they are monophthongal and coda clusters usually exhibit epenthesis. Thus, impressionistic transcription should be able to reveal differences in production. However, for sounds that were transcribed similarly across the two languages, transcription would not reveal discrepancies in production. For example, voicing implementation in plosives across languages varies. For this reason, stimuli including Libyan Arabic was deemed necessary to compare voicing and place of articulation. Another reason for eliciting Arabic data from the participants was because at around seven years of age, children do not reach adult-like phonetic maturity in all aspects of their speech production. For example, the findings of SA Lee and G. Iverson (2012) show that five-year-old Korean-English bilinguals distinguished word-initial plosives on the basis of VOT only, whereas the ten-year-olds (Korean-English bilinguals) distinguished word-initial plosives on the basis of VOT *and* vowel-onset fundamental frequency. Hazan and Barrett (2000) argue that reaching adult-like norms could take up to 12 years of age.

Thus, these kinds of stimuli were subdivided into Arabic and English. They were composed of monosyllabic words consisting of a bilabial or coronal plosive syllable-initially followed by each of the high front vowel /iː/, high back vowel /uː/ and low front vowel /a/. These vowels were comparable in Arabic and English and it has been demonstrated that acoustic values of various cues vary according to the following vowel such as VOT (B. Smith, 1978) and spectral tilt as it correlates with length of front cavity (Suchato, 2004b). See table 5.3 below.

Table 5.3 Test items 2 (Arabic and English)

	Arabic		gloss	English	
/p/	–			pad	/pad/
				pea	/piː/
				push	/pʊʃ/
/b/	bab	/baːb/	'door'	bat	/bat/
	beer	/biːr/	'well'	bee	/biː/
	booq	/buːq/	'trumpet'	bush	/bʊʃ/
/t/	taj	/taːʒ/	'crown'	tap	/tap/
	teen	/tiːn/	'fig'	tea	/tiː/
	toot	/tuːt/	'berries'	tool	/tuːl/
/d/	dal	/daːl/	'the letter d'	day	/deɪ/
	deek	/diːk/	'rooster'	dish	/dɪʃ/
	dood	/duːd/	'worms'	dog	/dɔg/

Apart from /buːq/, all the Arabic test items have a high frequency in LA, and the participants were expected to be familiar with them and produce them effortlessly. In addition to the above, spontaneous speech data were collected from a subgroup of the participants playing over an iPad game. Data from the Traditional teacher were also collected to allow for meaningful comparisons with her group of participants.

5.6 Data collection

The study examined these L2 learners' production. Therefore, data consisted of oral production in the form of audio recordings of the learners in all three conditions. This data collection technique was non-invasive, which makes it especially suitable for child participants. It was also suitable for recruiting a larger number of participants compared to other oral production data elicitation techniques. Another reason for opting for this data collection technique was its feasibility through using a portable recording device in the country where the study took place. The recordings were made using an Edirol R-05 in a quiet room in the language centre, where the training took place.[27] The recording device was placed at a 45 degree angle. Data from the Traditional condition learners were collected using the same recorder but in the primary school which participants attended. Some of these recordings had some background noise as a result.[28]

bee

Figure 5.4 An example of the presentation of the English test item 'bee' in the read aloud task

English data were elicited by means of two tasks, that is a read aloud task and a picture-naming task. Stimuli were presented on PowerPoint slides and randomised to elicit two productions from both tasks yielding two tokens per test word, with a total of 150 expected tokens.[29] In the read aloud task, each test item was presented individually on a slide. Consider figure 5.4. In the picture-naming task, again each test item was also presented individually on a slide. Consider figure 5.5.

---

[27] The room where the recordings took place was not sound-proofed and although great care was made to maintain a noise-free environment, occasional noises and noise from air conditioners were inevitable.

[28] Tokens which were affected by noise were discarded in the acoustic analysis.

[29] The stimuli were expected to render 75 tokens from the English read aloud task, 75 tokens from the English picture-naming task. See section 5.5.

Figure 5.5 An example of the presentation of the English test item ‘bee’ in the picture-naming task

A 5 seconds interval was allowed for each slide. Visual stimuli (pictures) for the delayed post-test were different from those for the post-test and the training visual materials to avoid familiarity and predictability of task. These two tasks did not yield enough tokens for analysis. The Arabic data was elicited in the same way as the English data. Children in the Arabic read aloud task were presented with each Arabic test item on a slide which lasted for 5 seconds before the next test item appeared on the following slide. In the picture-naming task, the picture of each test item was presented on a slide for 5 seconds before the next test item appeared on the subsequent slide.

Not all of the children in the sample had developed literacy in Arabic to read the words. Due to the fact that some participants struggled to read Arabic materials and then struggled to remember the English words from the training and read them during the read aloud task,[30] a delayed repetition task was added to gather more data. The delayed repetition task yielded the vast majority of tokens. In this task, the participants were presented with an audio clip – from DigLin – of each test item followed by an intervening phrase in Arabic /tawːa ʔawədha marːa taːnia/ meaning “now repeat it one more time”. This type of delayed repetition approach following Flege, Munro, et al. (1995a) arguably reduces the chances of relying on short-term phonological memory of the modelled audio clip. Not using an intervening phrase risks the use of close imitation. For those who struggled to read the Arabic testing materials (See table 5.3), the assistant researchers prompted them using delayed repetition. Reading the material from the slides posed an additional methodological issue:

---

[30] It was not clear whether children struggled to read the words in the English read aloud task because they did not remember them or because they did not develop literacy in English. We know that they struggled to remember words in the picture-naming task as well, but they remembered relatively more words in the picture-naming task compared to the read aloud task.

interference from Modern Standard Arabic (MSA) for those Arabic-literate participants. It seems that some literate participants were influenced by Arabic orthography especially for words that are used in both MSA and Libyan Arabic, despite not using Arabic diacritics. The sole Arabic test item that was exclusively colloquial was ‘beer’ in this case. For the rest of the Arabic materials, a sub-group of the participants who successfully managed to read the words, attached the indefinite suffix /-un/ to the test items making the data tokens bi-syllabic. Great care was taken to select a repetition[31] that was free of such inflection. This was to make sure all the tokens were monosyllabic and produced in the local variety of Libyan Arabic and not in Modern Standard Arabic. Establishing the participants’ literacy level in Arabic was based on their ability to read the Arabic materials. Literacy level in English however, could not be determined in a systematic way albeit there were some some clear cases of orthographic influence in the read aloud task data.[32] The data for each speaker was collected in a single session which lasted approximately between 20–30 minutes per speaker.

Table 5.4 demonstrates the number of analysable words per task across the three groups in the post-test. As mentioned above for each task, each participant was presented with 75 test items, in the form of written words in the read aloud task, pictures in the picture-naming task, and finally an audio clip of the word as it appeared in DigLin embedded in a carrier phrase. This culminates to 225 English stimuli per participant per test, 13050 target words (for the 58 participants) in the post-test and 10800 target words (for the 48 participants) in the delayed post-test. The percentages in tables 5.4 and 5.5 below were calculated by dividing the number of produced words by the expected total (derived from the stimuli).

---

[31] The vast majority of participants tended to repeat test items even when they were not directly asked to.

[32] This was judged based on some substitutions in sounds based on similarities in English letters. For example, the replacement of ‘d’ with ‘b’ and vice versa.

Table 5.4 Analysable words per task across training conditions during the post-test

	read aloud	picture-naming	delayed repetition	total
LISTEN AND SPEAK	15	67	1401	1483
n = 20	(1%)	(4.5%)	(93.4%)	(33%)
LISTEN ONLY	4	39	1242	1285
n = 18	(0.3%)	(2.9%)	(92%)	(31.7%)
TRADITIONAL	10	211	1365	1586
n = 20	(0.7%)	(14.1%)	(91%)	(35.3%)

The procedure carried out to elicit data during the delayed post-test was the same as that in the post-test. Table 5.5 below presents the number of analysable tokens per task across the three groups in the delayed post-test.

Table 5.5 Analysable words per task across training conditions during the delayed post-test

	read aloud	picture-naming	delayed repetition	total
LISTEN AND SPEAK	5	30	921	956
n = 13	(0.5%)	(3.1%)	(94.5%)	(32.7%)
LISTEN ONLY	3	28	1217	1248
n = 17	(0.2%)	(2.2%)	(95.5%)	(32.7%)
TRADITIONAL	2	142	1160	1304
n = 18	(0.1%)	(10.5%)	(85.9%)	(32.3%)

5.7 Instrumentation and measurements

Once data recordings were collected, they were digitised using Praat software (Boersma and Weeninck, 2016). Data were recorded at 44 kHz sampling rate, stereo channel. The data tokens were first transcribed impressionistically in Praat. The impressionistic transcription served to answer the research questions:

RQ1: Which training method will result in more target-like (and less L1 like) pronunciations of problematic sounds Libyan learners typically display learning English, including:

1. Affricates
2. CC Coda clusters
3. Diphthongs

4. Plosives

5. Voiceless dental fricatives

6. Rhotic approximants

RQ2: What developmental processes do beginner Arabic child learners of L2 English exhibit in each condition?

RQ3: Which training method will result in the most learned words?

It is thus hypothesised that:

H1: The Listen and Speak condition will result in the most target-like productions/least L1 like productions followed by the Listen Only, leaving the Traditional Teaching condition with the relatively least target-like productions/ most L1-like productions.

H2: The CAPT training conditions will resemble *relatively more* of the English child phonological developmental stages and *relatively less* L1 interference in comparison with the Traditional training condition.

In some cases, it might be impossible to decide whether a process observed in a given speaker's interlanguage phonology is a developmental process or the result of L1 interference if it is reflected in both English child language acquisition and adult interlanguage phonology. However, it is those patterns that can be seen in *either* child L1 acquisition *or* adult L2 acquisition that factor in the decision making. For example, whilst deaffrication is observed in both child L1 and adult L2 acquisition, the nature of deaffrication between the two groups varies: in British-English speaking children, deaffrication of /ʧ/ and /ʤ/ is in the form of [t, s, ts] and [d, z, dz] respectively (Ingram, Christensen, Veach, and Webster, 1980). No observation of deaffrication into [ʃ] or [ʒ] are observed (Ingram et al., 1980) (see Chapter 4). In Arabic adult L2 acquisition, it has been found that learners confuse /ʧ/ with /ʃ/ (Altaha, 1995; Kharma and Hajjaj, 1989) and /ʤ/ with /ʒ/ (Barros, 2003; Kharma and Hajjaj, 1989). Evans and Alshangiti (2018) found that Saudi-Arabic adult learners of English – both low and high proficiency – confuse /ʧ/ mainly with /ʃ/, and also less frequently with /ʒ/, /ɡ/, /ʤ/ and /h/. Their participants also confuse /ʤ/ with /ɡ/, and /ʒ/. Production-wise, studies on L2 learners of English from various L1 Arabic varieties (Ababneh, 2018; Al Yaqoobi, Ali, and Sulan, 2016; Hago and Khan, 2015; Jabali and Abuzaid, 2017) demonstrate that Arabic adult L2 learners of English tend to substitute English /ʧ/ with /ʃ/ almost always exclusively.

As for the acquisition of dental fricatives by adults and children, Saudi Arabic learners, for example, have been found to confuse /θ/ with /f, ð/, the latter being ascribed to inconsistency in letter-to-sound correspondence (Evans and Alshangiti, 2018). Shafiro, Levy, Khamis-Dakwar, and Kharkhurin (2013) found that early Arabic-English bilinguals from various native Arabic varieties and English dialectal backgrounds confuse /ð/ with /v/ in /ɑðɑ/ and /iði/ but not /uðu/ contexts. They attribute this to the higher spectrum as well as the greater difference in spectral centre of gravity between /ð/ and /v/ in the /u/ context (2407 and 4716 Hz respectively) compared with that in the other vocalic contexts. This rather more salient difference is what helps the learners identify the correct place of articulation for the two consonants in this context. The effect of universals is also evident in the relative difficulty of the problematic sounds to each other. It was shown earlier in section 2.2 and Chapter 4 that fricatives are less marked and acquired earlier by children than dental fricatives or rhotic approximants. As a result, it is expected that the participants would have less difficulty with affricates than they would with either dental fricatives or rhotic approximants and thus, this is an influence of language universals.

To answer the research questions, every token was assigned a token ‘target-like’ if its transcription holistically matched the IPA transcription for the item, or a ‘non-target-like’ if it did not, irrespective of the type of input received by the training condition. However, there were some exceptions to this. Mismatches in final voicing were not considered errors in this rating. This was because final devoicing in L1 is well-documented in children’s speech (Jakobson, 1968; Stoel-Gammon, 1987; Temp1in, 1957). An account of voicing hierarchy has been put forward to explain the role of universals especially in final position. Final position is considered the most marked for voicing whereas initial position is the least marked. In addition to final devoicing, /l/ vocalisation is not considered an error in the analysis either. This is because /l/ vocalisation/vowelisation is very common in children learning languages that employ velarised post-vocalic /l/ (Grunwell, 1985; Hodson and Paden, 1981; Stoel-Gammon and Dunn, 1985) and in adult speakers of SSBE, where some articulatory studies, among others, demonstrate a weakened tongue blade-alveolar ridge contact (Scobbie and Wrench, 2003; Trudgill, 1984; Tollfree, 1999; Hardcastle and Barry, 1985). It is also worth noting here that the productions of the DigLin speakers reflected recent changes that are common in SSBE. For example, the diphthong quality as realised by the DigLin speakers was [ɒʊ]. The diphthong /ɪə/ sounded monophthongised in the word 'hear' for example but not in 'beard'. This is a result of a process known as *smoothing* (Kortmann and Upton, 2008). As such, data tokens were compared directly to these DigLin realisations. These points aside, for a token to be rated target-like, the segments must match for place and manner of articulation as well as voicing in word-initial and word-medial positions. The target-likeness rating helped to answer the first research question. Owing to the fact that the Traditional teaching condition received a different type of input by the traditional foreign-accented teacher and that the input received by the experimental condition did not capture the full range of variability in SSBE, an additional measure, match rating was also carried out. In this measurement, special attention was focused on problematic sound class in each token. Tokens were given a label of the corresponding sound class of interest, vis–a–vis affricate, cluster, dental fricative, diphthong, plosive, rhotic. The experimental conditions' tokens were cross-examined with those of DigLin and the Traditional condition's tokens were cross-examined with their teacher's realisations and each were rated accordingly. It is worth-noting here that, unlike the experimental conditions whose input was less varied, the data from the Traditional teacher was only a snap-shot of the variability of her actual pronunciation. To minimise this issue as much as possible, the variations of each sound class of interest were taken into account when labelling child tokens as matching or non-matching. Variability in the realisation of problematic sounds was considered from a range of tokens. When a speaker had an overall lateralisation in all words including Arabic, this feature was not considered in the decision whether the production was 'non-target-like', 'matching' or 'non-matching'. Similarly, when a speaker had a lisp, it did not contribute to non-target-likeness.

It is also worth-mentioning here that some sounds from the test items in DigLin were not accurate with respect to the variety they were meant to represent. This could have been due to the low frequency sampling rate of recorded materials for the software and/or un-revised productions. Test items such as 'path', for instance, were perceived as /p^{h}ɑːs/ and a word like ‘rose’ was perceived and realised as /b̥rəʊz/. Consider figure 5.6 below, where the first portion of the waveform illustrates a transient of an intrusive plosive preceding the rhotic approximant. As such, these mispronunciations are tolerated from the participants and are marked target-like accordingly.

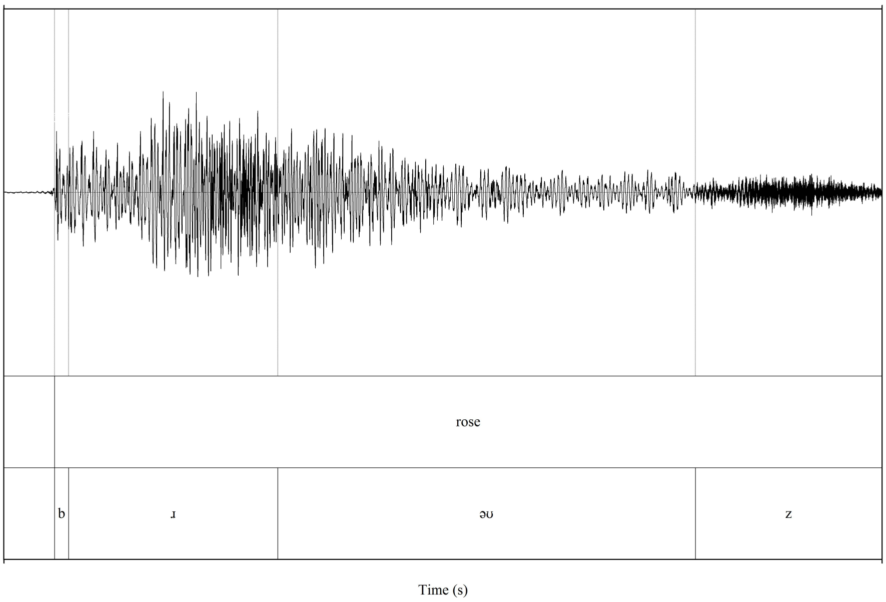


Figure 5.6 An illustration of an intrusive plosive prior to the test item /ɾəʊz/ as realised by one of the two DigLin speakers

Nonetheless, some participants managed to produce tokens of such DigLin sounds correctly. One possible reason for this could be that such participants were relying on orthography. As seen earlier in figure 5.3, participants have access to whole words and individual sounds in words. A comparison of the discrepancies in perception across sounds with matching denotations (in this case ‘th’) may have led learners to conclude that the odd example was a technological error. Rosenblum (2008) proposes that speech perception is multimodal when it comes to creating percepts of sounds. Learners rely on visual cues such as orthography and facial gestures alongside auditory cues. Although it was presumed that participants were at the start illiterate in the TL, developing initial stages of English literacy during training or support of literacy outside the training was not unlikely.[33] DigLin is designed to ascend in grapheme to phoneme correspondence

---

[33] Parents throughout the training asked assistant researchers to provide a list of the course words so that they could help their children practise at home. This is typical of the Libyan educational culture (Orafi and Borg, 2009). It is possible that when parents were denied these lists, they may have resorted to supporting their children with whatever materials they could use.

regularity throughout the word lists raising the user's phonemic awareness alongside other aspects of training (Cucchiarini, Dawidowicz, et al., 2015).

In terms of lexical learning, tokens that were produced by the children in full and were a full match to those of the input were labelled 'full'. Only tokens labelled 'full' were considered for the lexical learning analysis. These tokens represent full learning of test items and everything else considered partial learning was discarded from the analysis.

It should be pointed out here that the amount of data produced by learners differed in the delayed repetition task from both the read aloud and picture-naming tasks. In the delayed repetition task, learners were *generally* able to produce the vast majority of what they had heard from the audio prompt, but in the memory recall tasks – read aloud and picture-naming – learners could only produce the words they learned, and not all children had learned all the test words.

Following the collection of analysable tokens, they were filtered based on full versus partial realisations as defined above. Training conditions were compared to DigLin realisations. Traditional learners' tokens were compared to those produced by their teacher during data collection (not training period). It must be noted here that realisations of the Traditional teacher may not necessarily capture the whole variation in her speech throughout the teaching, thus reflecting the true input the children had received during training. Additionally, it is possible that the teacher's performance varied as a function of *register*. However, great care was taken to establish consistencies through the multiple realisations she produced during data collection. Based on this, only the tokens deemed as 'full' were considered for this part of the analysis.

For each task, each participant was presented with 75 test items, in the form of written words in the read aloud task, pictures in the picture-naming task, and finally an audio clip of the word as it appeared in DigLin embedded in a carrier phrase. The interpolation of such carrier phrase can help minimise direct imitation from sensory memory (Flege, Birdsong, Bialystok, Mack, Sung, and Tsukada, 2006). This culminated to 225 stimuli items per participant per test totalling 13050 target words (for the 58 participants) in the post-test and 10800 target words (for the 48 participants) in the delayed test.

As for the phonological processes that Arabic child learners of L2 English exhibit in their early learning stages, a comprehensive qualitative analysis described these and compared them to phonological processes exhibited by British RP children as well as phonological acquisition universals. Tokens labelled according to the respective sound class were grouped together under that sound class and cross-examined with their counterparts from the same subgroup. They were then examined for phonological processes and assigned a label of the phonological process. Table 5.6 illustrates a sample of phonological process coding for the sound class 'clusters' (/-ld/) and 'diphthongs' (/-əʊ-/) in the test item 'cold'.

Table 5.6 Sample coding of phonological processes in the test item 'cold' demonstrating the diphthong and CC coda cluster sound classes

item	Diphthong	Process	CC Cluster	Process	Transcription
cold	/ɒʊ/	monophthonging	LD	vowel epenthesis	/kɔːlɨd̥/
cold	/ɒʊ/	monophthonging	LD	substitution	/kɔˑnd/
cold	/ɒʊ/	correct	LD	reduction	/kə̱ʊd̥/
cold	/ɒʊ/	monophthonging	LD	correct	/kʰɔːɫdʰ/
cold	/ɒʊ/	correct	LD	reduction	/kʰʌ̃ʊdʰ/
cold	/ɒʊ/	correct	LD	deletion	/kʌu/
cold	/ɒʊ/	v2-gliding	LD	reduction	/kʰʌʊɨd/

These processes were also examined in the light of child phonological development discussed in Chapter 4 and relevant models of speech learning discussed in Chapter 2. Seventy-five test items were selected: 23 involving diphthongs; 16 involving CC codas; 11 involving dental fricatives; 21 including rhotic approximants, that is 14 word-initial and seven post-vocalic; 14 involving plosives (see table 5.2). There seems to be an imbalance for token count across test items with diphthongs having the most test items followed by rhotic approximants. A reason for this was the involvement of some test items in more than one sound class. For example 'beard' was analysed for the diphthong and post-vocalic rhotic approximant.

5.7.1 *Acoustic measurements*

For plosive-related measurements, Praat software was used to generate wide-band spectrograms and waveforms due to the better time resolution than narrow band spectorgrams (Abramson and Whalen, 2017). This was to procure visual inspection and segmentation of tokens of interest. Phonemic contrasts in plosives in RP English and Libyan Arabic were based on place of articulation and voicing categories. Both languages have *phonologically* voiceless and voiced velar stops /k, g/,[34] voiceless and voiced coronal stops /t, d/ and for the labial place contrast, there were voiceless and voiced stops in English /p, b/ but only voiced in Arabic for this place of articulation /b/. Despite the similarity in labelling across the two languages, the phonetic implementation of the contrast varies across languages (Tillmann, 1995). For voicing, the temporal properties of the release of the plosive relative to laryngeal activity (VOT) is considered one of the key correlates widely adopted for measuring voice contrast in a wide range of languages with binary distinction (Abramson and Whalen, 2017; Lisker and Abramson, 1964). If glottal pulsing initiates prior to the release it falls into the lead category, whereas if it initiates simultaneously with/after the release it falls in the lag category. English utilises a contrast between long lag (aspirated) for voiceless stops and short lag (plain) for voiced stops, whereas Arabic utilises a contrast between short lag (plain) for voiceless stops and lead voicing (pre-voiced) for voiced stops.[35] Thus, according to this specification, Arabic is considered a voicing language and English is an aspirating language (Abramson and Whalen, 2017). However, the voicing contrast based on VOT alone has been contested. Some studies on British English demonstrated a presence of prevoicing in *phonologically* voiced stops (Docherty, 1992). In Arabic, some evidence for partial devoicing in *phonologically* voiced plosives and aspiration in *phonologically* voiceless plosives was also found for Arabic (Flege and Port, 1981; Kulikov, 2016). This is one of the reasons VOT is argued to be insufficient to account for the voicing distinction and has led scholars to search for an alternative reliable measure of voice contrast. Kingston and Diehl (1994) in this respect argue that vowel-onset f0 is the most reliable cue for voicing. On the other hand, English coronal stops are also different in terms of the exact location of contact compared to Arabic. English coronal stops are usually described as alveolar, whereas Arabic coronal stops are described as alveo-dental in place.[36] Studies investigating acoustic correlates of place of articulation for plosives rely on the spectral shape of the release transient and burst (e.g. Stevens, Manuel, and Matthies, 1999; Suchato, 2004a).

---

[34] It was not possible to include training items commencing in velar plosives. This was because the word lists in DigLin did not have enough tokens of singleton onset velar plosives comparable across voice categories and vowel contexts.

[35] For further details see Chapter 3.

[36] See Chapter 3 for more detail.

The release of the flow of air, which is caused by the air pressure build-up during the closure, leads to frication noise exciting through the oral cavity starting from the point of constriction and ending outside the oral tract at the lips. This is usually referred to as frontal cavity. The length and shape of the front cavity vary as a result of place of constriction vis-à-vis bilabial, alveolar or velar in English. Dorsal plosives inherently have a longer cavity than their labial cognates. The succeeding vowel contributes further to the length of the frontal cavity and the spectral shape in terms of its frontness and lip rounding. The acoustic correlate of the burst-release examined in the present study was the measure of spectral tilt of the release burst spectrum. Suchato (2004a) presents the formula underlying the spectral tilt of the burst release as follows:

```
Ahi-A23(dB) = 20log( Ahi / A23)
```

The spectral tilt Ahi-A23 is the frequency differential between two frequency ranges. Ahi is the amplitude of the highest peak in the burst spectrum between the frequency range from 3 kHz for males and 3.5 kHz for females up to 8 kHz. A23 is average peak amplitude in the burst spectrum between the frequency range from 1.25 kHz to 3 kHz (Stevens et al., 1999; Suchato, 2004a). However, these values were designed for adult speakers. For children, the frequency range for A23 was calculated from the range between the mean F2 and the mean F3 as derived from the participants' data. Acoustic studies have demonstrated that with age comes a gradual reduction in formant frequency (and variability in formant frequency) alongside F1-F2 (Peterson and Barney, 1952; Vorperian and Kent, 2007). Vorperian and Kent (2007) demonstrate the presence of gender differences in formant frequency, which unfold by the age of four and become more prominent by the age of eight. Such differences are grounded in anatomic development of the vocal tract (Vorperian and Kent, 2007). It was predicted that the spectral tilt was the largest for alveolar plosives. Dental plosives were expected to have a lower spectral tilt and bilabial stops were expected to be the smallest. Within each plosive, the spectral tilt was predicted to be higher in back vowel contexts and lower when followed by a front vowel. Additionally, because children have shorter vocal tracts, their overall spectral tilt values were expected to be lower than those of their teachers. What really matters here is the relationship between their L1 and L2 spectral tilt values. Because English /t d/ are alveolar in place, their L2 spectral tilt values would be higher (provided the comparisons were made within the same vowel context) if learning took place but the same as L1 if no learning took place.

In the pilot study results, it has been found that there was covert contrast in /p/-/b/ distinction. L2 English productions of /p/ sounded to the native speaker more like /b/ (due to overlap across the two languages in the short lag VOT, which matches the voiceless plosive in Arabic and the voiced counterpart in English) and this is widely known of English L2 speakers with an L1 Arabic background but acoustically, the learners made a significant distinction between their native /b/ and target /p/ in VOT and their closure durations. The explanation is that syllable initial English /b/ was partially devoiced and classified as short lag and /p/ as long lag whereas Arabic /b/ has negative VOT. In attempt to deviate from their L1 /b/ negative VOT to the target /p/ long lag, the learners produce an intermediate lag, i.e. short lag making it sound to the native speaker short lag /b/. Also values for place of articulation of English /t/ compared to Arabic /t/ and L2 English /t/, tongue shape and sibilance (English /t/ has a fricated release compared to Arabic non-fricated release) showed interesting results. Learners exhibited intermediate values in their L2 productions of VOT, closure duration, spectral moments and spectral tilt in all three vowel contexts for the spectral measurements.

In the current study these values were compared in four stop sounds /b p t d/ across three language groups Arabic, and L2 English and the language of training input referred to as TL. The inclusion of /k g/ was not possible due to the lack of sufficient tokens for either plosive in DigLin. The target sounds were word-initial followed by one of three vowels /iː a uː/. The selection for these three vowels was for two reasons. The first was that vowel place and quality can have an influence on the acoustic cues measured including VOT (B. Smith, 1978). In the case of spectral tilt (Ahi-A23), it correlates with length of frontal cavity, which varies with each of these vowels. Vowel context was regarded an additional effect. Another reason was that these vowels exist in the vowel inventories of both languages and thus cross-linguistic comparisons were possible. 4 sounds x 3 vowels = 12 test items from DigLin (see table 5.3).

5.8 Segmentation

Data relating to plosives were analysed in Praat (Boersma and Weeninck, 2016). The acoustic cues that were compared across the three language groups were VOT, vowel-onset f0 both of which correlating with voicing contrast and spectral tilt (Ahi-A23) correlating with place of articulation (Stevens et al., 1999; Suchato, 2005). VOT in word-initial plosives is defined as the duration between the cusp of the release of a plosive closure and the onset of glottal pulsing illustrated as quasi-periodicity (Lisker and Abramson, 1964). Abramson and Whalen (2017) provide an updated definition to account for VOT in running speech/medial position. However, in the current study, we were interested in word-initial singleton plosives in citation form.

Studies on cues of voicing contrast in a wide range of languages indicate that f0 at the onset of periodic voicing correlates with voicing. That is vowel-onset fundamental frequency tends to be lower following voiced consonants – especially obstruents and plosives – and higher following voiceless consonants (Hombert, Ohala, and Ewan, 1979; House and Fairbanks, 1953; Kingston and Diehl, 1994; Kirby and Ladd, 2015; Lehiste and Peterson, 1961) especially in the absence or suppression of other cues of voicing contrast (Repp, 1982; Silverman, 1986).

Auditory and visual inspection was conducted to detect plosive segments in the speech signal. Once the transient was spotted, the segmentation criteria for the positive VOT for plosives were that the starting point was marked from the release of the transient marked from abrupt rise/fall in the waveform (see figure 5.7 left image). The interval labelled 'vot' from the bottom tier in the TextGrid marked the boundary which was aligned with the sharp fall (in this instance) in the waveform. The end point, which was the initiation of voicing was marked from the first positive peak of periodicity as seen in the waveform coinciding with the first visible striation in the spectrogram (see figure 5.7 right image). Figure 5.7 below:

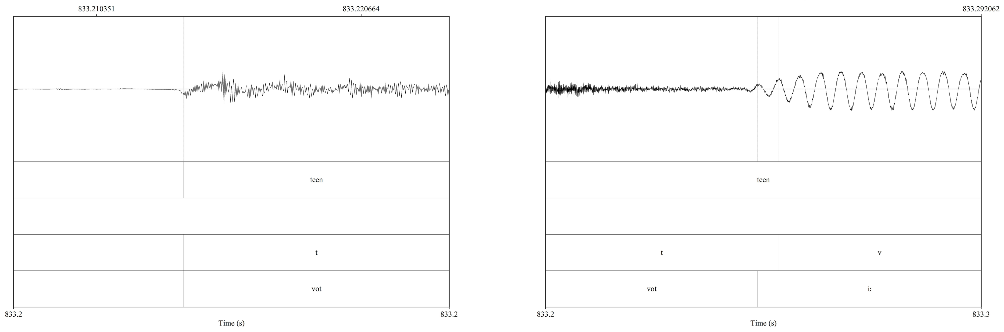


Figure 5.7 A token of word-initial voiceless plosive /t/ illustrating segmentation criteria for VOT boundaries

For L1 Arabic, VOT of voiced plosives was expected to be negative, in which case, the start was marked from the first visible striations in the spectrogram preceding the transient up to the release of the transient, which marks the end of the segment. Measurements in this case were assigned negative values. However, visual inspection revealed that some cases showed discontinued voicing during the closure duration. A similar case was observed in the voicing lead in monolingual Italians (MacKay et al., 2001). This could be due to language contact resulting from the history of Italian colonisation in Libya. In any case, to confirm the presence of gaps in voicing, a Praat script for detecting voiced and unvoiced segments developed by the first author in a study by J. Al-Tamimi and Khattab (2015) following J. Al-Tamimi (2007) was used. The script created an automated tier in the TextGrid with intervals marked as either 'V' for voiced intervals or 'UV' for unvoiced intervals based on f0 and intensity estimations to detect voicing. However, if the amplitude was low, this estimation-based procedure would fail to detect voicing. For this reason, instances which could not be determined based on this procedure were checked manually by changing the spectrogram frequency settings to 2 kHz and creating a narrow band using a 0.05 seconds window length to mimic the script settings which allowed to accurately evaluate the voicing components. Finally, activating pitch tracking and changing the analysis method to cross-correlation. The presence of low frequency activity vis-à-vis the first harmonic (H1) helped determine the presence of voicing in disputable cases. A dark H1, which was strong in amplitude indicated the presence of voicing and a faint H1 indicated unvoicing. Once this was determined, the negative VOT interval extends from this point until the release as defined above. Any potential unvoiced gaps between those two points were included in the duration.

Duration of VOT was extracted from the start to the end of this segment. Interlanguage (IL) productions were expected to be intermediate between L1 Arabic and TL English. Measurements related to the burst, that is spectral tilt were extracted from the first 10 ms of the VOT segment inspired by the segmentation procedure of Stevens et al. (1999). In the case of negative VOT, measurements were extracted from the first 10 ms post release of the VOT boundary.

As for vowel-onset f0, measurement was extracted from a point in the speech signal rather than an interval. The onset of the vowel was marked by a rise in amplitude following the preceding consonant and the emergence of formant structure following (Khattab and J. Al-Tamimi, 2014) which was facilitated using a wide-band spectrogram window.

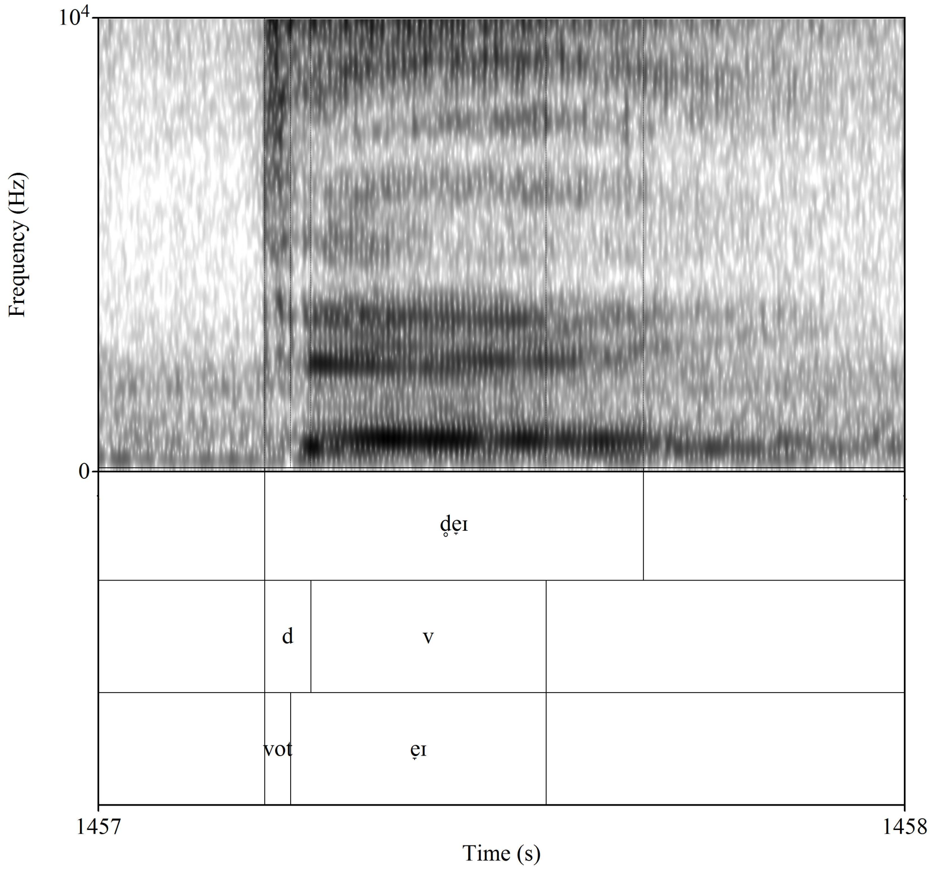


Figure 5.8 A token of the test item /deɪ/ illustrating segmentation criteria for the point of vowel onset

Measurements for the test sounds were extracted using a Praat script (Boersma and Weeninck, 2016) developed during the pilot study with the assistance of my supervisor, Jalal Al-Tamimi. The script was further modified to include spectral measurement ranges suitable for female speakers (for the DigLin female speaker and the Traditional teacher), child speakers as well as the command for vowel-onset f0 measurement extraction. A Kaiser-2 window was used to extract power spectra for the first 10 ms of the transient for Ahi-A23 using a fast power spectrum.

After data were extracted using Praat script, data were manually checked. For example, labels of Target-like, Non-Target-like were checked for consistency within words and across speakers.

5.9 Statistical analysis

5.9.1 *General modelling*

Two types of tests were used to carry out the statistical analyses, linear mixed effect models (LMM) and generalised linear mixed effect models (GLMM). These were applied using the `lmer` package (Bates, Mächler, Bolker, and Walker, 2015) in `R Statistical Software` (R Core Team, 2013).

The LMMs were suitable for acoustic data since the outcome values were continuous, whereas the GLMMs were more suited to categorical data. Linear mixed effect analyses were carried out with the outcome being each acoustic correlate of *voice* and *place*. The voice correlates measured were VOT durations, vowel-onset *f0*, and the *place* correlate measured was spectral tilt (Ahi-A23).

Generalised linear mixed effect analyses were carried out with the outcome being the binary categorical outcome target-likeness rating with two levels (target, non-target) using the function `family = binomial`. The same procedure was used for match rating outcome. Another GLMM analyses was conducted with the outcome being the count of learned words using the function `family = poisson`.

Following the recommendations of Winter (2019) for model specification practice in linguistics, the model specification was driven by theory and based on effects of interest, that is:

- Relating to the research questions, vis-à-vis Instruction, time of Test.

- Controlling for speaker-related effects: Gender, Age, Literacy (in Arabic).
- Controlling for effects relating to the measured outcome: for example, Vowel context, Place, Voice.

Model outcomes were reported individually for each outcome variable in subsequent chapters.

Additionally, due to the nature of the data, the acoustic analyses were divided into two main parts. The first part of the analysis (L1-L2-TL) encompasses the comparison of the participants' L2 to their L1 outputs and TL input. This was to discern any significant differences between their interlanguage (L2 English) compared to their native language (L1 Arabic) and the target input (TL) as a prerequisite to establish whether any learning may have taken place. In the first part of the analysis, each condition was examined individually. This was for several reasons. One reason was that training conditions varied in terms of their respective TL input. The experimental conditions received input from DigLin and the Traditional condition received input from Traditional teacher. Another reason for measuring results for conditions individually was the confounding factor of teacher vs learner on the one hand and the TL belonging to the teacher only and the L1, L2 belonging to the learners only risking the possibility of correlation amongst effects. For this reason, the effect of learner vs teacher was replaced with the effect of language group (Language). The second part of the analysis (L1-L2) entails comparisons between the learner's L2 outputs with their L1 outputs. This analysis explores differences in all training conditions in both tests. The GLMMs however, do not require comparisons with TL as that was already carried out during the labelling process. For example, the Traditional condition's tokens were labelled 'matching' or 'non-matching' based on their teacher's data. The same was true for the experimental conditions.

Fixed effects

Different effects were used for the various models depending on the acoustic measure and the type of outcome.

LMMs

For the first part of the analysis, that is L1-L2-TL (within-condition), effects such as moment of test and task were not comparable for the TL language group as teachers did not attend a post-test and delayed post-test or have their data elicited by the same tasks. For this reason, the data were subset to allow for each instruction type to be compared to its respective input individually. The data were further subset by moment of test for the same reason. There is a potential difference in performance depending on task, but mainly between the memory recall tasks (picture-naming and read aloud) and the delayed repetition task (because the latter – unlike the former two – may potentially involve reliance on phonological memory). That is why it is ideal to include this effect in the models. However, it could not be dealt with in the same way as moment of test and instruction type because tokens in the picture-naming and read-aloud tasks were small in number and did not cover all items by all speakers, which may increase the risk of convergence warnings. The models will have issues and the coefficients obtained may not reflect the data accurately. Because elicitation task was not something the research questions or hypotheses address directly, tokens are thus collated by dropping this effect from the L1-L2-TL models.

To allow for meaningful interpretation of the coefficients, contrast coding was used on all fixed effects (Schielzeth, 2010) using the following code:

```
contrasts(data$Gender) = contr.sum(2)
contrasts(data$Literacy) = contr.sum(2)
contrasts(data$Vowel) = contr.sum(4)
contrast(data$Language) = contr.sum(3)
```

The continuous predictor ‘Age’ was centred using the following code:

```
data$Age <- scale(data$Age, center = TRUE, scale=FALSE)
```

This step was to explore whether learners established new L2 phonetic categories that were distinguishable from their L1 prototypes or not following from the predictions made in the models of speech learning in Chapter 2. For the second of the analysis, by-condition comparisons, the fixed effects were instruction type with three levels (Listen and Speak, Listen Only, and Traditional teaching), language group with two levels (L1 Arabic, L2 English), task with three levels (picture-naming, read aloud, and delayed repetition), gender with two levels (male and female), literacy with two levels (literate and illiterate), and age. The fixed effect *moment of test* was added to see if learning was maintained or lost after ten weeks from the end of the training.

The fixed effect vowel with four levels (front-high, front-low, back-low and back-high) was added to both VOT and spectral analyses as VOT durations and spectral measurements vary by vowel context (Klatt, 1975; J. Ohala, 1981). However, it was not included in the analysis of vowel-onset *f0* due to its irrelevance.

The continuous predictor 'Age' was centred using the following code:

```
data$Age <- scale(data$Age, center = TRUE, scale=FALSE)
```

Sum-coding was used on all fixed effects using the following code:

```
contrasts(data$Gender) = contr.sum(2)
contrasts(data$Literacy) = contr.sum(2)
contrasts(data$Vowel) = contr.sum(4)
contrasts(data$Language) = contr.sum(3)
contrasts(data$Instruction) = contr.sum(3)
contrasts(data$Task) = contr.sum(3)
contrasts(data$Test) = contr.sum(2)
```

GLMMs

The GLMM analyses only included data from the L2 outputs as the aim was to compare results (target-likeness, match rating and lexical learning) across instruction types. The fixed effects were instruction type with three levels (Listen and Speak, Listen Only, and traditional teaching), task with three levels (picture-naming, read aloud, and delayed repetition), gender with two levels (male and female), literacy with two levels (literate and illiterate), and age. The fixed effect *moment of test* was again added to see if target-likeness ratings and lexical learning were maintained or lost after ten weeks from the end of the training.

Contrast coding was used on all fixed effects once again using the following code:

```
    contrasts(data$Gender) = contr.sum(2)
contrasts(data$Literacy) = contr.sum(2)
contrasts(data$Vowel) = contr.sum(4)
contrasts(data$Instruction) = contr.sum(3)
contrasts(data$Task) = contr.sum(3)
```

```
contrasts(data$Test) = contr.sum(2)
```

Age was centred by running the above code again.

Random effects

All the models utilised crossed random effects both for speaker and for item. By-speaker and by-item random intercepts were included to account for variability in relevant acoustic variables, target-likeness rating and amount of lexical learning in terms of speakers and test items. The specification of their random slopes, however, varied by model.

In the VOT and spectral measurements models an attempt was made to include by-speaker random slopes for Vowel, Place, Test, Voice, Task, Language, and Age in order to take into account the variability in speakers' VOT productions relative to each. However, these slopes were reduced gradually if the model failed to converge. An attempt was also made to keep models across training conditions as similar as possible in order to achieve comparable results. For the VOT, spectral measurements, and vowel-onset *f0* models pertaining to the comparison between the learners' L1 and L2 outputs by instruction and test, also attempted to include by-item random slopes for Language, Gender, Literacy, Age in order to account for the variability in test items relative to such effects. The test items for L1 Arabic were inherently different from those for L2 English.

The main objective from this step of the analysis was to explore any significant differences between each condition's L2 and L1 outputs.

Finally, Estimated marginal means (EMMs) (or Least-squares means) were applied using the package `emmeans` (Lenth, 2019) in `R` to compare EMMs of the predictors with one another. This was the most common post hoc analysis for models with factors as predictors. Not only does the emmeans function allow us to explore whether differences between levels of predictors were significant, it also allows us to make model-driven predictions which work similar to the `predict` function in `lme4`. The Satterthwaite method for calculation degrees of freedom in LMMs was applied to correct for multiple comparisons using the following code as an example:

```
emmeans(mdl, pairwise ~Instruction * Test * Language,
lmer.df = "satterthwaite")
```

For GLMMs, the function `type = "response"` was added to calculate based on log-odds using the following code:

```
emmeans(mdl, pairwise ~Instruction * Test * Language,
type = "response")
```

The graphs were generated using the `ggplot2` package (Hadley, 2016). Using the `ggpredict` function in the package `ggeffects` (Lüdecke, 2018) was also used to plot predictions visually.

Finally, to test whether the impact of a given effect was significant, the `afex` package (Singmann, Bolker, Westfall, Aust, and Ben-Shachar, 2019) in `R` was used to conduct a series of automated Likelihood ratio tests using the function `'method = LRT'` and by wrapping the function `mixed()` around the model. The object created by the `mixed()` function contains the full model as well as all 'nested' models that were used for likelihood ratio tests (see Winter, 2019: 263).

Chapter 6: Target-likeness and Match Ratings

6.1 Introduction

The aim of this chapter is twofold. The first is to to establish to what extent the participants learned the words they were expected to learn. However, in establishing the data for the remainder of the analyses in the current chapter for Target-likness rating and match rating and acoustic measurements in Chapters 7 and 8, and finally in Chapter 9 for the qualitative analysis of phonological processes, we considered including the words they fully and partially learned. This yields considerable data to provide insight into the variables we ask to measure in the present study. The second aim is to present the results of the target-likeness rating outcome as a function of Instruction type – Listen and Speak, Listen Only, and Traditional teaching. The impact of two conditions of computer–assisted pronunciation training and Traditional teaching on children is measured.

6.2 Lexical learning

For lexical learning, we seek to establish how many and which words the learners in the three conditions showed evidence of learning. We had asked earlier in Chapter 5:

RQ3: Which training method will result in the most *fully* learned words?

It was shown earlier in Chapter 5, tables 5.4 and 5.5, that the Traditional Teaching learners yielded the highest number of analysable tokens in the picture-naming task and that the number of tokens were three to four times as many as that found in either of the CAPT training conditions. Nevertheless, it was also stated in Chapter 5 that only the tokens which were produced in full and fully matched those of the input were considered in the analysis of lexical learning. This is because they represent full phonological learning of the test items. Everything else was considered partial learning and was therefore discarded from this analysis. Lexical learning will be compared between the post-test and the delayed test. If the participants recall the test words and are able to produce them accurately after 10 weeks of the end of training/teaching, as examined in the delayed test, this means that learning took place. Additionally, the most frequent recalled words will be presented and an attempt will be made to explain possible factors that influenced the participants' preference for them.

In the following sections, we present descriptive statistics on amount of lexical learning by time of Test and training/teaching condition. In doing so, we also explore the variations in amount of lexical learning as a function of elicitation Task. We also present the most recalled words and provide a discussion of possible explanations.

Amount of learned words

Table 6.1 demonstrates the mean of *fully* learned words per Instruction condition, time of Test and Task type. By examining the figures in the table, the answer to the research question 'Which training method will result in the most *fully* learned words?' is that overall, those following the Listen and Speak training condition learned the most words, followed by those following the Listen Only training condition, and finally, those following the Traditional Teaching condition learned the least amount of words by comparison.

Three overall themes can be deduced from table 6.1. First, the collective mean recall rate from the three tasks for each group indicates that there are considerable differences in the global rate of learned words across the three conditions.

Second, there is a considerable difference in the mean of learned words between the materials of the delayed repetition task on the one hand and those of the memory recall tasks on the other. Third, it seems that the effect of the test seems to interact with both task and instruction type. Whilst the mean rate of learned words exhibited in the delayed repetition task increases in the delayed test, that exhibited in the picture-naming decreases in the delayed test. Mean rate of learned words in the read aloud in the delayed test compared to the post-test however, seems to fluctuate depending on Instruction. It increased slightly for the Listen and Speak condition. It remained the same for Listen Only. Finally, it decreased for the Traditional condition.

Moreover, amongst the three instruction types, Traditional instruction has the

Table 6.1 Mean number of learned words grouped by Instruction, Test, and Task

Instruction	Test	Task	Mean
Listen and Speak	Post-test	delayed repetition	53.60
		picture-naming	3.85
		read aloud	2.00
	Delayed test	delayed repetition	54.50
		picture-naming	2.00
		read aloud	5.00
Listen Only	Post-test	delayed repetition	44.30
		picture-naming	1.83
		read aloud	1.00
	Delayed test	delayed repetition	48.10
		picture-naming	1.62
		read aloud	1.00
Traditional	Post-test	delayed repetition	24.30
		picture-naming	8.35
		read aloud	5.00
	Delayed test	delayed repetition	24.90
		picture-naming	7.13
		read aloud	1.00

highest recall rate in the picture-naming task but not in the read aloud and delayed repetition tasks. Listen Only generally demonstrated the poorest performance in the memory recall tasks. Amongst the two memory recall tasks, picture-naming task yielded consistently a marginally higher rate of words than the read aloud task. Thus the read aloud seems to be the most challenging for this age group of L2 learners.

Top most and least fully learned words

Table 6.2 *Fully* learned words produced 40+ times in the delayed test

#	item	total	Post-test	Delayed test
1	dog	149	80	69
2	bee	105	59	46
3	dish	103	57	46
4	bat	102	57	45
5	left	101	51	50
6	jam	99	53	46
7	milk	95	49	46
8	chin	94	53	41
9	chip	93	51	42
10	list	93	50	43
11	pad	93	53	40
12	pea	93	50	43

Table 6.2 shows the top most *fully* produced words and their frequencies

according to time of Test.[37] Research findings by Hansen (2017) indicate an advantage for words with certain psychological factors as well as linguistic factors that account for vocabulary development and learning. Two factors proposed by Hansen are frequency and imageability. All items in the experimental groups have had the same frequency.

The words in the Traditional group have also had similar frequencies across the test items. Therefore, this factor does not seem to apply to the current study. Similarly for imageability, all the test words are accompanied by images that reflect their meaning. Nonetheless, some pictures are more intuitive of the meaning than others (c.f. figure 6.1).

Gentner (1982) suggests an additional factor, which is word type. Nouns, he argues, are generally learned before verbs. Looking at the top 12 most *fully* learned words, most of the items are nouns with the exception of the verb 'left'. However, the 12 least *fully* learned words are also mainly, though not exclusively, nouns.

Thin Broth

Dish Tool

Figure 6.1 Some examples of test words and their associated images

Gentner (1982) also suggests that etymology, that is the historical origin of the word or morpheme may condition application or otherwise of the phonological or phonetic processes involved in learning. The commonality amongst these top learned words in comparison to the least learned words in table 6.3 can be explained on a linguistic basis.

---

[37] For a full list of the most *fully* learned words, see appendix C.

The sounds and structures in the top most learned words are relatively less marked and less phonologically-complex than those in the least learned words (e.g. /-nt/, /-lt/, /-mp/, /r/). In some cases, the words from the least *fully* learned words contain more than one problematic sound for Libyan learners. Examples of this is the test item ‘broth’, which not only requires the learning of a complex onset /br-/, but also the rhotic approximant, which is acquired late in British English speaking children (Mcleod and Crowe, 2018) and the dental fricative. The item ‘quilt’ similarly has two problematic areas, a complex onset and a complex coda. The item ‘rain’ also includes a rhotic approximant and a diphthong. This raises the possibility of being deemed partially learned and thus excluded more often than test items with less challenging problematic sounds. Further analysis is beyond the scope of the present thesis.

Table 6.3 *Fully* learned words produced less than 20 times in the delayed test

#	item	total	Post-test	Delayed test
1	paint	4	3	1
2	mount	11	9	2
3	wrap	17	9	8
4	point	24	12	12
5	quilt	24	11	13
6	rain	26	18	8
7	roll	27	14	13
8	pool	35	16	19
9	rose	36	18	18
10	jump	39	25	14
11	broth	40	20	20
12	march	41	23	18

We now turn to the second aim of this chapter in the following sections. That is to provide the answer to the first research question posed in Chapter 5:

RQ1: Which training method will result in more target-like (and less L1-like) pronunciations of problematic sounds Libyan learners typically display learning English? Namely: Affricates, Clusters, Dental fricatives, Diphthongs, Plosives, and Rhotic approximants.

The following section reports on the results of statistical analysis of target-likeness rating outcome as a function of time of test, training condition, elicitation task, the class of problematic sound, literacy level, gender, and age. However, due to the differences in the input received by training conditions versus the Traditional teaching condition, the analyses were two-fold. In the first part of the analyses, the aim of the impressionistic transcription of tokens was to compare L2 realisations *as a whole* to those of native speakers. Based on the transcriptions, each token was assigned a label of either ”target-like” labelled (T) if it matched that of a native speaker, or ”non-target-like” (NT) if it did not.

In the second part of the analyses, the aim of the labelling of tokens was to compare L2 realisations of problematic sounds in each training condition to the respective input from training. Thus, experimental groups' tokens were compared to those of DigLin and Traditional condition's tokens to their teacher. In this analysis, each token was assigned a label of either ”match” labelled (M) if it matched that of the training input, or ”non-match” (NM) if it did not.

6.3 Target-likeness

6.3.1 *Descriptive statistics*

Figure 6.2 and table 6.4 below show the percentages of average target rating scores by problematic sound class and by task for each instruction type in both tests. The figure shows that target rating scores varied between instruction types in terms of averages and rank. Cross-problematic sound class averages varied not only by instruction type but also by test and task. For the delayed repetition task, Listen and Speak exhibited the highest average target rating scores during the post-test, followed by Listen Only. Traditional instruction had the lowest average target rating scores. In the delayed test however, the Traditional instruction target rating scores for coda clusters (19%) and dental fricatives (22%) exceeded those for the Listen Only instruction (15%, 17% respectively). Additionally, the rank of difficulty for the Listen Only and Traditional instruction changed. For the Traditional instruction, in the post-test, the highest score was evident in affricates followed by plosives, clusters and dental fricatives (same score), diphthongs and finally rhotic approximants had the lowest score. In the delayed test, this changed for clusters. Similarly, the ranking of clusters and dental fricatives for the Listen Only instruction varied by test.

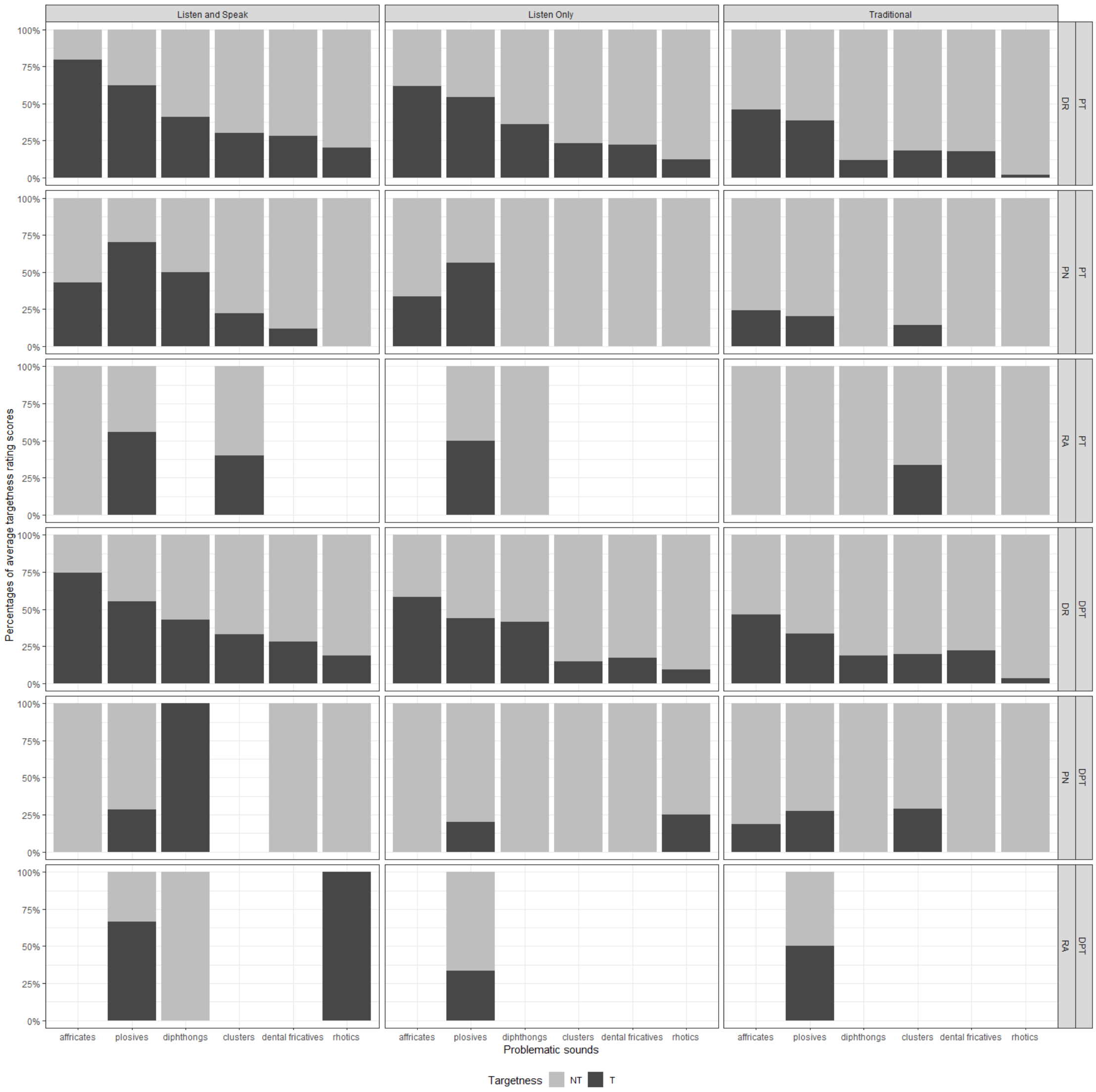


Figure 6.2 Proportions of target and non-target realisations grouped by problematic sound class and task across the three instruction types in the post-test and delayed test. PT stands for post-test. DPT stands for Delayed test. DR stands for delayed repetition. PN stands for picture-naming. RA stands for read aloud

For the picture-naming task, scores varied considerably. The highest average target rating score for Listen and Speak during the post-test was plosives (70%). In the delayed test it was diphthongs (100%). As for Listen Only instruction, the highest score was also evident in the class of plosives (56%) during the post-test. In the delayed test it was rhotic approximants (25%). The highest average target rating score for Traditional instruction during the post-test was affricates (24%). In the delayed test it was clusters (29%).

Materials from the read aloud task were scarce and figures did not represent scores very accurately as in the other elicitation materials. The Listen and Speak instruction yielded tokens containing affricates, plosives and clusters in the post-test and plosives, diphthongs, and rhotic approximants in the delayed test. Listen Only instruction yielded tokens containing plosives and diphthongs in the post-test and only plosives in the delayed test. Traditional instruction on the other hand yielded tokens containing all of the sound classes in the post-test and only plosives in the delayed test.

For Listen and Speak instruction type, plosives received the highest average target rating score in the read aloud task during the post-test (56%). In the delayed test, the class of rhotic approximants received the highest average target rating score (100%) within that group. Listen Only instruction exhibited a score of 50% for plosives in the post-test, whereas in the delayed test, this score declined to 33%. Traditional instruction exhibited the score of 33% for plosives in the post-test, which increased during the delayed test to 50%. It is clear from the above the interaction of the effects of instruction, test, sound class and task.

Table 6.4 Average target rating score per test, instruction, problematic sound class and task

	Post-test			Delayed test		
	Listen and Speak			Listen and Speak		
	delayed repetition (43%)	picture-naming (36%)	read aloud (47%)	delayed repetition (41%)	picture-naming (20%)	read aloud (60%)
Affricates	80%	43%	0%	75%	0%	
Plosives	62%	70%	56%	55%	29%	67%
Diphthongs	41%	50%		43%	100%	0%
Clusters	30%	22%	40%	33%		
Dental fricatives	28%	12%		28%	0%	
Rohtics	20%	0%		19%	0%	100%
	Listen Only (34%)			Listen Only (29%)		
	delayed repetition (34%)	picture-naming (26%)	read aloud (25%)	delayed repetition (30%)	picture-naming (11%)	read aloud (33.3%)
Affricates	62%	33%		58%	0%	
Plosives	54%	56%	50%	44%	20%	33%
Diphthongs	36%	0%	0%	42%	0%	
Clusters	23%	0%		15%	0%	
Dental fricatives	22%	0%		17%	0%	
Rohtics	12%	0%		9%	25%	
	Traditional (21%)			Traditional (23%)		
	delayed repetition (22%)	picture-naming (11%)	read aloud (10%)	delayed repetition (23%)	picture-naming (15%)	read aloud (50%)
Affricates	46%	24%	0%	46%	19%	
Plosives	39%	21%	0%	34%	27%	50%
Diphthongs	12%	0%	0%	19%	0%	
Clusters	18%	14%	33%	19%	29%	
Dental fricatives	18%	0%	0%	22%	0%	
Rohtics	2%	0%	0%	3%	0%	

Moving from the effects of problematic sound class and task, figure 6.3 below illustrates the effect of gender and literacy on average target rating scores. As shown, there was a clear interaction between gender and literacy that was consistent regardless of the instruction type. In the group of literates, males performed better than their female counterparts. In the group of illiterates, females outperformed their male counterparts. For the traditional instruction, comparisons between literate and illiterate females was not possible. Also comparisons between illiterate males and illiterate females was not possible as the group did not have illiterate females.

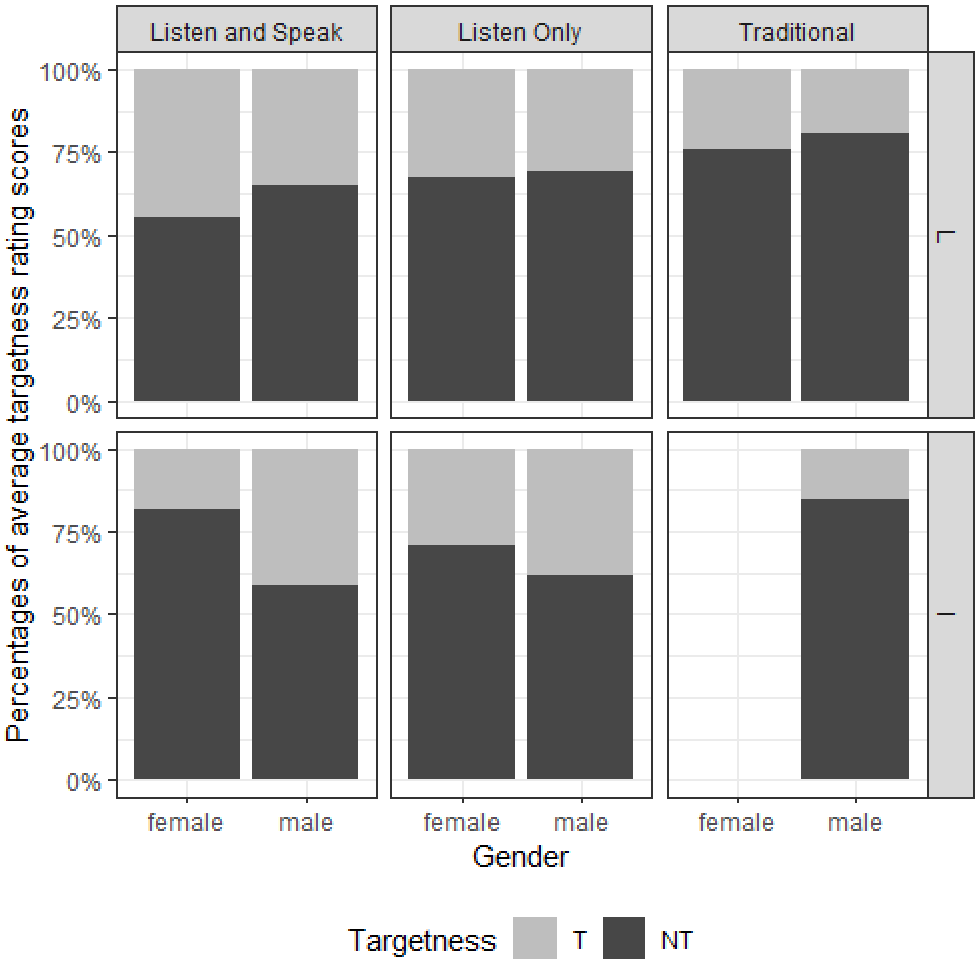


Figure 6.3 Proportions of target- and non-target realisations grouped by instruction type, gender and literacy. L stands for literate and I stands for illiterate

6.3.2 *Model specification for target-likeness rating*

This part of the analysis attempts to answer the first research question:

RQ1: Which training method will result in more target-like (and less L1 like) pronunciations of problematic sounds Libyan learners typically display learning English, including Affricates, CC Coda clusters, Diphthongs, Plosives, Voiceless dental fricatives, Rhotic approximants. It was hypothesised that the Listen and Speak condition will result in the most target-like productions/least L1 like productions (due to having access to the output-input loop feedback and native speaker input) followed by the Listen Only (due to having access to native speaker input), leaving the Traditional Teaching condition with the relatively least target-like productions/ most L1-like productions (due to being exposed to foreign-accented input). Thus, we were mainly interested in measuring the effect of instruction type between the three instruction types – Listen and Speak, Listen Only, and Traditional – and class of problematic sound on target-likeness rating scores. Following the recommendations of Winter (2019) for model specification practice in linguistics, and based on the observations in section 6.3.1, we also wanted to account for potential speaker-related effects of Literacy, age, and Gender on target-likeness rating scores. Finally, we want to account for were interested in the effects of task and 'time' on target-likeness rating scores to see whether pronunciation changed during the delayed test and if so, how they change.

We have also seen in section 6.3.1 the interaction between the class of problematic sound with time of test on the one hand and with instruction type on the other. To this end, the following generalised linear mixed-effects model was built. The model adds by-speaker and by-item random intercepts.

```
Target ~ Instruction * Test * Sound + Task + Literacy *
    Gender + Age + (1 | speaker) + (1 | item)
```

6.3.3 *Inferential statistics on target-likeness rating scores*

Table 6.5 shows the results of a generalised linear mixed-effects model for target-likeness rating scores in all instruction conditions using a binomial distribution. The reference for Instruction type was 'Listen and Speak', for Test 'Post-test', for Sound 'Affricate', Task 'Delayed repetition', for Literacy 'Literate', and for Gender it was 'Female'. The effect of Age was centred. All fixed effects were sum-coded.

The table includes coefficients for the simple effects of Instruction, Test and Sound given that they were part of an interaction. It also includes coefficients for the simple effects of Gender and Literacy. It also displays two-by-two and three-by-three interaction terms for the above mentioned. For this reason, differences between these are discussed later in the post hoc tests.

The GLMM output in table 6.5, shows that there was statistically significant increase (p <.0001) in the odds of target-like rating for Picture-naming materials compared to those of the delayed repetition task. It also shows that there was a statistically significant decrease (p <.0001) in the odds of target-like rating for Read aloud materials compared to that of the delayed repetition task.

Finally, the table shows that the odds of a target-like rating increase as a function of Age although it does not reach statistical significance (p >0.2).

Table 6.5 Output of a generalised linear mixed-effects model for target-likeness rating score

term	β	S.E	z	(p)
(Intercept)	-1.86	0.22	-8.34	**0.0000**
InstructionListen Only	0.64	0.09	6.73	**0.0000**
InstructionTraditional	0.12	0.09	1.3	0.1933
TestDelayed	0.07	0.04	1.63	0.1025
SoundCluster	1.7	0.32	5.38	**0.0000**
SoundDFricative	-0.46	0.3	-1.52	0.1286
SoundDiphthong	-0.44	0.33	-1.33	0.1852
SoundPlosive	-0.11	0.33	-0.32	0.7497
SoundRhotic	0.88	0.31	2.87	**0.0042**
TaskPicture-naming	0.570	0.15	3.79	**0.0002**
TaskRead aloud	-0.580	0.17	-3.47	**0.0005**
LiteracyLiterate	-0.11	0.1	-1.15	0.2497
GenderM	-0.06	0.09	-0.68	0.4984
Age	0.02	0.02	1.19	0.2323
InstructionListenOnly:TestDelayed	0.03	0.05	0.65	0.5144
InstructionTraditional:TestDelayed	0.12	0.05	2.58	**0.0098**
InstructionListenOnly:SoundCluster	0.09	0.09	0.96	0.3367
InstructionTraditional:SoundCluster	-0.14	0.09	-1.59	0.1115
InstructionListenOnly:SoundDFricative	-0.07	0.1	-0.77	0.4420
InstructionTraditional:SoundDFricative	-0.42	0.1	-4.24	**0.0000**
InstructionListenOnly:SoundDiphthong	-0.31	0.1	-3.09	**0.0020**
InstructionTraditional:SoundDiphthong	-0.22	0.1	-2.18	**0.0290**
InstructionListenOnly:SoundPlosive	0.13	0.11	1.26	0.2065
InstructionTraditional:SoundPlosive	0.52	0.1	4.99	**0.0000**
InstructionListenOnly:SoundRhotic	-0.17	0.09	-1.95	0.0517
InstructionTraditional:SoundRhotic	0.03	0.09	0.35	0.7281
TestDelayed:SoundCluster	0.05	0.06	0.73	0.4677
TestDelayed:SoundDFricative	0.04	0.07	0.53	0.5939
TestDelayed:SoundDiphthong	-0.02	0.07	-0.22	0.8275
TestDelayed:SoundPlosive	-0.19	0.08	-2.46	**0.0138**
TestDelayed:SoundRhotic	0.16	0.06	2.57	**0.0103**
LiteracyLiterate:GenderMale	-0.21	0.09	-2.34	**0.0190**
InstructionListenOnly:TestDelayed:SoundCluster	0.05	0.09	0.54	0.5865
InstructionTraditional:TestDelayed:SoundCluster	-0.11	0.09	-1.31	0.1918
InstructionListenOnly:TestDelayed:SoundDFricative	-0.16	0.09	-1.72	0.0860
InstructionTraditional:TestDelayed:SoundDFricative	0.19	0.1	1.89	0.0588
InstructionListenOnly:TestDelayed:SoundDiphthong	-0.02	0.1	-0.17	0.8685
InstructionTraditional:TestDelayed:SoundDiphthong	0.05	0.1	0.45	0.6512
InstructionListenOnly:TestDelayed:SoundPlosive	0.11	0.1	1.07	0.2852
InstructionTraditional:TestDelayed:SoundPlosive	-0.11	0.1	-1.04	0.3002
InstructionListenOnly:TestDelayed:SoundRhotic	0	0.09	-0.05	0.9580
InstructionTraditional:TestDelayed:SoundRhotic	-0.05	0.09	-0.580	0.5599

To find out model-predicted probabilities of target-likeness rating for each instruction type, in each test by sound class, the function `emmeans` (Lenth, 2019) in `R` was applied using `type = 'response'`. Predicted probabilities are also called fitted values because they result from fitting the linear regression model to a data set (Winter, 2019). If they are very close in value to the raw values, it means that they reflect the raw data and that the model fit is good. The predicted probabilities are used in pairwise comparisons using the `pairwise` function in `R`. The following code was used:

```
emmeans(mdl, pairwise ~ Instruction * Test | Sound,
                        type = 'response')
```

Intervals for predicted probabilities were back-transformed from the logit scale. The pairwise function allows for pairwise comparisons specified by the formula, that is within-Test pairwise comparisons by Instruction for each Sound class. The Tukey adjustment method for comparing a family of 6 estimates was used. Tests were performed on the log odds ratio scale.

Figure 6.4 shows model-predicted probabilities of target-likeness grouped by sound class, training condition, and time of test. The red line depicts predicted percentage rating during the post-test and the blue line predicts that in the delayed post-test. Affricates were the least challenging sound class, followed by plosives, and then diphthongs. The most challenging sound class was rhotic approximants especially for the Traditional Teaching condition. The figure also shows that overall, the Listen and Speak condition outperformed the other two training conditions across almost all the sound classes in both times of testing (an exception being in the class of diphthongs in the delayed post-test, where the two CAPT conditions were matched).

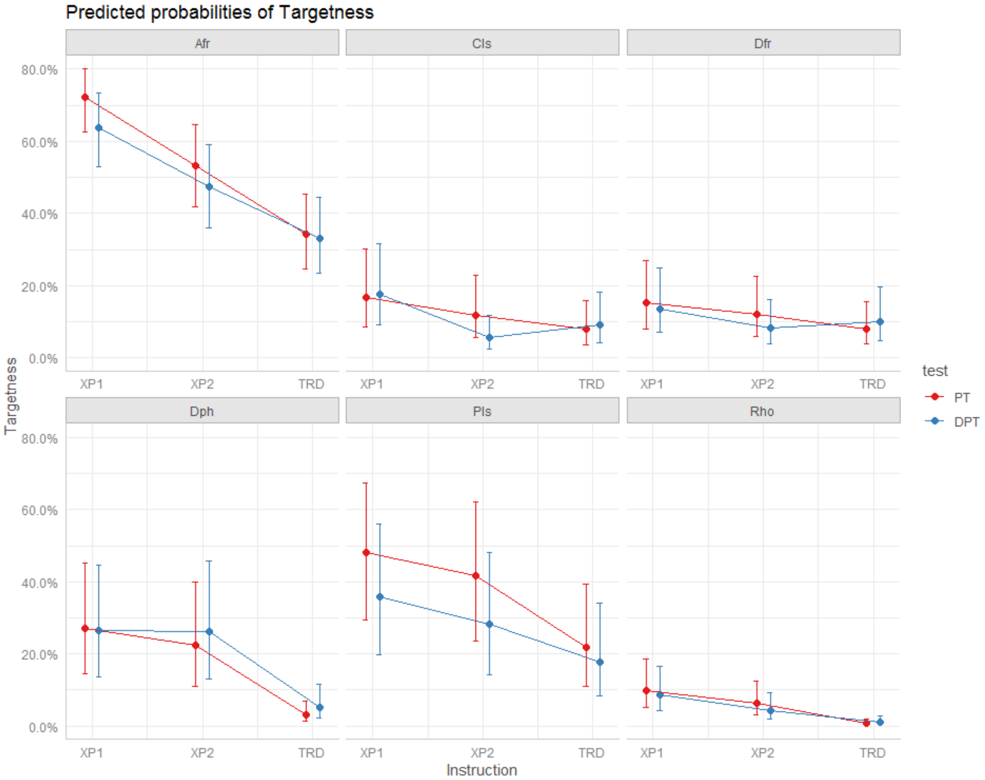


Figure 6.4 Predicted probabilities of Target-likeness by sound class, training condition, and time of test. Afr = affricates, Cls = clusters, Dfr = dental fricatives, Dph = diphthongs, Pls = plosives, Rho = rhotic approximants. XP1 = Listen and Speal, XP2 = Listen Only, TRD = Traditional. PT = post-test, DPT = delayed test

The post hoc within-test pairwise comparisons by training condition show that in the post-test, although the Listen and Speak condition outperformed its Listen Only counterpart, this difference was only statistically significant for the group of affricates (p <.03). The Listen and Speak condition also outperformed its Traditional counterpart and this difference was statistically significant for the group of affricates (p <.0001), coda clusters (p = 0.0157), diphthongs (p <.0001), plosives (p <.0001), and rhotic approximants (p <.0001), but not for dental fricatives. Finally, the Listen Only condition outperformed its Traditional counterpart during the post-test and the difference in the predicted probability of target-like rating between them was statistically significant for the group of affricates (p = 0.015), diphthongs (p <.0001), plosives (p = 0.003), and rhotic approximants (p <.001), but not for coda clusters, or dental fricatives.

Within the delayed post-test, the Listen and Speak condition once again outperformed its Listen Only counterpart – an exception was the class of diphthongs where their predicted probabilities of target-likeness rating were matched. However, this difference was statistically significant for the group of coda clusters only (p <.001). The Listen and Speak condition also outperformed its Traditional counterpart and the difference was statistically significant for the group of affricates (p = 0.0001), diphthongs (p <.0001), plosives (p = 0.0083), and rhotic approximants (p = 0.0001), but not for coda clusters, or dental fricatives. Finally, the Listen Only condition was outperformed by the Traditional condition in the delayed post-test within the groups of coda clusters and dental fricatives. However, this decline in performance did not illustrate a significant difference compared to the Traditional condition. The difference in the predicted probability of target-likeness rating was statistically significant between those latter conditions for the group of diphthongs only (p <.0001), where the Listen Only condition outperformed its Traditional counterpart. For detailed predicted probabilities and post hoc within test pairwise comparisons for target-likeness rating across sound classes, see Appendix D.

6.4 Match rating

In this part of the analysis, special attention was focused on problematic sound class in each token (unlike in target-likeness rating where the token was compared to target word as a whole. The experimental conditions' tokens were compared with those of DigLin and the Traditional condition's tokens were compared with their teacher's realisations and each were rated accordingly.

6.4.1 *Descriptive statistics*

Figure 6.5 below shows proportions of average match/non-match rating scores by instruction type in both tests. The figure shows that match rating scores varied between instruction types and to a lesser extent by time of test. Listen and Speak exhibited the highest average match rating scores during both tests, followed by Listen Only leaving the Traditional instruction with the lowest average match rating scores.

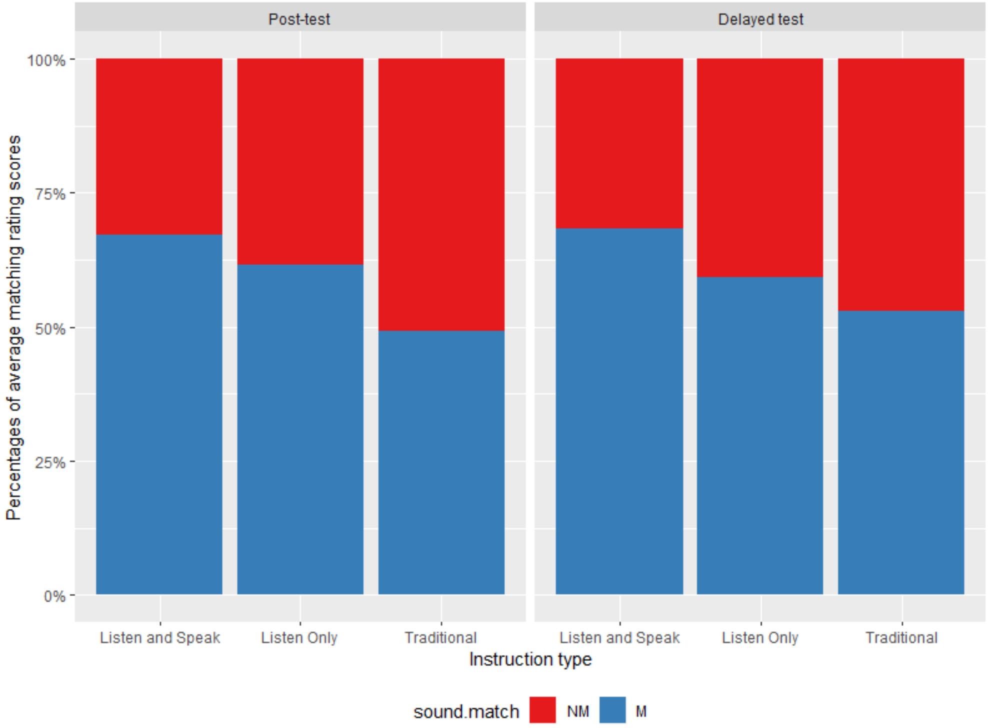


Figure 6.5 Proportions of matching and non-matching realisations grouped by instruction type in the post-test and delayed test

6.4.2 *Model specification for match rating*

The model specification for match rating follows the same as that for target-likeness to show parallel and comparable results. We were mainly interested in measuring the effect of instruction type – Listen and Speak, Listen Only, and Traditional – elicitation task, class of problematic sound on sound match rating. We also wanted to account for potential speaker-related effects of Literacy, age, and Gender on match rating. To see whether pronunciation changed with time and if so, how they change, Test was incorporated in the model.

To account for an interaction between problematic sound class with time of test on the one hand and with instruction type on the other, interaction terms for these effects were added. To this end, the following generalised linear mixed-effects model was built. The model adds by-speaker and by-item random intercepts.

```
sound.match ~ Instruction * Test * Sound + Task +
    Literacy + Gender + Age + (1 | speaker) + (1 | item)
```

6.4.3 *Inferential statistics on match rating*

Table 6.6 shows the results of a generalised linear mixed-effects model for match ratings in all instruction conditions using a binomial distribution. The reference for Instruction type was ‘Listen and Speak’, for Test ‘Post-test’, for Sound ‘Affricate’, Task ‘Delayed repetition’, for Literacy ‘Literate’, and for Gender it was ‘Female’. The effect of Age was centred. All fixed effects were sum-coded.

The table includes coefficients for the simple effects of Instruction, Test and Sound given that they were part of an interaction. It also includes coefficients for the simple effects of Gender and Literacy. It also displays two-by-two and three-by-three interaction terms for the above mentioned. For this reason, differences between these are discussed later in the post hoc tests.

The GLMM output in table 6.6 shows that there was an increase in the odds of match rating for Picture-naming materials compared to those of the delayed repetition task, but it does not reach statistical significance (p >0.05). It also shows that there was a decrease in the odds of match rating for Read aloud materials compared to that of the delayed repetition task, but it was not statistically significant either (p >0.4).

Finally, the table shows that the odds of a match rating increase as a function of Age although it does not reach statistical significance (p >0.05).

Table 6.6 Output of a generalised linear mixed-effects model for match rating score

term	β	S.E	*z*	(p)
(Intercept)	0.27	0.18	1.51	0.1310
InstructionListen Only	0.49	0.08	5.85	**0.0000**
InstructionTraditional	0.05	0.08	0.64	0.5251
TestDelayed	0.03	0.03	0.91	0.3626
SoundCluster	1.68	0.23	7.4	**0.0000**
SoundDFricative	-0.1	0.2	-0.51	0.6071
SoundDiphthong	-0.48	0.22	-2.18	**0.0293**
SoundPlosive	0.02	0.22	0.07	0.9425
SoundRhotic	0.32	0.21	1.54	0.1224
TaskPicture-naming	0.25	0.13	1.88	0.0596
TaskRead aloud	0.12	0.14	0.83	0.4086
LiteracyLiterate	-0.08	0.08	-0.97	0.3343
GenderM	-0.08	0.08	-1.12	0.2645
Age	0.03	0.02	1.92	0.0555
InstructionListenOnly:TestDelayed	-0.02	0.05	-0.46	0.6436
InstructionTraditional:TestDelayed	0.08	0.04	1.92	0.0543
InstructionListenOnly:SoundCluster	0.47	0.14	3.33	**0.0009**
InstructionTraditional:SoundCluster	-0.37	0.11	-3.34	**0.0009**
InstructionListenOnly:SoundDFricative	-0.11	0.09	-1.23	0.2169
InstructionTraditional:SoundDFricative	-0.11	0.08	-1.3	0.1944
InstructionListenOnly:SoundDiphthong	-0.15	0.09	-1.73	0.0843
InstructionTraditional:SoundDiphthong	0.19	0.08	2.39	**0.0168**
InstructionListenOnly:SoundPlosive	0.14	0.1	1.4	0.1613
InstructionTraditional:SoundPlosive	0.46	0.09	5.07	**0.0000**
InstructionListenOnly:SoundRhotic	-0.42	0.09	-4.81	**0.0000**
InstructionTraditional:SoundRhotic	0.25	0.08	3.02	**0.0025**
TestDelayed:SoundCluster	0.06	0.08	0.76	0.4471
TestDelayed:SoundDFricative	-0.1	0.06	-1.79	0.0732
TestDelayed:SoundDiphthong	0.11	0.06	1.86	0.0635
TestDelayed:SoundPlosive	-0.09	0.06	-1.44	0.1509
TestDelayed:SoundRhotic	0.13	0.06	2.23	**0.0259**
LiteracyLiterate:GenderMale	-0.12	0.08	-1.49	0.1351
InstructionListenOnly:TestDelayed:SoundCluster	0.02	0.14	0.16	0.8706
InstructionTraditional:TestDelayed:SoundCluster	-0.03	0.11	-0.27	0.7846
InstructionListenOnly:TestDelayed:SoundDFricative	-0.01	0.09	-0.16	0.8712
InstructionTraditional:TestDelayed:SoundDFricative	-0.03	0.08	-0.37	0.7137
InstructionListenOnly:TestDelayed:SoundDiphthong	0.09	0.08	1.04	0.2972
InstructionTraditional:TestDelayed:SoundDiphthong	-0.09	0.08	-1.07	0.2825
InstructionListenOnly:TestDelayed:SoundPlosive	-0.12	0.1	-1.27	0.2046
InstructionTraditional:TestDelayed:SoundPlosive	0.01	0.09	0.06	0.9491
InstructionListenOnly:TestDelayed:SoundRhotic	0.05	0.09	0.52	0.6005
InstructionTraditional:TestDelayed:SoundRhotic	0	0.08	0.05	0.9636

To find out model-predicted probabilities of match rating for each instruction type, in each test by sound class, the function `emmeans` (Lenth, 2019) in `R` was applied using `type = 'response'`. The following code was used:

```
emmeans(mdl, pairwise ~ Instruction * Test | Sound,
                        type = 'response')
```

Intervals for predicted probabilities were back-transformed from the logit scale. The pairwise functions allows for pairwise comparisons specified by the formula, that was within-Test pairwise comparisons by Instruction for each Sound class. The Tukey adjustment method for comparing a family of 6 estimates was used. Tests were performed on the log odds ratio scale.

The post hoc within-test pairwise comparisons by training condition show that in the post-test, although the Listen and Speak condition outperformed its Listen Only counterpart in most segmental categories - an exception is diphthongs and plosives - this difference was only statistically significant for the group of affricates (p <.03) and rhotic approximants (p = 0.037). The Listen and Speak condition outperformed its Traditional counterpart and this difference was statistically significant for the group of affricates (p <.0001), coda clusters (p = 0.025), dental fricatives (p <.0001), diphthongs (p <.0001), and rhotic approximants (p = 0.001), but not for plosives. Finally, the Listen Only condition outperformed its Traditional counterpart during the post-test and the difference in the predicted probability of target-like rating between them was statistically significant for the group of dental fricatives (p = 0.0019), diphthongs (p <.0001), and plosives (p = 0.0039), but not for affricates, coda clusters, or rhotic approximants.

Within the delayed post-test, the Listen and Speak condition once again outperformed its Listen Only counterpart – an exception was the class of plosives. However, this difference was only statistically significant for the group of affricates (p = 0.019), and rhotic approximants (p <.0001). The Listen and Speak condition also outperformed its Traditional counterpart and the difference was statistically significant for the group of affricates (p = 0.002), dental fricatives (p = 0.02), diphthongs (p <.0001), but not for coda clusters, plosives, or rhotic approximants. Finally, the Listen Only condition outperformed the Traditional condition in the delayed post-test within all the segmental groups apart from rhotic approximants. Nevertheless, the difference in the predicted probability of match rating was statistically significant between those latter conditions only for the group of dental fricatives (p = 0.015), and diphthongs (p

<.0001), where the Listen Only condition outperformed its Traditional counterpart. For detailed predicted probabilities and post hoc within test pairwise comparisons for match rating across sound classes, see Appendix E.

6.5 Summary

Table 6.7 below summarises the results for Target-likeness and Match rating analyses. It indicates that the Listen and Speak group statistically outperformed the Listen Only group in affricates in the post-test and clusters in the delayed post-test within the Target-likeness results and affricates and rhotic approximants in both tests within the Match results. The Listen Only group did not statistically outperform the Listen and Speak counterpart in any sound class in either Target-likeness or Match results. Within the Target-likeness results, the Listen and Speak statistically outperformed the Traditional Teaching group in all sound classes except from the dental fricatives in the post-test and all except the coda clusters and the dental fricatives in the delayed post-test. It also statistically outperformed the Traditional group in all the sounds but plosives in the post-test and all sounds but clusters, plosives and rhotic approximants in the delayed post-test within the Match results. The Traditional Teaching group did not statistically outperform the Listen and Speak group in any class in either the Target-likeness or Match results.

As for Listen Only, based on the Target-likeness results, it statistically outperformed the Traditional Teaching group in all sounds but clusters and dental fricatives in the post-test and the diphthongs only in the delayed post-test. Within the Match rating results, it statistically outperformed the Traditional Teaching group in the dental fricatives and the diphthongs in both tests and the plosives in the post-test only. The Traditional Teaching group did not statistically outperform the Listen Only group in any of the sounds tested.

The Target-likeness and Match data partially corroborate HYPOTHESIS ONE, which stated that the Listen and Speak condition will result in the most target-like pronunciations followed by the Listen Only condition, leaving the Traditional condition with the relatively least target-like/input-matching pronunciations. We argue '*partially*' because statistical differences in performance were only found for a number

Table 6.7 Summary results for Target-likeness and Match rating

	Listen and Speak							
	post-test							
	Target-likeness				**Match**			
	Exceeds?	**Significant?**	**Exceeds?**	**Significant?**	**Exceeds?**	**Significant?**	**Exceeds?**	**Significant?**
	Listen Only		**Traditional**		**Listen Only**		**Traditional**	
Affricates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	Yes	No	Yes	Yes	Yes	No	Yes	Yes
Dental fricatives	Yes	No	Yes	No	Yes	No	Yes	Yes
Diphthongs	Yes	No	Yes	Yes	No	No	Yes	Yes
Plosives	Yes	No	Yes	Yes	No	No	Yes	No
Rhotics	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
	delayed post-test							
	Target-likeness				**Match**			
	Exceeds?	**Significant?**	**Exceeds?**	**Significant?**	**Exceeds?**	**Significant?**	**Exceeds?**	**Significant?**
	Listen Only		**Traditional**		**Listen Only**		**Traditional**	
Affricates	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	Yes	Yes	Yes	No	Yes	No	Yes	No
Dental fricatives	Yes	No	Yes	No	Yes	No	Yes	Yes
Diphthongs	Same	No	Yes	Yes	Yes	No	Yes	Yes
Plosives	Yes	No	Yes	Yes	No	No	Yes	No
Rhotics	Yes	No	Yes	Yes	Yes	Yes	Yes	No
	Listen Only							
	post-test							
	Target-likeness				**Match**			
	Exceeds?	**Significant?**	**Exceeds?**	**Significant?**	**Exceeds?**	**Significant?**	**Exceeds?**	**Significant?**
	Listen and Speak		**Traditional**		**Listen and Speak**		**Traditional**	
Affricates	No	Yes	Yes	Yes	No	Yes	Yes	No
Clusters	No	No	Yes	No	No	No	Yes	No
Dental fricatives	No	No	Yes	No	No	No	Yes	Yes
Diphthongs	No	No	Yes	Yes	Yes	No	Yes	Yes
Plosives	No	No	Yes	Yes	Yes	No	Yes	Yes
Rhotics	No	No	Yes	Yes	No	Yes	Yes	No
	delayed post-test							
	Target-likeness				**Match**			
	Exceeds?	**Significant?**	**Exceeds?**	**Significant?**	**Exceeds?**	**Significant?**	**Exceeds?**	**Significant?**
	Listen and Speak		**Traditional**		**Listen and Speak**		**Traditional**	
Affricates	No	No	Yes	No	No	Yes	Yes	No
Clusters	No	Yes	Yes	No	No	No	Yes	No
Dental fricatives	No	No	No	No	No	No	Yes	Yes
Diphthongs	Same	No	Yes	Yes	No	No	Yes	Yes
Plosives	No	No	Yes	No	Yes	No	Yes	No
Rhotics	No	No	Yes	No	No	Yes	No	No

of sound categories and not uniformly across all problematic sounds even though the raw percentages were mostly in favour of the Listen and Speak condition. It also confirms that there was an improvement in CAPT training conditions compared to the Traditional teaching. An exception to this was the Listen Only condition outperforming the two other conditions in the probabilities of matching productions for the class of diphthongs and plosives though no statistical difference was observed between this condition and the Listen and Speak condition. Overall, the target-likeness as well as the match data show that the Listen and Speak condition generally had the highest probability for a target-like and match rating in both tests. Followed by this was the Listen Only condition, whereas the probabilities for Traditional condition seem to be the lowest.

However, in the post-test, the probabilities of a target-like pronunciation for the experimental groups were not statistically different except in the case of affricates. Additionally, the Listen Only's probabilities of a target-like pronunciation for the class of clusters were not statistically different from those of the Traditional condition.

In the delayed test, the dynamics change for affricates, clusters, plosives, and approximant rhotics. The statistical difference between the experimental groups for the class of affricates was lost, whilst at the same time the Listen Only condition was not statistically different from the Traditional condition in this sound class. For the class of plosives and rhotic approximants, the Listen Only condition was no longer statistically different from the Traditional condition. It is not surprising that the Traditional condition had the lowest probability for a target-like rating since the input of the training was mostly – though not entirely – non-target-like. However, for the experimental conditions, despite Listen Only receiving native input, their performance was not statistically different for the class of affricates, clusters, plosives and approximant rhotics suggesting the important role of speech production practice.

The match data show further statistical differences between the experimental conditions. This was evident in the class of affricates and rhotic approximants. This suggests that output practice was particularly crucial for affricates and rhotic approximants. It also shows the interactional relationship between training condition and sound class.

The approximant rhotics, dental fricatives, and affricates[38] according to Flege's (1995) SLM are new sounds since their phonetic configuration was not similar to any of the learner's L1 sounds. According to Kuhl's (1994) NLM, their phonetic configuration does not fall in the psycho-acoustic space of any of the first language's prototypes.[39] Yet, learners from the three training conditions found difficulty learning the dental fricatives and approximant rhotics but not affricates. A study by Diehl and Lindblom (2004) suggests that not all sounds are equal in their degree of difficulty to produce. Studies of relative articulatory complexity reveal that initiating and maintaining voicing for instance was harder in obstruents compared to sonorants (J. Ohala, 1997). Additionally, voicing in dorsal places of articulation was harder than that in posterior locations in the oral cavity owing to the comparatively shorter cavity behind the constriction (Colantoni and Steele, 2008).

---

[38] The discussion of clusters will be dealt with later as theories of speech learning make predictions for segments and segmental contrast not sound sequences/phonotactic constraints.

[39] An exception was made for the class of Dental fricatives, which will be dealt with in Chapter 10.

6.6 Discussion

The results of the target-likeness rating revealed that overall, delayed speaking (Listen Only) did not have an advantage over *pressure* to speak (Listen and Speak) when it comes to target-like realisations . The Listen and Speak condition overall had the most target-like pronunciations. Although in many cases the experimental conditions were not statistically different, overall oral production practice does have an advantage over delayed production (particularly in affricates and coda clusters). This seems to dispute Krashen's (1982; 1985; 1994) proposal that not only does pressure to speak not improve learning but also inhibits it particularly in the early stages of L2 exposure, that is three weeks of training. Krashen (2018: 99) claims that,

> When acquirers were forced to produce language that they have not yet acquired, known as "forced speech," they often experience anxiety. I argue here that forced speech is not only uncomfortable, it makes no direct contribution to language acquisition.

The Listen and Speak training condition had an advantage over the Listen Only training condition as they were doubly consolidating their learning through listening *and* speaking practice. In other words, they were storing exemplars based on perception and tuning these and changing them based on production and practice.

The hybrid model (Colantoni and Steele, 2008) proposes that practising speaking produces output, which subsequently provides feedback to the learner. This feedback enables a learner to modify his/her speech accordingly. The continued modification allows mental representations to evolve although it does not guarantee accent-free speech.

Chapter 7: Voicing

7.1 Introduction

The aim of this chapter is to present the results of the phonetic learning outcome mainly as a function of Instruction type – Listen and Speak, Listen Only, and Traditional condition. The first research question posed in Chapter 5 relates to all problematic sounds. This chapter deals with the phonetic learning, specifically that in plosives. Thus, this chapter seeks to answer a sub-part of the research question:

RQ1: Which training method will result in more target-like (and less L1 like) pronunciations of problematic sounds Libyan learners typically display learning English?

- **Plosives**

Phonetic learning will also be compared across the two times of testing. If participants maintain learning after 10 weeks of the end of training – as examined in the delayed test – this means that learning took place. The phonetic parameters explored are divided into voicing contrasts and place contrasts. Voicing cues include voice onset time and vowel-onset fundamental frequency. As for place, spectral tilt (Ahi-A23) in bilabial and coronal plosives is explored.

The chapter is divided into the following sections. Section 7.2 deals with VOT duration outcome. It first explores various effects within each condition and compares the learners' L2 against their L1 and the training input they received. Because the effect of test is not directly comparable with the teacher's TL, given that the factor does not apply, data for each training condition is subset to post-test and delayed test and compared separately from one another. At the end of the analyses, all training conditions are compared to each other without data for the TL to test differences between training conditions and differences across time and across tasks. Section 7.3 deals with vowel-onset f0 outcome. It follows the same procedure carried out for VOT durations. Section 7.5 provides a summary of the results and concludes with a discussion and implications.

7.2 Voice Onset Time

7.2.1 *Descriptive statistics*

Table 7.1 below shows mean VOT durations by for the target input for the experimental groups and the Traditional condition. The Traditional teacher clearly shows non-native VOT durations.

Table 7.1 Mean voice onset times (ms) of the DigLin speakers and the Traditional teacher

		DigLin speakers	Traditional teacher
/p/	(+)	103	27
/b/	(+)	6	3
	(-)		-76
/t/	(+)	125	35
/d/	(+)	19	
	(-)		-65

Figure 7.1 shows VOT durations for L1 and L2. Although overall L2 VOT values are higher than those for L1 (exceptions are the voiced categories in the Listen and Speak group, the voiceless categories in the Traditional, each of which during the delayed post-test), they do not seem to vary considerably from one another. It is also noticed that the L1 VOT durations for the experimental groups, and more notably for the Listen and Speak condition, are longer on the positive side and shorter on the negative side.

7.2.2 *Model Specification*

The main goal from this step of the analysis is to explore differences in VOT values between the language groups L2, L1 and TL (Language). We want to account for potential speaker-related effects of Literacy and Gender. In terms of VOT related effects, we want to control for voicing category, Place of articulation and Vowel context. Adding

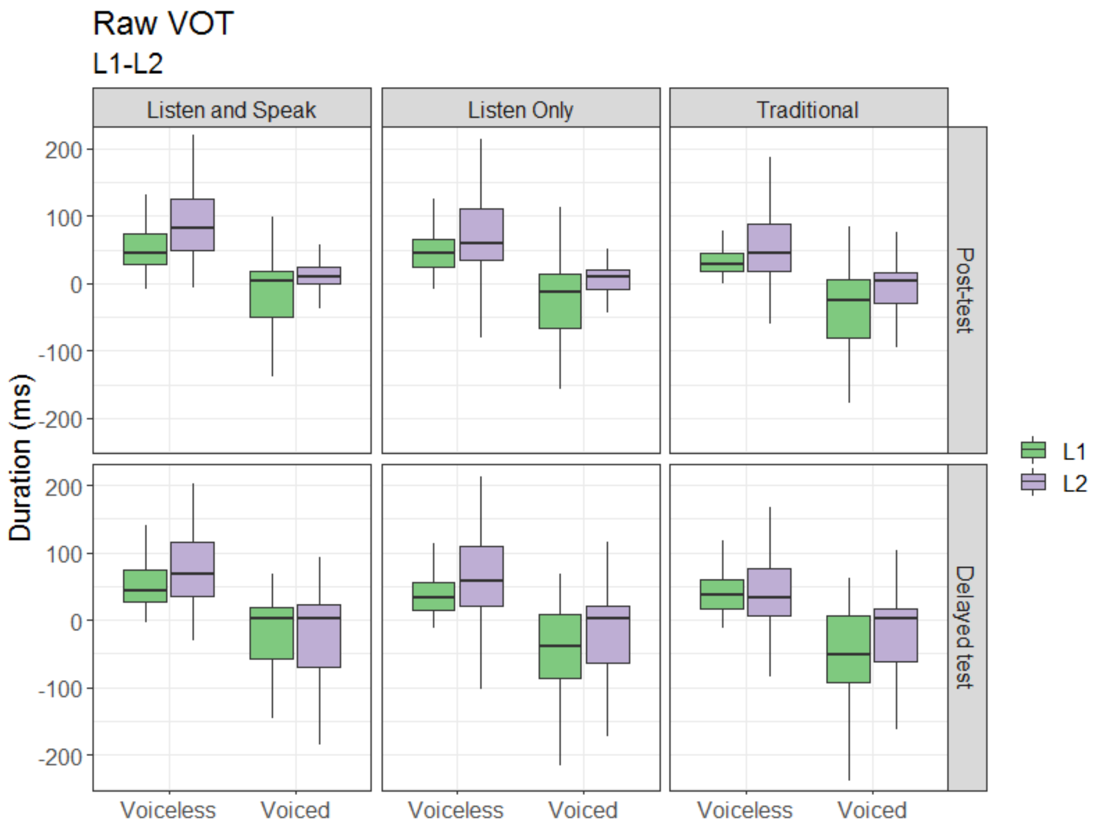


Figure 7.1 VOT duration (ms). The line inside the boxplot represents the mean

the effect of 'Age' created convergence issues. When the function `all_fit(mdl)` in the package `afex` (Singmann et al., 2019) is used, the model converged but standard errors for Language were very large. The Estimated Marginal Means in the post hoc were not a true reflection of the data as they were quite extreme particularly for the Target Language. Interaction terms for Language and Voice are added as we want the model to capture variance in voice within each Language. Additional interaction terms for Gender and Literacy are also added where relevant. Adding the effect of 'Task' created the following warning messages:

```
contrasts dropped from factor Language due to missing levels
```

This has caused the model to drop the TL level from the 'Language' effect since data elicited from DigLin could not be assigned similar tasks to those of the learners and data elicited from the Traditional teacher was using read aloud and picture-naming but no delayed repetition is used since it was not needed. As a result, this effect is dropped from the model as unlike the TL language level, it is not considered a priority.

The model added by-speaker and by-item random intercepts. By-speaker random slopes for Voice, Vowel, and Place are added. Adding by-item random slopes of Language, Literacy, and Gender did not allow the model to converge even after decorrelating slopes from the intercept. To this end, a linear mixed-effects model with all of the fixed effects and random effects and slopes is built. The following model is considered:

```
Duration ~ Language * Voice + Gender * Literacy + Place
          + Vowel + (1 + Voice + Vowel + Place | speaker)
          + (1 | item)
```

An effort was made to keep the model uniform across the training conditions and across the times of testing. Where the model is modified owing to failure to converge, this is pointed out in the relevant section.

7.2.3 *Estimated Marginal Means and post hoc tests*

The effects of Language and Voice in the above model are simple effects since they are involved in interaction terms. Therefore, differences in mean VOT duration will be dealt with in a follow-up post hoc test for all of the models which include interaction terms. The same applies to the effects of Gender and Literacy: they are simple effects since they are part of an interaction. To find out Estimated Marginal Means for VOT values for each Language group, for each Voice category, the function `emmeans` (Lenth, 2019) in `R` is applied. The following code is used:

```
emmeans(mdl, pairwise ~ Language * Voice)
```

Results are averaged over the levels of Gender, Literacy, Place, and Vowel. Degrees-of-freedom are calculated using the Kenward-Roger method using confidence intervals of 0.95. The pairwise functions allows for pairwise comparisons specified by the formula, that is pairwise comparisons by Language and Voice. The Tukey adjustment method for P values is used for comparing a family of six estimates.

Finally, to find out whether differences in mean VOT values as a function of Language and Voice category are statistically significant, post hoc pairwise comparisons were conducted.

7.2.4 *Listen and Speak*

Post-test

Inferential statistics on VOT durations in Listen and Speak condition during post-test

Table 7.2 shows the results of a linear mixed-effects model for VOT by the learners of Listen and Speak condition during the post-test compared to Diglin. The reference for Language is ‘L2’, for Gender ‘Male’, for Vowel ‘Back-High’, for Literacy ‘Literate’, for Voice ‘Voiceless’, and for Place ‘Bilabial’. In this model, the intercept represents mean VOT values for a Voiceless Bilabial plosive followed by a Back-High vowel and produced by Male Literate participants from Listen and Speak condition. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients.

The LMM output in table 7.2, shows that there is an increase by an average of 2 ms in the VOT duration of Coronals compared to their Bilabial cognates, but it is not statistically significant (p >0.5). For differences in VOT duration as a function of Vowel, there is an overall marginal decrease by an average of less than 1 ms in the VOT of plosives preceding a Back-Low vowel compared to a Back-High vowel. There is also an overall decrease by an average of 11 ms in the VOT of plosives preceding a Front-High vowel compared to a Back-High vowel. Finally, there is an overall increase by an average of 7 ms in the VOT of plosives followed by a Front-Low vowel compared to a Back-High vowel. However, neither of these trends is statistically significant.

Table 7.2 Output of a linear mixed-effects model for VOT in Listen and Speak post-test

term	β	S.E	df	*t*	(p)
(Intercept)	39.03	8.76	32.35	4.46	**9.41e-05**
LanguageL1	0.49	7.36	36.31	0.07	0.94718
LanguageTL	-25.76	8.32	36.94	-3.10	**0.00371**
VoiceVoiced	44.59	6.44	54.07	6.93	**5.37e-09**
GenderF	-0.30	5.29	19.21	-0.06	0.95493
LiteracyIlliterate	-1.36	5.33	18.88	-0.26	0.80092
PlaceCoronal	2.47	4.61	20.96	0.54	0.59780
VowelBack-Low	-0.52	5.95	14.76	-0.09	0.93163
VowelFront-High	-11.45	8.21	28.42	-1.39	0.17393
VowelFront-Low	6.51	6.16	18.76	1.06	0.30337
LanguageL1:VoiceVoiced	-1.13	5.80	103.63	-0.19	0.84636
LanguageTL:VoiceVoiced	-8.81	6.89	35.92	-1.28	0.20891
GenderF:LiteracyIlliterate	-7.48	5.29	19.24	-1.41	0.17359

We are interested in differences between L2 and L1 on the one hand and L2 and TL on the other. We are also interested in whether the learners statistically distinguish voiceless from voiced VOT categories.

The results of the post hoc within-voiced and within-voiceless pairwise comparisons (see table F.1 in Appendix F) show that the learners following the Listen and Speak condition failed to establish L2 categories that are statistically independent of their L1 categories. At the same time, their L2 categories were not statistically independent of the TL categories either. Table 7.3 indicates that the learners managed however to establish a voicing contrast of the TL evident in their L2 VOT values and this contrast is statistically significant (p <.0001).

Table 7.3 Results of post hoc pairwise comparisons for VOT durations by Language and Voice for Listen and Speak in Post-test

Language	contrast	estimate	SE	df	t.ratio	p.value
L2	voiceless - voiced	86.9	11.8	28.3	7.342	**<.0001**

Delayed test

In the delayed post-test, the interaction between Gender and Literacy was removed as the model failed to converge for this subset of the data.

Inferential statistics on VOT durations in Listen and Speak condition during delayed test

Table 7.4 shows the results of a linear mixed-effects model for VOT by the learners of Listen and Speak condition during the delayed test compared to Diglin. The reference for Language is 'L2', for Gender 'Male', for Vowel 'Back-High', for Literacy 'Literate', for Voice 'Voiceless', and for Place 'Bilabial'. In this model, the intercept represents mean VOT values for a Voiceless Bilabial plosive followed by a Back-High vowel and produced by Male Literate participants from Listen and Speak condition. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients.

The LMM output in table 7.4, shows that there is an increase by an average of 2 ms in the VOT duration of Coronals compared to their Bilabial cognates, but it is not statistically significant (p >0.7). For differences in VOT duration as a function of Vowel, there is an overall decrease by an average of 8 ms in the VOT of plosives preceding a Back-Low vowel compared to a Back-High vowel. There is also an overall decrease by an average of 6 ms in the VOT of plosives preceding a Front-High vowel compared to a Back-High vowel. However, none of these trends is statistically significant (see table 7.4). Finally, there is an overall statistically significant increase by an average of 16 ms in the VOT of plosives followed by a Front-Low vowel compared to a Back-High vowel (p <.05).

Table 7.4 Output of a linear mixed-effects model for VOT in Listen and Speak delayed test

term	β	S.E	df	t	(p)
(Intercept)	41.25	9.11	25.30	4.53	**0.000123**
LanguageL1	-12.44	7.28	36.30	-1.71	0.095974
LanguageTL	-20.91	8.73	31.93	-2.40	**0.022618**
VoiceVoiced	46.49	6.93	35.63	6.71	**8.43e-08**
GenderF	-7.75	6.20	12.60	-1.25	0.233617
LiteracyIlliterate	-9.83	6.97	12.00	-1.41	0.184179
PlaceCoronal	1.88	5.04	16.47	0.37	0.713454
VowelBack-Low	-7.53	7.40	18.58	-1.02	0.321644
VowelFront-High	-5.68	9.05	40.76	-0.63	0.533874
VowelFront-Low	15.62	6.82	15.97	2.29	**0.035964**
LanguageL1:VoiceVoiced	2.78	6.13	61.01	0.45	0.651330
LanguageTL:VoiceVoiced	-9.72	7.76	27.48	-1.25	0.220838

We are interested in differences between L2 and L1 on the one hand and L2 and TL on the other and whether the learners continued to statistically distinguish voiceless from voiced L2 VOT categories. The results of the post hoc within-voiced and within-voiceless pairwise comparisons (see table F.1 in Appendix F) demonstrate that the Listen and Speak learners continued to use intermediary L2 VOT categories as none of them was statistically independent of their L1 VOT categories whilst at the same time, the L2 VOT categories were not statistically independent of the TL VOT categories either. Table 7.5 indicates that the learners maintained a voicing contrast of the TL plosives evident in their L2 VOT values and this contrast is statistically significant (p <.0001).

Table 7.5 Results of post hoc pairwise comparisons for VOT durations by Language and Voice for Listen and Speak in Delayed test

Language	contrast	estimate	SE	df	t.ratio	p.value
L2	voiceless - voiced	98.55	14.3	25.5	6.869	**<.0001**

7.2.5 *Listen Only*

Inferential statistics on VOT durations in Listen Only condition during post-test

Table 7.6 shows the results of a linear mixed-effects model for VOT by the learners of Listen Only condition during the post-test compared to Diglin. The reference for Language is 'L2', for Gender 'Male', for Vowel 'Back-High', for Literacy 'Literate', for Voice 'Voiceless', and for Place 'Bilabial'. In this model, the intercept represents mean VOT values for a Voiceless Bilabial plosive followed by a Back-High vowel and produced by Male Literate participants from Listen Only condition. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients.

The effects of Language and Voice are simple effects since they are involved in interaction terms. Therefore, differences in mean VOT duration will be dealt with in a follow-up post hoc test. The same applies to the effects of Gender and Literacy: they are simple effects since they are part of an interaction.

The LMM output in table 7.6, shows that there is an increase by an average of 2 ms in the VOT duration of Coronals compared to their Bilabial cognates, but it is not statistically significant (p >0.5). For differences in VOT duration as a function of Vowel, there is an overall statistically significant decrease (p <.03) by an average of 12 ms in the VOT of plosives preceding a Back-Low vowel compared to a Back-High vowel. There is an overall increase by an average of 6 ms in the VOT of plosives preceding a Front-High vowel compared to a Back-High vowel, but it is not statistically significant (p >.3). Finally, there is an overall statistically significant increase by an average of 11 ms in the VOT of plosives followed by a Front-Low vowel compared to a Back-High vowel (p <.05).

Table 7.6 Output of a linear mixed-effects model for VOT in Listen Only post-test

term	β	S.E	df	*t*	(p)
(Intercept)	47.04	8.81	30.91	5.34	**8.12e-06**
LanguageL1	-5.92	6.78	44.18	-0.87	0.3878
LanguageTL	-25.57	7.39	44.28	-3.46	**0.0012**
VoiceVoiced	45.96	6.01	47.71	7.65	**7.65e-10**
GenderF	2.19	5.77	16.96	0.38	0.7095
LiteracyIlliterate	-12.44	5.80	16.81	-2.15	**0.0467**
PlaceCoronal	2.31	3.47	19.15	0.66	0.5144
VowelBack-Low	-11.92	4.84	20.14	-2.46	**0.0229**
VowelFront-High	6.33	7.25	36.39	0.87	0.3888
VowelFront-Low	11.41	4.97	19.70	2.30	**0.0326**
LanguageL1:VoiceVoiced	-4.93	5.67	68.88	-0.87	0.3872
LanguageTL:VoiceVoiced	-6.30	6.35	45.55	-0.99	0.3271
GenderF:LiteracyIlliterate	4.41	5.77	16.97	0.76	0.4553

We are interested in differences between L2 and L1 on the one hand and L2 and TL on the other and whether a statistical distinction between voiceless and voiced L2 VOT categories was made. The results of the post hoc within-voiceless and within-voiced pairwise comparisons (see table F.2 in Appendix F) indicate that those following the Listen Only training did not establish L2 VOT categories independent from their L1 VOT categories immediately after the training. Instead, their L2 VOT categories were intermediate between L1 and TL VOT values, and were not statistically different from the TL VOT categories either.

Table 7.7 shows that the learners established a statistically significant voicing contrast ($p <.0001$) of the TL plosives evident in their L2 VOT values.

Table 7.7 Results of post hoc pairwise comparisons for VOT durations by Language and Voice for Listen Only in Post-test

Language	contrast	estimate	SE	df	t.ratio	p.value
L2	voiceless - voiced	82.1	10.64	29.9	7.715	**<.0001**

Delayed test

In the delayed post-test, the interaction between Gender and Literacy was removed as the model failed to converge for this subset of the data.

Inferential statistics on VOT durations in Listen Only condition during delayed test

Table 7.8 shows the results of a linear mixed-effects model for VOT by the learners of Listen Only condition during the delayed test compared to Diglin. The reference for Language is 'L2', for Gender 'Male', for Vowel 'Back-High', for Literacy 'Literate', for Voice 'Voiceless', and for Place 'Bilabial'. In this model, the intercept represents mean VOT values for a Voiceless Bilabial plosive followed by a Back-High vowel and produced by Male Literate participants from Listen Only condition. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients.

The effects of Language and Voice are simple effects since they are involved in interaction terms. Therefore, differences in mean VOT duration will be dealt with in a follow-up post hoc test.

The LMM output in table 7.8, shows an overall decrease by an average of 5 ms in the VOT values of Females compared to Males, but it is not statistically significant (p >0.3). There is also an overall statistically significant decrease (p <.05) by an average of 16 ms in the VOT durations of Illiterates compared to Literates. The table also shows that there is an increase by an average of 5 ms in the VOT duration of Coronals compared to their Bilabial cognates, but it is not statistically significant (p >0.3). For differences in VOT duration as a function of Vowel, there is an overall decrease by an average of 6 ms in the VOT of plosives preceding a Back-Low vowel compared to a Back-High vowel. There is also an overall marginal decrease by an average of 2 ms in the VOT of plosives preceding a Front-High vowel compared to a Back-High vowel. Finally, there is an overall increase by an average of 9 ms in the VOT of plosives followed by a Front-Low vowel compared to a Back-High vowel. However, neither of these trends is statistically significant.

Table 7.8 Output of a linear mixed-effects model for VOT in Listen Only delayed test

term	β	S.E	df	*t*	(p)
(Intercept)	41.98	10.55	27.08	3.98	**0.000466**
LanguageL1	-8.73	7.80	37.97	-1.12	0.269927
LanguageTL	-30.44	9.01	37.27	-3.38	**0.001712**
VoiceVoiced	47.22	7.59	45.02	6.22	**1.48e-07**
GenderF	-4.83	5.51	16.73	-0.88	0.393461
LiteracyIlliterate	-16.34	7.50	16.00	-2.18	**0.044674**
PlaceCoronal	4.56	4.59	16.63	0.99	0.334096
VowelBack-Low	-6.31	6.82	19.06	-0.93	0.366104
VowelFront-High	-1.86	8.67	59.65	-0.21	0.830803
VowelFront-Low	9.42	6.74	19.86	1.40	0.177757
LanguageL1:VoiceVoiced	-2.19	6.64	72.77	-0.33	0.742441
LanguageTL:VoiceVoiced	-4.98	7.97	37.87	-0.63	0.535677

We are interested in differences between L2 and L1 on the one hand and L2 and TL on the other and whether a voicing distinction in L2 VOT is made.

The results of the post hoc within-voiceless and within-voiced pairwise comparisons demonstrate that learners from the Listen Only condition continued to merge their L2 VOT categories in either voiceless or voiced plosives with their respective L1 VOT categories (see table F.2 in Appendix F). However, the L2 VOT values were not statistically independent of the TL VOT categories. Table 7.9 indicates that the learners established a statistically significant voicing contrast ($p <$.0001) of the TL plosives evident in their L2 VOT values.

Table 7.9 Results of post hoc pairwise comparisons for VOT durations by Language and Voice for Listen Only in Delayed test

Language	contrast	estimate	SE	df	t.ratio	p.value
L2	voiceless - voiced	90.1	14.3	30.1	6.298	**<.0001**

7.2.6 *Traditional*

Post-test

In the post-test, the interaction between Gender and Literacy was removed as the model failed to converge for this subset of the data. Adding a by-speaker random slope for Place did not allow the model to converge even after decorrelating slopes from the intercept. Therefore, it was dropped.

Inferential statistics on VOT durations in Traditional condition during post-test

Table 7.10 shows the results of a linear mixed-effects model for VOT by the learners of Traditional teaching condition during the post-test compared to the Traditional teacher. The reference for Language is 'L2', for Gender 'Male', for Vowel 'Back-High', for Literacy 'Literate', for Voice 'Voiceless', and for Place 'Bilabial'. In this model, the intercept represents mean VOT values for a Voiceless Bilabial plosive followed by a Back-High vowel and produced by Male Literate participants from Traditional condition. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients.

The effects of Language and Voice are simple effects since they are involved in interaction terms. Therefore, differences in mean VOT duration will be dealt with in a follow-up post hoc test.

The LMM output in table 7.10, shows an overall marginal decrease by an average of 1 ms in the VOT values of Females compared to Males, but it is not statistically significant (p >0.7). There is an overall increase by an average of 6 ms in the VOT durations of Illiterates compared to Literates, but it is statistically non-significant (p >0.4). The table also shows that there is a marginal increase by an average of less than 1 ms in the VOT duration of Coronals compared to their Bilabial cognates, but it is not statistically significant (p >0.8). For differences in VOT duration as a function of Vowel, there is an overall decrease by an average of 5 ms in the VOT of plosives preceding a Back-Low vowel compared to a Back-High vowel. There is an overall increase by an average of 7 ms in the VOT of plosives preceding a Front-High vowel compared to a Back-High vowel. Finally, there is an overall marginal decrease by an average of 2 ms in the VOT of plosives followed by a Front-Low vowel compared to a Back-High vowel. However, none of these trends is statistically significant.

Table 7.10 Output of a linear mixed-effects model for VOT in Traditional post-test

term	β	S.E	df	*t*	(p)
(Intercept)	-3.35	9.79	27.74	-0.34	0.73486
LanguageL1	22.40	6.73	27.70	3.33	**0.00247**
LanguageTL	1.80	7.65	27.80	0.24	0.81544
VoiceVoiced	39.50	5.81	15.05	6.80	**5.91e-06**
GenderF	-1.39	4.39	20.57	-0.32	0.75520
LiteracyIlliterate	5.84	7.55	25.68	0.77	0.44578
PlaceCoronal	0.59	3.71	13.87	0.16	0.87605
VowelBack-Low	-4.74	5.69	21.42	-0.83	0.41379
VowelFront-High	7.11	7.37	40.47	0.96	0.34037
VowelFront-Low	-2.22	7.32	13.42	-0.30	0.76611
LanguageL1:VoiceVoiced	-6.86	5.66	20.96	-1.21	0.23920
LanguageTL:VoiceVoiced	-1.13	6.87	21.27	-0.16	0.87071

We are interested in differences between L2 and L1 on the one hand and L2 and TL on the other and whether a voicing contrast was established in L2 VOT categories. The results of within-voiceless and within-voiced pairwise comparisons (see table F.3 in Appendix F) show that, like the learners in the CAPT conditions, Traditional Teaching learners did not create L2 VOT categories that are independent of their L1. However, they were not statistically independent of the Traditional teacher's VOT categories either.

Table 7.11 shows that the learners exhibited a statistically significant voicing contrast (p <.0001) in their L2 VOT values, despite the Traditional teacher's lack of such statistically significant contrast in her mean VOTs (p >0.1).

Table 7.11 Results of post hoc pairwise comparisons for VOT durations by Language and Voice for Traditional in Post-test

Language	contrast	estimate	SE	df	t.ratio	p.value
L2	voiceless - voiced	65.3	10.6	17.63	6.166	**0.0001**
TL	voiceless - voiced	95.0	32.5	10.35	2.922	0.1123

Delayed test

In the delayed post-test, the interaction between Gender and Literacy was again removed as the model failed to converge for this subset of the data. Adding a by-speaker random slope for Place however allowed the model to converge Therefore, it was kept.

Inferential statistics on VOT durations in Traditional condition during delayed test

Table 7.12 shows the results of a linear mixed-effects model for VOT by the learners of Traditional teaching condition during the delayed test compared to the Traditional teacher. The reference for Language is 'L2', for Gender 'Male', for Vowel 'Back-High', for Literacy 'Literate', for Voice 'Voiceless', and for Place 'Bilabial'. In this model, the intercept represents mean VOT values for a Voiceless Bilabial plosive followed by a Back-High vowel and produced by Male Literate participants from Traditional condition. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients.

The effects of Language and Voice are simple effects since they are involved in interaction terms. Therefore, differences in mean VOT duration will be dealt with in a follow-up post hoc test.

The LMM output in table 7.12, shows an overall marginal decrease by an average of 2 ms in the VOT values of Females compared to Males, but it is not statistically significant (p >0.6). There is also an overall decrease by an average of 3 ms in the VOT durations of Illiterates compared to Literates, but it is statistically non-significant (p >0.8). The table also shows that there is an increase by an average of 6 ms in the VOT duration of Coronals compared to their Bilabial cognates, but it is not statistically significant (p >0.2). For differences in VOT duration as a function of Vowel, there is an overall increase by an average of 3 ms in the VOT of plosives preceding a Back-Low vowel compared to a Back-High vowel. There is also an overall decrease by an average of 4 ms in the VOT of plosives preceding a Front-High vowel compared to a Back-High vowel. Finally, there is an overall marginal increase by an average of less than 1 ms in the VOT of plosives followed by a Front-Low vowel compared to a Back-High vowel. However, neither of these trends is statistically significant.

Table 7.12 Output of a linear mixed-effects model for VOT in Traditional delayed test

term	β	S.E	df	*t*	(p)
(Intercept)	-15.06	15.41	27.55	-0.98	0.3368
LanguageL1	18.02	10.45	25.01	1.73	0.0968
LanguageTL	2.04	10.90	26.66	0.19	0.8527
VoiceVoiced	40.79	6.38	24.77	6.39	**1.12e-06**
GenderF	-2.07	5.26	22.18	-0.39	0.6980
LiteracyIlliterate	2.60	11.88	24.71	0.22	0.8287
PlaceCoronal	-5.82	4.53	18.47	-1.29	0.2143
VowelBack-Low	3.29	6.64	26.56	0.50	0.6243
VowelFront-High	-4.16	9.73	43.96	-0.43	0.6713
VowelFront-Low	0.75	7.22	21.63	0.10	0.9185
LanguageL1:VoiceVoiced	-12.66	6.20	37.07	-2.04	**0.0484**
LanguageTL:VoiceVoiced	4.32	7.31	26.93	0.59	0.5596

We are interested in differences between L2 and L1 on the one hand and L2 and TL on the other and whether a voicing distinction was made. The results of within-voiceless and within-voiced pairwise comparisons (see table F.3 in Appendix F) show that, once again the Traditional Teaching participants did not create L2 VOT categories that are independent of their L1. However, their L2 VOT categories were not statistically independent of the Traditional teacher's VOT categories in this test either. Table 7.13 shows that the learners established a statistically significant voicing contrast ($p <.0021$) of the TL plosives evident in their L2 VOT values.

Table 7.13 Results of post hoc pairwise comparisons for VOT durations by Language and Voice for Traditional in Delayed test

Language	contrast	estimate	SE	df	t.ratio	p.value
L2	voiceless - voiced	56.260	12.8	26.0	4.386	**0.0021**

7.2.7 *By-Condition results*

Model specification

The main goal from this step of the analysis is to explore differences in VOT values between training conditions 'Instruction' and time of test 'Test', the language groups L1 and L2. We want to account for potential speaker-related effects of Gender and Age. We also want to control for elicitation task effect. In terms of VOT related effects, we want to control for Voice category, Place, Vowel context, Gender, and potentially Age. Interaction terms for Instruction, Language and Voice are added as we want the model to capture variance in Voice category within each Language within each training condition. We also want to capture interaction between Place of articulation and Vowel context.

The model added by-speaker and by-item random intercepts. By-speaker random slopes for Language would not allow the model to converge. Therefore, only by-speaker random slopes for interaction terms of Language and Voice as well as random slopes for Test, and interaction terms for Place and Vowel are added. Adding by-item random slopes for Language, Instruction, Test, Gender and Age would not allow the model to converge. To this end, a linear mixed-effects model with all of the fixed effects and random effects (intercepts and slopes) is built. The following model is considered:

```
Duration ~ Instruction * Language * Voice + Gender + Place *
        Vowel + Test + Age + Task + (1 + Language * Voice + Test +
        Place * Vowel | speaker) + (1 | item)
```

Inferential statistics on VOT in all training conditions

Table 7.14 shows the results of a linear mixed-effects model for VOT durations for all training conditions in both tests. The reference for Instruction is 'Traditional', for Language 'L1', for Voice 'Voiceless', for Gender 'Male', for Test 'Post-test', and for Task 'Delayed repetition'. In this model, the intercept represents mean VOT values for Voiceless bilabial plosives preceding a Back-High vowel, produced by Male Literate participants from the Traditional condition in the Delayed repetition task during the Post-test. All fixed effects are sum coded. The effect of Age is centred.

The effects of Instruction, Language, Voice, Place and Vowel are simple effects since they are involved in interaction terms. The two- and three-way interaction terms are also simple effects. Therefore, differences in mean VOT values for these effects will be dealt with in a follow-up post hoc test.

In terms of the effects of Gender, Test, Age, and Task, the LMM output in table 7.14 shows that there is an overall marginal decrease by an average of 2 ms in the mean VOT values of females compared to males, but it is not statistically significant (p >0.3). There is an overall statistically significant increase (p <.001) by an average of 5 ms in the mean VOT of plosives in the delayed test compared to those in the post-test. For Age, the table shows that there is no difference in mean VOT as a function of Age. There is an overall decrease by an average of 4 ms in the mean VOT as produced by Illiterates compared to their Literate peers, but it does not reach statistical significance (p >0.2). However, there is an overall statistically significant increase (p <.01) by an average of 10 ms in the mean VOT values of plosives produced in the Picture-naming task compared to those in the delayed repetition task. Also, there is an overall statistically significant decrease (p <.01) by an average of 7 ms in the mean VOT of plosives in the Read aloud task compared to those in the delayed repetition task.

Table 7.14 Output of a linear mixed-effects model for VOT in all training conditions

term	β	S.E	df	*t*	(p)
(Intercept)	18.82	4.88	42.17	3.86	**0.000384**
InstructionListen and Speak	-4.52	3.41	59.36	-1.33	0.190092
InstructionListen Only	6.55	3.36	54.61	1.95	0.056916
LanguageL2	-2.22	4.10	29.26	-0.54	0.592040
VoiceVoiced	41.09	3.64	17.79	11.27	1.56e-09
GenderF	-1.89	2.19	54.47	-0.87	0.390212
PlaceCoronal	0.67	3.52	15.26	0.19	0.850471
VowelBack-Low	-3.65	4.41	25.88	-0.83	0.414814
VowelFront-High	1.97	4.51	90.19	0.44	0.663854
VowelFront-Low	2.74	4.83	18.82	0.57	0.577160
TestDelayed test	5.04	1.78	68.39	2.84	**0.005974**
Age	-0.31	0.72	58.74	-0.43	0.666036
Literacy	-3.59	3.09	60.10	-1.16	0.249281
TaskPicture-naming	9.73	3.17	2470.38	3.07	**0.002173**
TaskRead aloud	-7.05	1.78	2537.23	-3.97	**7.54e-05**
InstructionListen and Speak:LanguageL2	0.07	1.97	61.10	0.03	0.973505
InstructionListen Only:LanguageL2	0.49	1.89	53.77	0.26	0.798049
InstructionListen and Speak:VoiceVoiced	-3.22	2.24	65.15	-1.43	0.156770
Instruction2:VoiceVoiced	2.54	2.21	62.15	1.15	0.254834
Language1:Voice1	0.82	3.39	13.68	0.24	0.811443
PlaceCoronal:VowelBack-Low	3.65	4.32	24.39	0.85	0.405923
PlaceCoronal:VowelFront-High	-7.02	4.34	79.24	-1.62	0.109504
PlaceCoronal:VowelFront-Low	-3.36	4.90	20.07	-0.69	0.500557
InstructionListen and Speak:LanguageL2:VoiceVoiced	6.92	1.58	69.71	4.38	4.12e-05
InstructionListen Only:LanguageL2:VoiceVoiced	-5.31	1.53	63.19	-3.47	**0.000938**

To find out whether differences in mean VOT values as a function of Instruction type within-Language groups and within-Voice categories are statistically significant, post hoc pairwise comparisons are conducted. We are interested in differences between training conditions within-Voice and within-Language. The results of the post hoc show that there are no statistical differences in the model-predicted mean VOT durations between Instruction types in any of the Voicing categories or the Language groups (see table F.4 in Appendix F).

7.3 Vowel-onset f0

7.4 Model Specification

The main goal from this step of the analysis is to explore differences in vowel-onset f0 values between the language groups L2, L1 and TL (Language). We want to account for potential speaker-related effect of Gender. Adding the effect of 'Age' yielded very large standard errors for the 'Language' effect. The Estimated Marginal Means in the post hoc did not seem to be a true reflection of the data as they were quite extreme particularly for the Target Language. In terms of f0 related effects, we want to control for Voice category and Gender. Interaction terms for Language, Gender and Voice are added as we want the model to capture variance in Gender and Voice category within each Language. Adding the effect of 'Task' created the following warning messages:

```
contrasts dropped from factor Language due to missing levels
```

This has caused the model to drop the TL level from the 'Language' effect since data elicited from DigLin could not be assigned similar tasks to those of the learners and data elicited from the Traditional teacher was using read aloud and picture-naming but no delayed repetition is used since it was not needed. As a result, this effect is dropped from the model as unlike the TL language level, it is not considered a priority. This is true of all the models used in this chapter.

The models added by-speaker and by-item random intercepts. By-speaker random slopes for Language would not allow the model to converge. Therefore, only by-speaker random slopes for Voice are added. Similarly, adding by-item random slopes for Language and Gender would not allow the model to converge even after decorrelating slopes from the intercept. Models in the vowel-onset f0 analyses behaved differently with this respect. However, an effort is made to keep the models as uniform as possible for comparable results. To this end, a linear mixed-effects model with all of the fixed effects and random effects (intercepts and slopes) is built. The following model is considered:

```
f0.start ~ Language * Gender * Voice +
            (1 + Voice | speaker) + (1 | item)
```

7.4.1 *Listen and Speak*

Descriptive statistics

Table 7.15 illustrates vowel-onset f0 data for the Listen and Speak condition in the post-test and delayed test. It shows that the mean vowel-onset f0 preceded by phonologically voiceless plosives is higher than that preceded by phonologically voiced plosives. The mean vowel-onset f0 for female speakers in the post-test does not exhibit any voice contrast within the group's L2 data. In the delayed test, female learners show a mean difference of 4 Hz only between the voice categories of their L2 data.

The table also shows that in the post-test, female learners tend to have higher mean vowel-onset f0, whereas in the delayed test, males tend to have a higher mean vowel-onset f0. The figures also indicate that based on standard deviations, a) male learners generally tend to have greater variability than females, b) in the post-test, L1 vowel-onset f0 generally exhibits greater variability than L2, and c) in the delayed test, L2 vowel-onset f0 exhibits greater variability than L1. The figures also indicate that the learners' vowel-onset f0 means are closer to those of the female speaker in DigLin than they are to the male speaker, who illustrated considerably the lowest f0 means.

Table 7.15 Descriptive statistics for vowel-onset fundamental frequency (Hz) with mean and standard deviation (SD) for the two-way contrast within Listen and Speak condition

		Male		Female	
Language	Voice	Mean	SD	Mean	SD
TL	voiceless	141.	5.15	248.	3.20
	voiced	119.	2.96	240.	9.92
Post-test					
L1	voiceless	267.	37.1	284.	30.9
	voiced	255.	32.1	270.	31.8
L2	voiceless	267.	29.7	270.	31.0
	voiced	259.	44.5	270.	27.0
Delayed Post-test					
L1	voiceless	271.	24.1	264.	13.0
	voiced	268.	24.0	255.	24.2
L2	voiceless	279.	49.2	259.	33.9
	voiced	256.	35.2	255.	35.2

Post-test

Inferential statistics on vowel-onset f0 in Listen and Speak condition during post-test

Table 7.16 shows the results of a linear mixed-effects model for vowel-onset f0 by the learners of Listen and Speak condition during the post-test compared to Diglin. The reference for Language is 'L2', for Gender 'Male', and for Voice 'voiceless'. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients. The effects of Language, Gender and Voice are simple effects since they are involved in interaction terms. Therefore, differences in mean vowel-onset f0 values will be dealt with in a follow-up post hoc test.

Table 7.16 Output of a linear mixed-effects model for vowel-onset f0 in Listen and Speak condition during post-test

term	β	S.E	df	*t*	(p)
(Intercept)	241.06	7.98	20.83	30.22	**<2e-16**
LanguageL1	27.97	7.67	23.13	3.64	**0.00134**
LanguageTL	25.89	7.51	21.50	3.45	**0.00236**
GenderFemale	-23.12	3.89	41.48	-5.94	**5.02e-07**
VoiceVoiced	5.86	3.28	33.47	1.79	0.08303
LanguageL1:GenderFemale	15.73	2.95	184.43	5.33	**2.88e-07**
LanguageTL:GenderFemale	19.76	2.98	189.08	6.64	**3.31e-10**
LanguageL1:VoiceVoiced	0.50	3.56	40.07	0.14	0.88981
LanguageTL:VoiceVoiced	-2.35	3.20	43.58	-0.73	0.46630
GenderFemale:VoiceVoiced	1.82	2.50	308.23	0.73	0.46596
LanguageL1:GenderFemale:VoiceVoiced	-2.82	2.55	577.74	-1.11	0.26920
LanguageTL:GenderFemale:VoiceVoiced	0.27	2.58	575.41	0.11	0.91601

The results of the post hoc pairwise comparison in table 7.17 show that within the group of males, the difference in fundamental frequency values between L2 and L1 within the voiceless as well as the voiced categories are above the JND for pitch discrimination proposed as 1 Hz for a frequency range up to 1 kHz (Stevens, 2000), albeit these differences are statistically non-significant. JND is short for Just Noticeable Difference, a measure of psychoacoustic perception typically based on logarithmic characteristics of frequency but also applicable for VOT durations amongst other acoustic measurements. A difference in frequency or duration can be statistically different but not necessarily high/long enough to be noticeable for a listener to perceive a difference. In terms of differences between L2 and TL, there is a statistically significant increase by an average of 129 Hz (p <.001) and 141 Hz (p <.001) in the L2 f0 values compared to that of the TL within the voiceless and voiced category respectively. For within-language comparisons, although the contrast between voiceless and voiced categories within each language group exceeded the JND for pitch discrimination, none of the language groups show a statistically significant voice contrast.

Table 7.17 Results of post hoc within-Gender pairwise comparisons for vowel-onset f0 values by Language and Voice for Listen and Speak during the post-test

Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Male	voiceless	L2 - L1	2.191	6.50	35.2	0.337	0.9994
		L2 - TL	128.536	24.51	27.1	5.245	**0.0002**
	voiced	L2 - L1	1.703	5.55	35.6	0.307	0.9996
		L2 - TL	141.496	27.93	30.4	5.066	**0.0003**

As for within the group of females, none of the differences in fundamental frequency values between a) L1 and L2 within the voiceless category, b) L1 and L2 within the voiced category, c) L2 and TL within the voiceless category, d) L2 and TL within the voiced category, or e) the voiceless and voiced categories within the L2 language group, reaches statistical significance.[40]

Delayed test

Inferential statistics on vowel-onset f0 in Listen and Speak condition during delayed test

Table 7.18 shows the results of a linear mixed-effects model for vowel-onset f0 by the learners of Listen and Speak condition during the delayed test compared to Diglin. The reference for Language is ‘L2’, for Gender ‘Male’, and for Voice ‘voiceless’. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients. The effects of Language, Gender and Voice are simple effects since they are involved in interaction

---

[40] See table H.1, Appendix H.

terms. Therefore, differences in mean vowel-onset f0 values will be dealt with in a follow-up post hoc test.

Table 7.18 Output of a linear mixed-effects model for vowel-onset f0 in Listen and Speak condition during delayed test

term	β	S.E	df	*t*	(p)
(Intercept)	237.89	8.12	11.97	29.31	**1.62e-12**
LanguageL1	26.44	7.51	12.97	3.52	**0.00378**
LanguageTL	24.26	7.51	12.96	3.23	**0.00661**
GenderFemale	-16.11	4.47	19.68	-3.60	**0.00181**
VoiceVoiced	5.76	3.05	22.67	1.89	0.07167
LanguageL1:GenderFemale	20.87	3.22	78.21	6.47	**7.71e-09**
LanguageTL:GenderFemale	19.94	3.26	80.77	6.12	**3.14e-08**
LanguageL1:VoiceVoiced	0.74	3.14	33.13	0.24	0.81444
LanguageTL:VoiceVoiced	-2.69	3.14	37.76	-0.86	0.39649
GenderFemale:VoiceVoiced	1.97	2.57	159.53	0.77	0.44493
LanguageL1:GenderFemale:VoiceVoiced	2.40	2.66	384.54	0.90	0.36699
LanguageTL:GenderFemale:VoiceVoiced	-3.71	2.69	383.88	-1.38	0.16937

The results of the post hoc pairwise comparison in table 7.19 show that within the group of males, the difference in fundamental frequency values between L2 and L1 within the voiceless as well as the voiced categories are above the JND for pitch discrimination proposed as 1 Hz for a frequency range up to 1 kHz (Stevens, 2000), albeit these differences are statistically non-significant. In terms of differences between L2 and TL, there is a statistically significant increase by an average of 128 Hz ($p <.01$) and 148 Hz ($p <.001$) in the L2 f0 values compared to that of the TL within the voiceless and voiced category respectively. For within-language comparisons, although the contrast between voiceless and voiced categories within each language group exceeded the JND for pitch discrimination, none of the language groups show a statistically significant voice contrast.

Table 7.19 Results of post hoc within-Gender pairwise comparisons for vowel-onset f0 values by Language and Voice for Listen and Speak during delayed test

Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Male	voiceless	L2 - L1	-8.30	5.87	93.0	-1.413	0.7193
		L2 - TL	128.23	26.57	15.4	4.826	**0.0023**
	voiced	L2 - L1	10.79	5.14	97.0	2.097	0.2976
		L2 - TL	147.55	26.54	20.0	5.559	**0.0002**

As for within the group of females, none of the differences in fundamental frequency values between any of the language groups either within-voiced or within-voiceless, or across-voicing for the L2 language group were statistically significant.[41]

7.4.2 *Listen Only*

Descriptive statistics

Table 7.20 illustrates vowel-onset f0 data for the Listen Only condition in the post-test and delayed test.The figures in table 7.20 show that the mean vowel-onset f0 preceded by phonologically voiceless plosives is greater than that preceded by phonologically voiced plosives. The table also shows that in either test, female learners tend to have surprisingly lower mean vowel-onset f0 values. The figures also indicate that based on standard deviations, a) the greatest variability is evident in vowel-onset f0 values within the L2 language group produced by male learners during the delayed test, and b) large variability is also observable in vowel-onset f0 values within the L1 language group produced by female learners during the post-test. The figures also indicate that the learners' vowel-onset f0 means in the delayed test are higher than those in the post-test with the exception of the L2 f0 values of the voiced context as produced by male learners.

---

[41] See table H.1 Appendix H.

Table 7.20 Descriptive statistics for vowel-onset fundamental frequency (Hz) with mean and standard deviation (SD) for the two-way voice contrast within Listen Only condition

		Male		Female	
Language	Voice	Mean	SD	Mean	SD
TL	voiceless	141.	5.15	248.	3.20
	voiced	119.	2.96	240.	9.92
Post-test					
L1	voiceless	268.	26.5	256.	46.3
	voiced	261.	30.7	247.	30.5
L2	voiceless	273.	26.9	260.	35.6
	voiced	268.	36.5	249.	33.8
Delayed Post-test					
L1	voiceless	272.	44.2	267.	38.4
	voiced	262.	33.9	253.	37.0
L2	voiceless	271.	34.6	263.	34.5
	voiced	256.	48.0	254.	30.9

Post-test

Inferential statistics on vowel-onset f0 in Listen Only condition during post-test

Table 7.21 shows the results of a linear mixed-effects model for vowel-onset f0 by the learners of Listen Only condition during the post-test compared to Diglin. The reference for Language is ‘L2’, for Gender ‘Male’, and for Voice ‘voiceless’. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients. The effects of Language, Gender and Voice are simple effects since they are involved in interaction terms. Therefore, differences in mean vowel-onset f0 values will be dealt with in a follow-up post hoc test.

Table 7.21 Output of a linear mixed-effects model for vowel-onset f0 in Listen Only condition during post-test

term	β	S.E	df	*t*	(p)
(Intercept)	236.39	10.20	16.92	23.18	**2.92e-14**
LanguageL1	27.49	9.50	17.16	2.89	**0.01005**
LanguageTL	21.71	9.59	17.75	2.26	**0.03634**
GenderFemale	-14.90	4.67	22.09	-3.19	**0.00422**
VoiceVoiced	5.03	2.92	33.63	1.72	0.09461
LanguageL1:GenderFemale	22.02	2.96	55.74	7.43	**6.79e-10**
LanguageTL:GenderFemale	21.31	2.93	53.85	7.26	**1.58e-09**
LanguageL1:VoiceVoiced	-1.69	2.83	42.29	-0.60	0.55245
LanguageTL:VoiceVoiced	-0.99	3.10	40.17	-0.32	0.75160
GenderFemale:VoiceVoiced	0.76	2.07	211.67	0.36	0.71588
LanguageL1:GenderFemale:VoiceVoiced	-2.91	2.09	506.47	-1.39	0.16582
LanguageTL:GenderFemale:VoiceVoiced	-1.07	2.06	505.90	-0.52	0.60227

The results of the post hoc pairwise comparison in table 7.22 show that within the group of males, the difference in fundamental frequency values between L2 and L1 within the voiceless as well as the voiced categories are above the JND for pitch discrimination proposed as 1 Hz for a frequency range up to 1 kHz (Stevens, 2000), albeit these differences are statistically non-significant. In terms of differences between L2 and TL, there is a statistically significant increase by an average of 131 Hz (p <.01) and 153 Hz (p <.001) in the L2 f0 values compared to that of the TL within the voiceless and voiced category respectively. For within-language comparisons, although the contrast between voiceless and voiced categories within each language group exceeded the JND for pitch discrimination, none of the language groups show a statistically significant voice contrast.

Table 7.22 Results of post hoc within-Gender pairwise comparisons for vowel-onset f0 values by Language and Voice for Listen Only during post-test

Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Male	voiceless	L2 - L1	3.95	5.07	26.7	0.779	0.9688
		L2 - TL	130.77	31.35	18.5	4.172	**0.0061**
	voiced	L2 - L1	9.03	4.43	29.6	2.039	0.3455
		L2 - TL	153.29	30.45	20.8	5.034	**0.0007**

As for within the group of females, the difference in fundamental frequency values between L1 and L2 within either voicing category are above the JND for pitch discrimination, but this difference is not statistically significant. In terms of differences between L2 and TL, there is an increase by an average of 14 Hz and 9 Hz in the L2 f0 values compared to that of the TL within the voiceless and voiced category respectively. However, neither reaches statistical significance. The contrast between voiceless and voiced categories within each language group is well above the JND for pitch discrimination. However, none of the language groups show a statistically significant voice contrast.[42]

Delayed test

Inferential statistics on vowel-onset f0 in Listen Only condition during delayed test

Table 7.23 shows the results of a linear mixed-effects model for vowel-onset f0 by the learners of Listen Only condition during the delayed test compared to Diglin. The reference for Language is 'L2', for Gender 'Male', and for Voice 'voiceless'. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients. The effects of Language, Gender and Voice are simple effects since they are involved in interaction terms.

Table 7.23 Output of a linear mixed-effects model for vowel-onset f0 in Listen Only condition during delayed test

term	β	S.E	df	*t*	(p)
(Intercept)	237.52	11.13	16.25	21.35	**2.56e-13**
LanguageL1	24.54	10.34	16.53	2.37	**0.0301**
LanguageTL	25.78	10.49	17.42	2.46	**0.0247**
GenderFemale	-18.23	5.28	22.57	-3.45	**0.0022**
VoiceVoiced	6.39	3.00	45.95	2.13	**0.0385**
LanguageL1:GenderFemale	20.27	3.49	66.61	5.80	**1.98e-07**
LanguageTL:GenderFemale	21.43	3.47	64.81	6.18	**4.74e-08**
LanguageL1:VoiceVoiced	-0.68	2.93	75.00	-0.23	0.8174
LanguageTL:VoiceVoiced	-0.64	3.41	41.98	-0.19	0.8522
GenderFemale:VoiceVoiced	1.94	2.44	374.83	0.80	0.4258
LanguageL1:GenderFemale:VoiceVoiced	-0.33	2.59	507.58	-0.13	0.8987
LanguageTL:GenderFemale:VoiceVoiced	-2.79	2.56	507.43	-1.09	0.2751

---

[42] See table H.2, Appendix H.

The results of the post hoc pairwise comparison in table 7.24 show that within the group of males, there is no difference in fundamental frequency values between L2 and L1 within the voiceless category. However, within the voiced category, there is a decrease in the fundamental frequency value by an average of 5 Hz of L2 compared to that of L1 and such difference is beyond the JND for pitch discrimination proposed as 1 Hz for a frequency range up to 1 kHz (Stevens, 2000), albeit these differences are statistically non-significant.

Table 7.24 Results of post hoc within-Gender pairwise comparisons for vowel-onset f0 values by Language and Voice for Listen Only condition during delayed test

Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Male	voiceless	L2 - L1	0.0111	6.61	28.9	0.002	1.0000
		L2 - TL	131.3805	34.86	18.2	3.769	**0.0147**
	voiced	L2 - L1	-4.8356	5.85	33.3	-0.827	0.9604
		L2 - TL	142.2822	32.69	21.6	4.352	**0.0031**

In terms of differences between L2 and TL, there is a statistically significant increase by an average of 131 Hz ($p = 0.01$) and 142 Hz ($p < .01$) in the L2 f0 values compared to that of the TL within the voiceless and voiced category respectively. For within-language comparisons, although the contrast between voiceless and voiced categories within each language group exceeded the JND for pitch discrimination, none of the language groups show a statistically significant voice contrast.

As for within the group of females, the difference in fundamental frequency values between L1 and L2 within either voicing category are above the JND for pitch discrimination, but this difference is not statistically significant. In terms of differences between L2 and TL, there is an increase by an average of 14 Hz and 11 Hz in the L2 f0 values compared to that of the TL within the voiceless and voiced category respectively. However, neither reaches statistical significance. The contrast between voiceless and voiced categories within each language group is well above the JND for pitch discrimination. However, none of the language groups show a statistically significant voice contrast.[43]

---

[43] See table H.2, Appendix H.

7.4.3 *Traditional*

Descriptive statistics

Table 7.25 Descriptive statistics for vowel-onset fundamental frequency (Hz) with mean and standard deviation (SD) for the two-way voice contrast within Traditional condition

		Male		Female	
Language	Voice	Mean	SD	Mean	SD
TL	voiceless	–	–	226.	18.2
	voiced	–	–	216.	13.4
Post-test					
L1	voiceless	277.	63.0	250.	92.4
	voiced	278.	23.9	270.	55.1
L2	voiceless	291.	33.4	275.	50.5
	voiced	280.	42.5	281.	42.8
Delayed Post-test					
L1	voiceless	280.	63.1	267.	46.2
	voiced	267.	50.8	264.	34.7
L2	voiceless	270.	43.3	277.	55.4
	voiced	275.	38.0	269.	50.9

Post-test

Inferential statistics on vowel-onset f0 in Traditional condition during post-test

Table 7.26 shows the results of a linear mixed-effects model for vowel-onset f0 by the learners of Traditional condition during the post-test compared to the Traditional teacher. The reference for Language is ‘L2’, for Gender ‘Male’, and for Voice ‘voiceless’. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients. The effects of Language, Gender and Voice are simple effects since they are involved in interaction terms. Therefore, differences in mean vowel-onset f0 values will be dealt with in a follow-up post hoc test. Male TL non-estimable because there is only one female teacher.

Table 7.26 Output of a linear mixed-effects model for vowel-onset f0 in Traditional condition during post-test

term	β	S.E	df	*t*	(p)
(Intercept)	259.00	11.77	16.61	22.00	**1.02e-13**
LanguageL1	21.28	11.40	17.18	1.87	0.0791
LanguageTL	7.78	11.91	20.47	0.65	0.5208
GenderFemale	5.95	8.41	23.92	0.71	0.4861
VoiceVoiced	1.73	3.71	120.55	0.47	0.6424
LanguageL1:GenderFemale	-3.03	4.82	324.73	-0.63	0.5291
LanguageL1:VoiceVoiced	-1.12	3.86	145.20	-0.29	0.7713
LanguageTL:VoiceVoiced	-6.88	5.17	238.03	-1.33	0.1845
GenderFemale:VoiceVoiced	3.32	4.10	271.46	0.81	0.4186
LanguageL1:GenderFemale:VoiceVoiced	-1.11	4.72	322.34	-0.24	0.8142

The results of the post hoc pairwise comparison in table H.3, Appendix H show that within the group of males, the difference in fundamental frequency values between L2 and L1 within the voiceless and voiced category exceeds the JND for pitch discrimination (Stevens, 2000). However, these differences are statistically non-significant. For within-language comparisons, although the contrast between voiceless and voiced categories within each language group exceeded the JND for pitch discrimination, none of the language groups show a statistically significant voice contrast.

As for within the group of females, the difference in fundamental frequency values between L2 and L1 within either voicing category are way beyond the JND for pitch discrimination, but this difference is not statistically significant. In terms of differences between L2 and TL, there is an increase by an average of 50 Hz and 63 Hz in the L2 f0 values compared to that of the TL within the voiceless and voiced category respectively. However, neither reaches statistical significance. The contrast between voiceless and voiced categories within each language group is well above the JND for pitch discrimination. However, none of the language groups show a statistically significant voice contrast.

Delayed test

Inferential statistics on vowel-onset f0 in Traditional condition during delayed test

Table 7.27 shows the results of a linear mixed-effects model for vowel-onset f0 by the learners of Traditional condition during the delayed test compared to the Traditional teacher. The reference for Language is 'L2', for Gender 'Male', and for Voice 'voiceless'. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients. The effects of Language, Gender and Voice are simple effects since they are involved in interaction terms. Therefore, differences in mean vowel-onset f0 values will be dealt with in a follow-up post hoc test.

Table 7.27 Output of a linear mixed-effects model for vowel-onset f0 in Traditional condition during delayed test

term	β	S.E	df	*t*	(p)
(Intercept)	257.44	13.32	15.50	19.32	**2.88e-12**
LanguageL1	12.64	12.40	16.17	1.02	0.3231
LanguageTL	14.00	12.37	15.97	1.13	0.2742
GenderFemale	4.13	8.53	15.80	0.48	0.6351
VoiceVoiced	5.11	2.84	176.30	1.80	0.0737
LanguageL1:GenderFemale	-5.71	3.03	532.78	-1.88	0.0604
LanguageL1:VoiceVoiced	-6.16	3.16	408.50	-1.95	0.0518
LanguageTL:VoiceVoiced	-0.66	3.02	78.94	-0.22	0.8284
GenderFemale:VoiceVoiced	1.60	1.98	459.30	0.81	0.4196
LanguageL1:GenderFemale:VoiceVoiced	-5.02	3.02	532.71	-1.66	0.0969

The results of the post hoc pairwise comparison in table H.3, Appendix H show that within the group of males, the difference in fundamental frequency values between L2 and L1 within the voiceless and voiced category exceeds the JND for pitch discrimination (Stevens, 2000). However, these differences do not reach statistical significance. For within-language comparisons, although the contrast between voiceless and voiced categories within each language group exceeded the JND for pitch discrimination, none of the language groups show a statistically significant voice contrast.

As for within the group of females, the difference in fundamental frequency values between L2 and L1 within either voicing category are way beyond the JND for pitch discrimination, but this difference is not statistically significant. In terms of differences between L2 and TL, there is an increase by an average of 48 Hz and 54 Hz in the L2 f0 values compared to that of the TL within the voiceless and voiced category respectively. However, neither reaches statistical significance. The contrast between voiceless and voiced categories within each language group is well above the JND for pitch discrimination. However, none of the language groups show a statistically significant voice contrast.

7.4.4 *By-Condition results*

Model specification

The main goal from this step of the analysis is to explore differences in vowel-onset f0 values between training conditions 'Instruction' and time of test 'Test', the language groups L1 and L2. We want to account for potential speaker-related effects of Gender and Age. We also want to control for elicitation task effect. In terms of f0 related effects, we want to control for Voice category, Gender, and potentially Age. Interaction terms for Instruction, Language, Voice and Gender are added as we want the model to capture variance in Gender and Voice category within each Language within each training condition.

The model added by-speaker and by-item random intercepts. By-speaker random slopes for Language would not allow the model to converge. Therefore, only by-speaker random slopes for interaction terms of Test and Voice as well as random slopes for Age and Task are added. By-item random slopes for Language are added. Adding by-item random slopes for Language, Instruction, Test, Gender and Age would not allow the model to converge. To this end, a linear mixed-effects model with all of the fixed effects and random effects (intercepts and slopes) is built. The following model is considered:

```
Onset.f0 ~ Instruction * Language * Voice * Gender + Test +
     Age + Task + (1 + Test * Voice + age + Task | speaker)
     + (1 | item)
```

Inferential statistics on vowel-onset f0 in all training conditions

Table 7.28 shows the results of a linear mixed-effects model for vowel-onset f0 for all training conditions in both tests. The reference for Instruction is ‘Traditional’, for Language ‘L1’, for Voice ‘Voiceless’, for Gender ‘Male’, for Test ‘Post-test’, and for Task ‘Delayed repetition’. In this model, the intercept represents mean onset f0 values for vowel following Voiceless plosives produced by Male participants from the Traditional condition in the Delayed repetition task during the post-test. All fixed effects are sum coded. The effect of age is centred.

The effects of Instruction, Language, Voice and Gender are simple effects since they are involved in interaction terms. The two-, three-, and four-way interaction terms are also simple effects. Therefore, differences in mean vowel-onset f0 values for these effects will be dealt with in a follow-up post hoc test.

In terms of the effects of Test, Age, and Task, the LMM output in table 7.28 shows that there is an overall marginal increase by an average of just over 1 Hz in the f0 values of the delayed test compared to the post-test, but it is not statistically significant (p >0.5). For differences in mean f0 as a function of Age, there is an overall marginal decrease by an average of less than 1 Hz in the mean f0 with Age and it is not statistically significant (p >0.8). There is also an overall slight decrease by an average of just over 1 Hz in the mean f0 of vowel onsets in the Picture-naming task compared to those in the delayed repetition task, but it is statistically non-significant (p >0.5). However, there is an overall slight decrease by an average of approximately 2 Hz in the f0 of vowel onsets in the Read aloud task compared to those in the delayed repetition task, but it is not statistically significant (p >0.1).

Table 7.28 Output of a linear mixed-effects model for vowel-onset f0 in all training conditions

term	β	S.E	df	*t*	(p)
(Intercept)	269.82	3.35	50.16	80.63	**<2e-16**
InstructionListen and Speak	10.09	4.29	42.76	2.35	**0.0233**
InstructionListen Only	-2.07	4.28	37.93	-0.48	0.6310
LanguageL2	-1.79	1.79	58.29	-1.00	0.3222
VoiceVoiced	3.86	1.39	21.49	2.78	**0.0112**
GenderF	4.48	3.17	39.27	1.41	0.1660
TestDelayed test	1.17	1.95	37.10	0.60	0.5508
Age	-0.22	1.13	26.72	-0.19	0.8495
Taskpicture-naming	-1.33	2.23	227.86	-0.59	0.5527
TaskRead aloud	1.59	1.19	119.95	1.34	0.1837
InstructionListen and Speak:LanguageL2	-0.30	1.18	156.79	-0.26	0.7989
InstructionListen Only:LanguageL2	0.89	1.19	134.41	0.75	0.4549
InstructionListen and Speak:VoiceVoiced	-1.25	0.89	56.12	-1.40	0.1677
InstructionListen Only:VoiceVoiced	1.11	0.88	49.71	1.26	0.2131
LanguageL2:VoiceVoiced	0.83	1.31	17.32	0.63	0.5347
InstructionListen and Speak:GenderF	-0.17	4.29	41.79	-0.04	0.9690
InstructionListen Only:GenderF	-3.43	4.30	35.31	-0.80	0.4297
LanguageL2:GenderF	0.43	0.84	126.95	0.51	0.6102
VoiceVoiced:GenderF	0.48	0.62	50.70	0.77	0.4438
InstructionListen and Speak:LanguageL2:VoiceVoiced	0.02	0.75	2662.78	0.02	0.9829
InstructionListen Only:LanguageL2:VoiceVoiced	0.77	0.73	2690.01	1.06	0.2877
InstructionListen and Speak:LanguageL2:GenderF	2.11	1.17	163.63	1.81	0.0729
InstructionListen Only:LanguageL2:GenderF	-1.87	1.19	133.10	-1.58	0.1162
InstructionListen and Speak:VoiceVoiced:GenderF	0.51	0.89	55.72	0.57	0.5725
InstructionListen Only:VoiceVoiced:GenderF	-0.26	0.88	49.70	-0.30	0.7666
LanguageL2:VoiceVoiced:GenderF	0.36	0.52	2687.97	0.70	0.4870
InstructionListen and Speak:LanguageL2:VoiceVoiced:GenderF	0.69	0.75	2667.07	0.92	0.3601
InstructionListen Only:LanguageL2:VoiceVoiced:GenderF	-0.06	0.73	2688.73	-0.09	0.9294

To find out Estimated Marginal Means for vowel-onset f0 means for each Instruction type, Language group, and Voice category, the function `emmeans` (Lenth, 2019) in `R` is applied. The following code is used:

```
emmeans(mdl, pairwise ~ Instruction * Language * Voice,
                              lmer.df = 'satterthwaite')
```

Results are averaged over the levels of Gender, Test and Task . Degrees-of-freedom are calculated using the Satterthwaite method using confidence intervals of 0.95. The pairwise functions allows for pairwise comparisons specified by the formula, that is pairwise comparisons by Instruction, Language and Voice. The Tukey adjustment method for P values is used for comparing a family of twelve estimates.

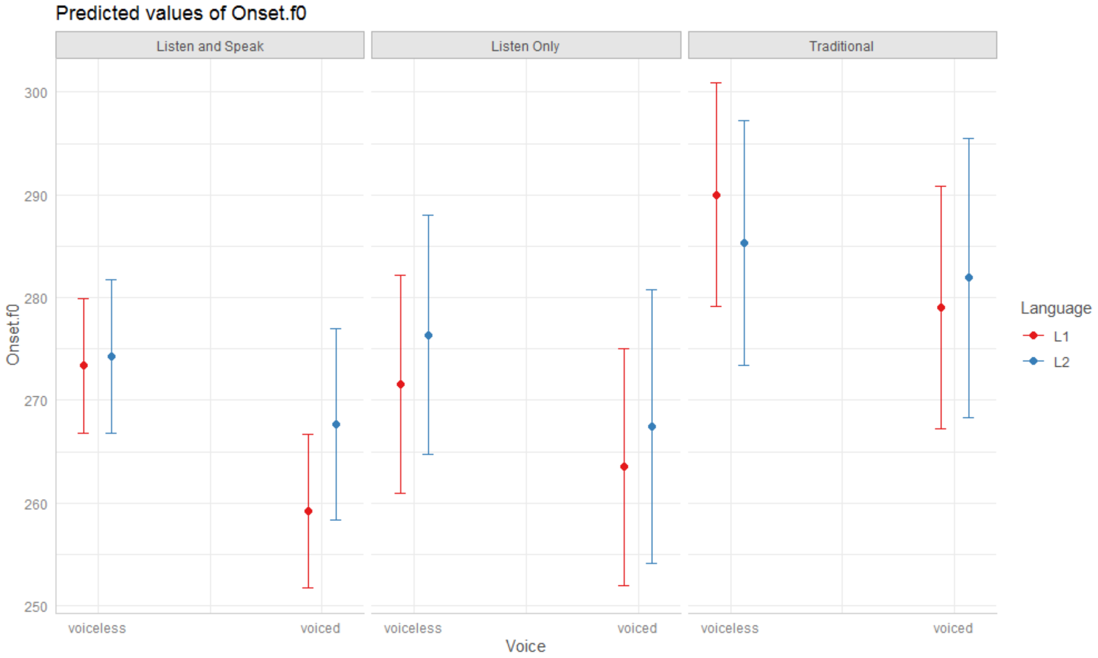


Figure 7.2 Predicted values of vowel-onset f0 by Instruction, Language group, and Voice

Within-voice pairwise comparisons show that the difference in L2 vowel-onset f0 within voiceless contexts and voiced contexts across the training conditions is not statistically different.[44] Neither of the training conditions exhibited a vowel-onset f0-based voicing contrast (across-voicing categories) in either L1 or L2. It can be concluded that there is no significant effect of training condition on vowel-onset f0 in either test.

7.5 Summary and discussion

7.5.1 *Voice Onset Time*

The voicing contrast reflected by L2 VOT is illustrated by all training conditions and is maintained throughout the post-test and delayed test. For the Traditional condition, this contrast is successfully manifested despite the Traditional teacher's lack of statistically significant L2 contrast between the two voicing categories. Despite the lack of a phonological contrast based on voicing for bilabial plosives in Libyan Arabic, whereby only the voiced bilabial plosive is incorporated in the phonemic inventory, the

---

[44] See table H.4 in Appendix H.

learners seem to have extrapolated the VOT contrast from other places of articulation to the bilabial one. A study by Bond and Fokes (1991) on the perception of the English voicing contrast of plosives by Arabic learners of English (among other L1 backgrounds) revealed that despite the lack of consistency in identifying voicing contrast in different plosives, the performance of Arabic learners was better with alveolar plosives compared to their bilabial cognates. They attribute this difficulty to the lack of a bilabial phonemic contrast in their L1 phonological system. A case similar to the current study was observed in Saudi Arabic learners of English residing in the United States (Flege and Port, 1981). The voicing contrast illustrated by all the training conditions is also evidence that despite the lack of a *phonemically* voiceless bilabial plosive in L1 Arabic, the learners did not assimilate the TL contrast into a single category. Rather, a Two Category TC assimilation is observed here (Best, 1994; Best, 1995).

As for the establishment of L2 VOT categories, in the post-test and based on within-Voice (not across-Voice) comparisons, neither the experimental conditions – Listen and Speak and Listen Only – nor the Traditional condition create new L2 VOT categories that are statistically distinct from either L1 or TL for either voiceless or voiced plosives. In the delayed test, the experimental conditions – Listen and Speak and Listen Only – and the Traditional learners continue to assimilate the VOT category of L1 and L2, which is also statistically indistinguishable from TL for either voiceless and voiced plosives. It was discussed in Chapter 3 that Arabic voiceless plosives (Khattab, 2002b; Kriba, 2009) and English voiced plosives (Lisker and Abramson, 1964) fall into the short-lag category. We have also seen in Chapter 2 what this proximity means for the perception of such TL sounds by non-native learners.

Based on Perceptual Assimilation Model-L2 (Best and Tyler, 2007), L2 learners create a common phonological space linking the phonetic and phonological learning through the concept of goodness of fit. The extended model also argues for a bidirectional interaction between L1 and TL, whereby learning to perceive TL categories or contrasts and distinguishing them from L1 categories may have an impact on how L1 sounds are perceived (Best and Tyler, 2007). This could explain the relatively slightly higher L1 VOT values for the voiceless plosives and the relatively shorter L1 lead voicing illustrated by the experimental conditions compared to the Traditional condition.

According to NLM's *Perceptual Magnet effect* by Kuhl (1991) and Kuhl, K. Williams, et al. (1992), the VOT of English voiced plosives is *harder* to acquire. This is because the perceptual space surrounding the native sound category is *warped*. Based on one mechanism, any TL sound category, for example the VOT of /b, d/, that is acoustically close (short-lag) to a proto-typical sound category – in this case the proto-type of L1 Arabic VOT of /t/ – is *attracted* to it, reducing the perceived distance between them and consequently assimilating them into a single category. Conversely, the opposing mechanism, *maximises* the perceived distance around the edges of a phonetic boundary – for example TL long-lag /p, t/ – which is not in the vicinity of any proto-type – in this case L1 Arabic VOT categories of /t, b, d/ – and making it as dissimilar as possible (Kuhl and P. Iverson, 1995: 141). That is because the acoustic signal embedded in an L2 sound contrast is processed through native language cue-weighting resulting in incorrect perceptual representations of the L2 sound contrast as well as longer processing time (P. Iverson, Kuhl, et al., 2003).

Similarly, according to the SLM's *Equivalence classification* by Flege (1995), TL English Voiced (short lag) VOT is considered an *old/similar* sound since it is utilised for the Voiceless VOT categories in Arabic and is thus considered *harder* to acquire. Due to the effect of equivalence classification, perceptually similar sounds block new category formation. Instead a merger category evolves, reflecting both L1 and TL phonetic input, which also alters the L1 category.

Also based on SLM, the TL English Voiceless (long lag) VOT category is considered a *new* sound since it is neither utilised for the Voiceless or Voiced VOT categories in Arabic and is thus considered *easier* to acquire. Therefore, L2 Voiceless and TL voiced VOT means are expected to be statistically indistinguishable but L2 mean VOT is expected to be *more exaggerated* and longer than that for TL. L2 voiced /b, d/ will be indistinguishable from L1 voiceless. L2 voiceless will be statistically different from L1 voiced. Although L2 mean VOT for the voiceless category is statistically indistinguishable from that for TL, it is also indistinguishable from L1 but not Longer than TL. For Listen and Speak learners, the temporal gap between the mean VOT of L2 and L1 on the one hand and L2 and TL on the other for the voiceless and voiced plosives surpass the Just Noticeable Difference (**JND**) for temporal discrimination (Stevens, 2000: 228-9) in the post-test. In the delayed test, the temporal gap between the mean VOT of L2 and L1 for the voiced plosives is smaller than the **JND**, however.

On the other hand, for Listen Only learners, the temporal gap between the mean VOT of L2 and L1 on the one hand and L2 and TL on the other for the voiceless and voiced plosives surpass the **JND** in the post-test and delayed test.

For Traditional learners, the temporal gap between the mean VOT of L2 and L1 on the one hand and L2 and TL on the other for the voiceless and voiced plosives surpass the **JND** in the post-test. In the delayed test, the temporal gap between the mean VOT of L2 and L1 for the voiceless plosives is however, smaller than the **JND** for temporal discrimination.

Thus the Listen Only condition was the only condition that managed to establish two L2 VOT categories for voiced and voiceless plosives that are beyond the **JND** for temporal discrimination and have managed to maintain this distance ten weeks after the training.

Thus, it can be said that despite the L2 VOT categories for the voiceless and voiced plosives resembling both L1 and TL categories, the fact that mean VOT values for either category are intermediate in both experimental conditions indicates some sort of learning that took place during the training and that is maintained after ten weeks of the end of training.

A different observation can be made for the Traditional condition. Again despite the mean L2 VOT for voiceless and voiced plosives being statistically indistinguishable from either L1 or that of the teacher, it is not intermediate. This could be due to the effect of relying on phonological memory during the delayed repetition task. The effect of this task could not be incorporated in the models due to the lack of corresponding matches in the Target language tokens for either DigLin or the Traditional teacher. Studies show that L2 learners, including those with very low proficiency or minimal exposure have the ability to distinguish native from foreign accent (Major, 2007; Neufeld, 1980; O'Brien, 2014).

To sum up, the results above show that each experimental condition managed to produce L2 VOT values that are statistically indistinguishable from those of the TL for voiced plosives. Both experimental conditions managed to maintain this performance in the delayed test. However, the Listen and Speak condition's mean L2 VOT for voiced plosives was closer to that of the TL than their L1 during the post-test but not during the delayed post-test. The Listen Only learners' mean L2 VOT for voiced plosives was closer to that of their L1 than the TL in both tests.

7.5.2 *Vowel-onset f0*

The findings indicate that vowel-onset f0 perturbations are lower after voiceless plosives, a phenomenon observed in a broad range of languages (Hombert et al., 1979; House and Fairbanks, 1953; Kirby and Ladd, 2015; Lehiste and Peterson, 1961) and is referred to as the onset voicing effect (Kirby and Ladd, 2015). An exception to this is the Traditional learners who exhibited a lower vowel-onset f0 following voiceless plosives in L1 (male and female speakers) and L2 (female speakers) during the post-test and lower vowel-onset f0 following voiceless plosives in L2 (male speakers only) during the delayed test in comparison to vowel-onset f0 following a voiced plosive.

In the seminal study by Kingston and Diehl (1994), it is argued that vowel-onset f0 is considered the most reliable correlate of phonological voicing contrast across languages, though not the primary voicing cue. A study by Dmitrieva, Llanos, Shultz, and Francis (2015) explored the co-variation between vowel-onset f0 and VOT in English and Spanish. They chose two languages which have different VOT implementations. That is for voiceless and voiced plosives English utilises a long VOT (aspirated) and short VOT respectively, whilst Spanish utilises a short VOT and lead VOT respectively. Their results demonstrate that vowel onset f0 and VOT had a significant inter-voicing category co-dependency – irrespective of within-category phonetic variability[45] – for both English and Spanish. This shows that the implementation of vowel-onset f0 depends on the phonological category. However, in the results above despite f0 perturbations following a voiceless plosive being (almost) consistently higher, neither of the language groups showed categorical contrast that is statistically significant. The `afex` model comparison function from the afex package (Singmann et al., 2019) in `R` is performed to test these results in the Listen and Speak condition in the post-test:

---

[45] The production of voiced plosives in their English data varied between short lag and lead voicing

Table 7.29 Mixed Model Anova Table (Type 3 tests, LRT-method) for Listen and Speak PT

Effect	df	Chisq	p.value
Language	2	12.01	**.002**
Gender	1	26.33	**<.0001**
Voice	1	3.46 +	.06
Language:Gender	2	39.01	**<.0001**
Language:Voice	2	0.99	.61
Gender:Voice	1	0.50	.48
Language:Gender:Voice	2	2.32	.31

Results of the model comparison confirm that the effect of Voice does not reach statistical significance ($p = 0.06$), whereas Language, Gender and the interaction terms between Language and Gender have a significant impact on f0 perturbations. The findings also show that within each of the voice categories, L2 vowel-onset f0 perturbations are statistically non-distinguishable from either L1 or TL. An exception to this is the within-male L2 to TL comparisons. This difference is found in both experimental conditions and in both times of testing. As for the Traditional condition, such comparison is not possible due to having a female teacher only. Before going back to the comparisons of language, an explanation is offered with regards to this variation in performance as a result of Gender. At face value it may be assumed that the female learners behaved differently from male learners. However, I argue that this difference is not due to Gender but rather to age. You may recall that it was not possible to incorporate the effect of age in the model – as important as it is for measurements – due to the exaggerated predicted values of the TL language group and due to the large standard error.[46] It is also seen in the descriptive results section how vowel-onset f0 values of male learners did not vary considerably from those of the female learners. The significant effect of Gender displayed in table 7.29 is most likely exclusive to the TL language group, where there is a considerable difference in f0 values between the DigLin male and female speaker. Several studies have demonstrated the lack of gender effect on f0 measurements in children. Eguchi and Hirsh (1969) examined acoustic features of Australian vowels, including fundamental frequency, in prepubescent children in four age groups, that is children aged five, seven, nine and eleven. Their

---

[46] This is because age was confounded with language group. The TL data come from adult speakers (DigLin speakers and the Traditional teacher), whereas the L1 and L2 data come from children. A high standard error – relative to the coefficient – means either that a) the coefficient is close to 0, or b) the coefficient is not well estimated, or some combination.

results indicate that there is no gender differences in f0 between males and females in either of the age groups. There is no consensus as to what age the gender distinction emerges (Lee, Potamianos, and Narayanan, 1999). A study by Busby and Plant (1995) indicates that the lack of a gender effect extends to children aged between 11 and 13. However, findings from Perry, Ohde, and Ashmead (2001) suggest that f0 differences are distinguished by gender after the age of 12. Differences in mean f0 values illustrated in table 7.15 are above the JND in pitch discrimination proposed as 1 Hz difference for complex tones with frequencies between 80 and 500 Hz (Stevens, 2000).

Chapter 8: Place contrast: Spectral Tilt Ahi-A23

8.1 Listen and Speak

8.1.1 *Descriptive statistics*

Table 8.1 shows Ahi-A23 values for the language groups L1 and L2 from the Listen and Speak condition and TL from the DigLin programme. Cells with a mean showing '—' indicate the unavailability of measurable tokens, whereas that in SD cells indicate that the value is based on a single speaker. This is because the target items in DigLin are produced by either the female or male speaker but not both. Overall, the table and figure show that the mean of Ahi-A23 of coronal plosives is higher than that for bilabial cognates within all of the vowel contexts supporting the findings of Stevens et al. (1999) and Suchato (2004a). Secondly, the figures show that the mean of Ahi-A23 values for adults is, as expected, consistently higher than that for the learners. This is due to difference in size of oral cavity between adults and children, with adults having a larger/longer vocal tract compared to children. Thirdly, variability as indicated by standard deviation is *generally* greater in the post-test compared to the delayed test. Moreover, variability is *generally* greater for coronals than bilabials. Fourthly, there is a *general* tendency for the learners, mean Ahi-A23 to increase with time[47]. This is likely due to the increased oral cavity as a function of age.

---

[47] Exceptions to this pattern are found in L2 means in the bilabial/back-high vowel context, L1 means in the bilabial/front-low vowel context, and L1 means in the coronal/front-low vowel context.

Table 8.1 Descriptive statistics for Ahi-A23 (dB) with mean and standard deviation (SD) for the Listen and Speak condition

			Post-test		Delayed test	
Vowel	Place	Language	Mean	SD	Mean	SD
Back-High	bilabial	L1	-6.1	6.3	-5.5	4.7
		L2	-4.2	6.3	-4.4	5.7
		TL	-5.6	6.5	-5.6	6.5
	coronal	L1	2.8	7.8	4.5	6.9
		L2	4.7	7.3	5.7	5.4
		TL	11.4	8.4	11.4	8.4
Back-Low	bilabial	L2	-3.1	8.0	-1.9	6.5
		TL	-18.0	—	-18.0	—
	coronal	L2	3.4	6.3	8.7	6.0
		TL	10.0	—	10.0	—
Front-High	bilabial	L1	-3.5	6.5	-1.8	7.8
		L2	-3.7	7.0	-2.4	4.0
		TL	-7.2	1.1	-7.2	1.1
	coronal	L1	7.7	7.6	8.7	6.8
		L2	7.0	6.9	8.4	7.0
		TL	10.0	8.1	10.0	8.1
Front-Low	bilabial	L1	0.1	5.3	-1.5	6.6
		L2	-4.7	7.2	2.2	6.4
		TL	5.3	—	5.3	—
	coronal	L1	7.7	7.5	5.0	6.0
		L2	4.9	8.7	9.2	8.9
		TL	18.4	2.0	18.4	2.0

Within the back-high vowel context, the mean Ahi-A23 of the L2 bilabial plosives is higher than that for L1 and TL in both tests. The mean Ahi-A23 of the L2 coronal plosives is intermediate between those of L1 and TL in both tests. Within the back-low vowel context, the mean Ahi-A23 of the L2 bilabial plosives is higher than that for TL in both tests, but the mean Ahi-A23 of the L2 coronal plosives is lower than that for TL coronal plosives also in both tests. Within the front-high vowel context, the mean Ahi-A23 of the L2 bilabial plosives is intermediate between those of the L1 and TL in both tests. However, for the coronal plosives, the mean Ahi-A23 of L2 is lower in both tests. Finally, within the front-low vowel context, the mean Ahi-A23 of L2 plosives is lower than either that for L1 or TL in the post-test. In the delayed test, it is rather intermediate between the two. For the coronal plosives, the mean Ahi-A23 for the L2 plosives is lower than either that for L1 or TL, but rather intermediate between the two in the delayed test.

8.1.2 *Post-test*

Model Specification

The main goal from this step of the analysis is to explore differences in Ahi-A23 values between the language groups L2, L1 and TL (Language). We want to account for potential speaker-related effect of Gender. Adding the effect of 'Age' yielded very large standard errors for the 'Language' effect. The model specification for this subset of the data will be applied for the subsequent subsets of the data to allow for more comparable outcomes. In terms of Ahi-A23 related effects, we want to control for Place of articulation and Vowel context. Interaction terms for Language, Place and Vowel are added as we want the model to capture variance in place of articulation and vowel context within each Language. Adding the effect of 'Task' created the following warning messages:

```
contrasts dropped from factor Language due to missing levels
```

The Language variable has three levels: L1, L2, and TL. When eliciting the L1 data, we used two tasks: picture-naming and read aloud. When eliciting the L2 data, we used three tasks: picture-naming reading aloud, and delayed repetition. However, TL data comes from the DigLin speakers' audio files in DigLin and/or the Traditional teacher using picture-naming and read aloud only. This creates an issue for the model as the missing levels (for example the delayed repetition task is missing in the L1 and TL data) from the Task variable do not allow the model to conduct the required comparison. Task, despite having a potential effect on the outcome, is not something the research question addresses. Since the research question is concerned mainly with differences across the training conditions and in differences between the L2 and L1, and L2 and the TL, these effects were prioritised over the effect of task within the model. This is also true of the subsequent models used in this chapter.

The model added by-speaker and by-item random intercepts. By-speaker random slopes for Language, Place, and Vowel would not allow the model to converge. Similarly, adding by-item random slopes for Language and Gender would not allow the model to converge even after decorrelating slopes from the intercept. Models in the spectral tilt analyses behaved differently with this respect. However, an effort is made to keep the models as uniform as possible for comparable results. To this end, a linear mixed-effects model with all of the fixed effects and random effects is built. The following model is considered:

```
Tilt ~ Language * Place * Vowel + Gender +
        (1 | speaker) + (1 | item)
```

Inferential statistics on spectral tilt in Listen and Speak condition during post-test

Table 8.2 shows the results of a linear mixed-effects model for Ahi-A23 by the learners of Listen and Speak condition during the post-test compared to Diglin. The reference for Language is ‘L2’, for Place ‘Bilabial’, for Vowel ‘Back-High’, and for Gender ‘Male’. In this model, the intercept represents mean Ahi-A23 values for an L2 Bilabial plosive followed by a Back-High vowel and produced by Male participants from Listen and Speak condition during the post-test. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients. The effects of Language, Place and Vowel are simple effects since they are involved in interaction terms. Therefore, differences in mean spectral tilt values will be dealt with in a follow-up post hoc test.

The LMM output in table 8.2, shows that the effect of Gender on the mean value of Ahi-A23 does not reach statistical significance (p >0.05).

Table 8.2 Output of a linear mixed-effects model for spectral tilt in Listen and Speak condition during post-test

term	β	S.E	df	*t*	(p)
(Intercept)	2.11	1.42	44.63	1.49	0.1442
LanguageL1	-1.58	1.33	60.82	-1.19	0.2398
LanguageTL	0.72	1.78	34.17	0.41	0.6863
PlaceCoronal	-6.81	1.14	38.50	-5.99	**5.67e-07**
VowelBack-Low	-1.34	1.36	35.51	-0.99	0.3302
VowelFront-High	-0.74	2.90	52.19	-0.26	0.7982
VowelFront-Low	-0.76	1.75	88.41	-0.44	0.6641
GenderFemale	-1.14	0.57	21.46	-2.01	0.0576
LanguageL1:PlaceCoronal	2.49	1.07	81.91	2.31	**0.0233**
LanguageTL:PlaceCoronal	0.53	1.59	24.95	0.33	0.7425
LanguageL1:VowelBack-Low	1.56	1.38	58.09	1.14	0.2607
LanguageTL:VowelBack-Low	-3.30	2.19	24.09	-1.50	0.1459
LanguageL1:VowelFront-High	0.23	2.74	96.50	0.08	0.9325
LanguageTL:VowelFront-High	5.22	3.93	28.24	1.33	0.1948
LanguageL1:VowelFront-Low	1.55	1.54	562.45	1.00	0.3156
PlaceCoronal:VowelBack-Low	0.97	1.36	36.15	0.71	0.4801
PlaceCoronal:VowelFront-High	-3.10	2.91	52.62	-1.07	0.2913
PlaceCoronal:VowelFront-Low	0.71	1.76	88.63	0.41	0.6862
LanguageL1:PlaceCoronal:VowelBack-Low	-1.07	1.38	58.92	-0.78	0.4387
LanguageTL:PlaceCoronal:VowelBack-Low	0.83	2.19	24.13	0.38	0.7070
LanguageL1:PlaceCoronal:VowelFront-High	3.71	2.75	97.79	1.35	0.1802
LanguageTL:PlaceCoronal:VowelFront-High	-1.91	3.93	28.25	-0.49	0.6312
LanguageL1:PlaceCoronal:VowelFront-Low	-1.22	1.55	566.49	-0.79	0.4317

We are interested in whether a contrast in Place of articulation is made. To check this, post hoc pairwise comparisons between language groups were carried out. To do this for each Place of articulation by Vowel, the function `emmeans` (Lenth, 2019) in `R` is applied. The following code is used:

```
emmeans(mdl, pairwise ~ Language * Place | Vowel)
```

The pairwise function allows for pairwise comparisons specified by the formula, that is pairwise comparisons by Language, Place and Vowel. The Tukey adjustment method for P values is used for comparing a family of six estimates. Table I.1 in Appendix I indicates that within each of the vowel contexts and within either place of articulation, none of the contrasts between L2 and L1 nor L2 and TL are statistically significant.

8.1.3 *Delayed test*

Inferential statistics on spectral tilt in Listen and Speak condition during delayed test

The coefficients in table 8.3 are the results of the model presented above. In this model, the intercept represents mean Ahi-A23 values for an L2 Bilabial plosive followed by a Back-High vowel and produced by Male participants from Listen and Speak condition during the delayed test. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients. The effects of Language, Place and Vowel are simple effects since they are involved in interaction terms. Therefore, differences in mean spectral tilt values will be dealt with in a follow-up post hoc test. The LMM output in table 8.3, shows that there is no significant effect of Gender (p >0.4) on the mean value of Ahi-A23.

Table 8.3 Output of a linear mixed-effects model for spectral tilt in Listen and Speak delayed test

term	β	S.E	df	*t*	(p)
(Intercept)	3.42	1.32	31.25	2.59	**0.0143**
LanguageL1	-0.42	1.24	44.47	-0.34	0.7340
LanguageTL	0.85	1.66	24.50	0.51	0.6137
PlaceCoronal	-6.28	1.07	29.65	-5.88	**<.0001**
VowelBack-Low	-2.30	1.27	28.29	-1.81	0.0809
VowelFront-High	-0.20	2.69	38.65	-0.07	0.9421
VowelFront-Low	-0.80	1.63	60.77	-0.49	0.6254
GenderFemale	0.47	0.60	13.85	0.77	0.4539
LanguageL1:PlaceCoronal	1.54	1.02	62.64	1.51	0.1369
LanguageTL:PlaceCoronal	1.32	1.50	18.54	0.88	0.3894
LanguageL1:VowelBack-Low	0.11	1.31	46.67	0.08	0.9355
LanguageTL:VowelBack-Low	-2.44	2.07	18.29	-1.18	0.2535
LanguageL1:VowelFront-High	0.07	2.58	71.01	0.03	0.9785
LanguageTL:VowelFront-High	8.23	3.74	21.46	2.20	**0.0388**
LanguageL1:VowelFront-Low	0.77	1.42	371.15	0.54	0.5890
PlaceCoronal:VowelBack-Low	-0.03	1.27	28.89	-0.03	0.9787
PlaceCoronal:VowelFront-High	-1.85	2.70	39.35	-0.68	0.4976
PlaceCoronal:VowelFront-Low	-0.28	1.63	61.45	-0.17	0.8661
LanguageL1:PlaceCoronal:VowelBack-Low	-0.31	1.32	47.69	-0.23	0.8172
LanguageTL:PlaceCoronal:VowelBack-Low	-0.04	2.07	18.29	-0.02	0.9856
LanguageL1:PlaceCoronal:VowelFront-High	1.90	2.59	72.35	0.73	0.4656
LanguageTL:PlaceCoronal:VowelFront-High	0.51	3.74	21.49	0.14	0.8927
LanguageL1:PlaceCoronal:VowelFront-Low	-0.77	1.43	374.26	-0.54	0.5885

The formula specified for the post hoc pairwise comparison here follows the same as that for the post-test for this subset of the data. The same formula will applied to the subsequent subsets of the data. Table I.1 in Appendix I indicates that within each of the vowel contexts and within either place of articulation, none of the contrasts between L2 and L1 nor L2 and TL are statistically significant.

8.2 Listen Only

8.2.1 *Descriptive statistics*

Table 8.4 show Ahi-A23 values for the language groups L1 and L2 from the Listen Only condition and TL from the DigLin programme. Cells with an SD showing '—'

indicate that the mean value is based on a single speaker. This is because the target items in DigLin are produced by either the female or male speaker but not both. Overall, the table and figure show that the mean of Ahi-A23 of coronal plosives is higher than that for bilabial cognates within all of the vowel contexts supporting the findings of Stevens et al. (1999) and Suchato (2004a) and in line with data observed in the Listen and Speak condition. Secondly, the figures show that the mean of Ahi-A23 values for children is generally, but not consistently, lower than that for the adults. This is due to difference in size of oral cavity between adults and children, with adults having a larger/longer vocal tract compared to children. There are a number of exceptions to this almost always occurring in the bilabial place of articulation. Thirdly, variability as indicated by standard deviation is *generally* greater in the post-test compared to the delayed test. Fourthly, there is a tendency for the learners' mean Ahi-A23 to increase with time. This is likely due to the increased oral cavity as a function of age. Within the back-high vowel context, the mean Ahi-A23 of the L2 bilabial plosives is intermediate between that for L1 and TL in the post test, but higher than both L1 and TL in the delayed test. The mean Ahi-A23 of the L2 coronal plosives is intermediate between those of L1 and TL in both tests. Within the back-low vowel context, the mean Ahi-A23 of the L2 bilabial plosives is higher than that for TL in both tests, but the mean Ahi-A23 of the L2 coronal plosives is higher than that for TL coronal plosives in both tests. Within the front-high vowel context, the mean Ahi-A23 of the L2 bilabial plosives is higher than either that of the L1 or TL in the post-test, but intermediate between those of the L1 and TL in the delayed test. However, for the coronal plosives, the mean Ahi-A23 of L2 is smaller than L1 and TL in post-test, but intermediate between their L1 and the TL. Finally, within the front-low vowel context, the mean Ahi-A23 of L2 plosives is lower than either that for L1 or TL in the post-test. In the delayed test, it is rather intermediate between the two. For the coronal plosives, the mean Ahi-A23 for the L2 plosives is lower than either that for L1 or TL, but rather intermediate between the two in the delayed test.

Table 8.4 Descriptive statistics for Ahi-A23 (dB) with mean and standard deviation (SD) for the Listen Only condition

			Post-test		Delayed test	
Vowel	Place	Language	Mean	SD	Mean	SD
Back-High	bilabial	L1	-7.5	6.2	-4.6	8.0
		L2	-6.3	7.4	-2.1	5.6
		TL	-5.6	6.5	-5.6	6.5
	coronal	L1	2.8	6.5	5.6	6.2
		L2	4.1	6.5	7.3	5.7
		TL	11.4	8.4	11.4	8.4
Back-Low	bilabial	L2	-4.2	6.3	-1.2	4.2
		TL	-18.0	—	-18.0	—
	coronal	L2	4.6	6.5	5.8	4.5
		TL	10.0	—	10.0	—
Front-High	bilabial	L1	-4.6	6.1	-1.1	5.5
		L2	-3.6	4.0	-1.6	5.6
		TL	-7.2	1.1	-7.2	1.1
	coronal	L1	7.1	7.8	12.7	6.9
		L2	4.9	6.2	10.4	8.0
		TL	10.0	8.1	10.0	8.1
Front-Low	bilabial	L1	-1.6	6.3	-0.6	6.2
		L2	-2.1	4.5	1.6	7.0
		TL	5.3	—	5.3	—
	coronal	L1	4.5	8.3	6.3	7.1
		L2	3.3	6.9	7.8	6.7
		TL	18.4	2.0	18.4	2.0

8.2.2 *Post-test*

Inferential statistics on spectral tilt in Listen Only condition during post-test

In this model, the intercept represents mean Ahi-A23 values for an L2 Bilabial plosive followed by a Back-High vowel and produced by Male participants from Listen Only condition during the post-test. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients. The effects of Language, Place and Vowel are simple effects since they are involved in interaction terms. Therefore, differences in mean spectral tilt values will be dealt with in a follow-up post hoc test. The LMM output in table 8.3, shows that there is no significant effect of Gender (p >0.1) on the mean value of Ahi-A23.

Table 8.5 Output of a linear mixed-effects model for spectral tilt in Listen Only post-test

term	β	S.E	df	*t*	(p)
(Intercept)	1.79	1.22	49.00	1.46	0.14988
LanguageL1	-1.41	1.19	64.89	-1.18	0.24275
LanguageTL	0.21	1.47	61.55	0.14	0.88741
PlaceCoronal	-6.68	0.96	100.76	-6.95	**3.62e-10**
VowelBack-Low	-1.85	1.13	86.36	-1.64	0.10462
VowelFront-High	-0.43	2.50	130.86	-0.17	0.86481
VowelFront-Low	-0.66	1.57	202.70	-0.42	0.67507
GenderFemale	-0.85	0.54	18.64	-1.57	0.13344
LanguageL1:PlaceCoronal	2.62	0.97	221.77	2.71	**0.00732**
LanguageTL:PlaceCoronal	0.42	1.29	58.15	0.33	0.74470
LanguageL1:VowelBack-Low	0.48	1.20	157.39	0.40	0.68913
LanguageTL:VowelBack-Low	-2.36	1.76	54.05	-1.34	0.18574
LanguageL1:VowelFront-High	0.73	2.47	244.94	0.29	0.76931
LanguageTL:VowelFront-High	5.76	3.23	64.84	1.78	0.07900
LanguageL1:VowelFront-Low	1.02	1.51	506.83	0.68	0.49933
PlaceCoronal:VowelBack-Low	0.57	1.13	88.05	0.50	0.61682
PlaceCoronal:VowelFront-High	-3.59	2.51	132.80	-1.43	0.15462
PlaceCoronal:VowelFront-Low	0.45	1.57	203.96	0.28	0.77715
LanguageL1:PlaceCoronal:VowelBack-Low	-1.38	1.21	159.65	-1.14	0.25485
LanguageTL:PlaceCoronal:VowelBack-Low	0.55	1.76	53.99	0.31	0.75425
LanguageL1:PlaceCoronal:VowelFront-High	2.98	2.48	248.54	1.20	0.23015
LanguageTL:PlaceCoronal:VowelFront-High	-1.22	3.23	64.84	-0.38	0.70685
LanguageL1:PlaceCoronal:VowelFront-Low	-0.74	1.52	510.05	-0.49	0.62576

The results of the post hoc pairwise comparisons (using the same formula for this subset of the data as the previous one) indicate that within each of the vowel contexts and within either place of articulation, none of the contrasts between L2 and L1 nor L2 and TL are statistically significant with the exception of an overall statistically significant decrease ($p < .05$) by an average of 15 dB in the mean Ahi-A23 value of the L2 coronal plosives compared to their TL counterparts within the front-low vowel context. See table I.2 in Appendix I.

8.2.3 *Delayed test*

Inferential statistics on spectral tilt in Listen Only condition during delayed test

In this model, the intercept represents mean Ahi-A23 values for an L2 Bilabial plosive followed by a Back-High vowel and produced by Male participants from Listen Only condition during the delayed test. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients. The effects of Language, Place and Vowel are simple effects since they are involved in interaction terms. Therefore, differences in mean spectral tilt values will be dealt with in a follow-up post hoc test. The LMM output in table 8.6, shows that there is no significant effect of Gender (p >0.4) on the mean value of Ahi-A23.

Table 8.6 Output of a linear mixed-effects model for spectral tilt in Listen Only delayed test

term	β	S.E	df	t	(p)
(Intercept)	4.22	1.16	44.37	3.63	**0.00073**
LanguageL1	-0.82	1.15	56.34	-0.71	0.47950
LanguageTL	2.03	1.35	102.19	1.51	0.13482
PlaceCoronal	-6.82	0.86	509.70	-7.90	**1.67e-14**
VowelBack-Low	-2.20	1.01	503.51	-2.18	**0.02977**
VowelFront-High	0.47	2.27	501.07	0.21	0.83590
VowelFront-Low	-0.36	1.44	501.35	-0.25	0.80427
GenderFemale	-0.45	0.55	18.13	-0.81	0.42737
LanguageL1:PlaceCoronal	2.43	0.89	509.41	2.73	**0.00655**
LanguageTL:PlaceCoronal	0.31	1.14	501.01	0.28	0.78342
LanguageL1:VowelBack-Low	1.40	1.09	503.44	1.28	0.19973
LanguageTL:VowelBack-Low	-3.59	1.55	501.47	-2.31	**0.02105**
LanguageL1:VowelFront-High	-1.50	2.28	501.17	-0.66	0.50983
LanguageTL:VowelFront-High	9.00	2.86	501.62	3.15	**0.00174**
LanguageL1:VowelFront-Low	1.25	1.44	501.35	0.87	0.38416
PlaceCoronal:VowelBack-Low	0.79	1.01	511.56	0.78	0.43374
PlaceCoronal:VowelFront-High	-2.63	2.28	508.12	-1.15	0.24928
PlaceCoronal:VowelFront-Low	-0.33	1.44	504.53	-0.23	0.82022
LanguageL1:PlaceCoronal:VowelBack-Low	-1.02	1.09	510.61	-0.94	0.34978
LanguageTL:PlaceCoronal:VowelBack-Low	0.69	1.55	501.34	0.44	0.65667
LanguageL1:PlaceCoronal:VowelFront-High	3.85	2.29	508.08	1.68	0.09337
LanguageTL:PlaceCoronal:VowelFront-High	-1.61	2.86	501.47	-0.56	0.57425
LanguageL1:PlaceCoronal:VowelFront-Low	-1.28	1.44	504.53	-0.88	0.37679

The results of the post hoc pairwise comparisons as shown in table I.2 in Appendix I indicate that within each of the vowel contexts and within either place of articulation, none of the contrasts between L2 and L1 nor L2 and TL are statistically significant.

8.3 Traditional

8.3.1 *Descriptive statistics*

Table 8.7 show Ahi-A23 values for the language groups L1 and L2 from the Traditional condition and TL from their teacher. Overall, the table and figure show that the mean of Ahi-A23 of coronal plosives is higher than that for bilabial cognates within all of the vowel contexts supporting the findings of Stevens et al. (1999) and Suchato (2004a) and in line with data observed in the experimental conditions. Secondly, the figures show that the mean of Ahi-A23 values for children is consistently, higher than that for the adults. This is surprising given the difference in size of oral cavity between adults and children, with adults having a larger/longer vocal tract compared to children. Thirdly, variability as indicated by standard deviation is *generally* greater in the post-test compared to the delayed test. Fourthly, there is a tendency for the learners' mean Ahi-A23 to increase with time[48]. As explained before, this is likely due to the increased oral cavity as a function of age. Within the back-high vowel context, the mean Ahi-A23 of the L2 bilabial plosives is higher than that for L1 and TL in both tests. The same is true for the mean Ahi-A23 of the L2 coronal plosives also in both tests. Within the back-low vowel context, the mean Ahi-A23 of the bilabial and coronal plosives is only estimable in the L2 language group as no such tokens were produced for this context in either the L1 or TL data. Within the front-high vowel context, the mean Ahi-A23 of the L2 bilabial plosives is intermediate between that for L1 and TL in both tests. However, for the coronal place of articulation, the mean Ahi-A23 of the L2 is higher than either that for L1 or TL in both tests. Finally, within the front-low vowel context, the mean Ahi-A23 of the L2 bilabial plosives is intermediate between that for L1 and TL in the post-test. In the delayed test, it is rather higher than either that for L1 or L2. As for coronal plosives, the mean Ahi-A23 of the L2 is higher than either that for L1 or TL.

---

[48] There are three exceptions to this, all of which in are within the bilabial place of articulation.

Table 8.7 Descriptive statistics for Ahi-A23 (dB) with mean and standard deviation (SD) for the Traditional condition

			Post-test		Delayed test	
Vowel	Place	Language	Mean	SD	Mean	SD
Back-High	bilabial	L1	-6.3	9.0	-6.4	6.3
		L2	-6.1	5.4	-4.3	5.8
		TL	-10.0	4.9	-10.0	4.9
	coronal	L1	0.4	6.6	2.3	6.0
		L2	3.0	7.1	3.0	6.0
		TL	-1.7	8.8	-1.7	8.8
Back-Low	bilabial	L2	-4.1	4.0	-3.7	5.7
	coronal	L2	3.9	5.4	6.0	6.7
Front-High	bilabial	L1	-1.0	7.2	-2.3	4.9
		L2	-2.0	6.7	-3.2	5.5
		TL	-11.2	6.6	-11.2	6.6
	coronal	L1	2.8	6.2	7.7	7.3
		L2	7.8	8.1	9.2	7.9
		TL	-0.2	11.0	-0.2	11.0
Front-Low	bilabial	L1	-3.1	5.5	-2.8	5.2
		L2	-3.8	4.1	-1.2	5.6
		TL	-9.9	2.5	-10.0	2.5
	coronal	L1	2.7	5.4	4.2	5.8
		L2	5.1	8.9	8.6	7.9
		TL	-6.7	6.9	-6.7	6.9

8.3.2 *Post-test*

Inferential statistics on spectral tilt in Traditional condition during post-test

In this model, the intercept represents mean Ahi-A23 values for an L2 Bilabial plosive followed by a Back-High vowel and produced by Male participants from the Traditional condition during the post-test. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients.

The effects of Language, Place and Vowel are simple effects since they are involved in interaction terms. Therefore, differences in mean spectral tilt values will be dealt with in a follow-up post hoc test.

The LMM output in table 8.8, shows that there is no significant effect of Gender (p >0.3) on the mean value of Ahi-A23.

Table 8.8 Output of a linear mixed-effects model for spectral tilt in Traditional post-test

term	β	S.E	df	*t*	(p)
(Intercept)	-3.62	1.60	32.99	-2.27	**0.030047**
LanguageL1	4.19	1.52	34.94	2.75	**0.009406**
LanguageTL	1.05	1.84	38.04	0.57	0.570872
PlaceCoronal	-4.33	1.09	41.59	-3.98	**0.000274**
VowelBack-Low	-0.46	1.28	56.66	-0.36	0.719662
VowelFront-High	-2.86	2.74	65.35	-1.05	0.299259
VowelFront-Low	2.47	0.96	10.82	2.58	**0.025805**
GenderFemale	-0.82	0.84	15.75	-0.98	0.339815
LanguageL1:PlaceCoronal	-0.35	1.03	89.06	-0.34	0.734885
LanguageTL:PlaceCoronal	1.03	1.43	24.87	0.72	0.479889
LanguageL1:VowelBack-Low	-2.47	1.35	87.00	-1.83	0.071101
LanguageTL:VowelBack-Low	-0.12	1.85	25.08	-0.07	0.947927
LanguageL1:VowelFront-High	3.12	2.55	105.80	1.22	0.224445
LanguageTL:VowelFront-High	-1.37	3.42	26.68	-0.40	0.691481
PlaceCoronal:VowelBack-Low	0.56	1.28	57.10	0.44	0.662796
PlaceCoronal:VowelFront-High	-1.05	2.73	65.15	-0.38	0.702360
PlaceCoronal:VowelFront-Low	-0.58	0.96	10.92	-0.61	0.556572
LanguageL1:PlaceCoronal:VowelBack-Low	-0.14	1.35	85.01	-0.10	0.917388
LanguageTL:PlaceCoronal:VowelBack-Low	-0.22	1.84	24.40	-0.12	0.904391
LanguageL1:PlaceCoronal:VowelFront-High	1.79	2.54	105.56	0.70	0.482730
LanguageTL:PlaceCoronal:VowelFront-High	0.84	3.39	26.21	0.25	0.805216

The results of the post hoc pairwise comparisons as shown in table I.3, Appendix I indicate that within each of the vowel contexts and within either place of articulation, none of the contrasts between L2 and L1 nor L2 and TL are statistically significant.

8.3.3 *Delayed test*

Inferential statistics on spectral tilt in Traditional condition during delayed test

In this model, the intercept represents mean Ahi-A23 values for an L2 Bilabial plosive followed by a Back-High vowel and produced by Male participants from Traditional condition during the delayed test. All fixed effects are sum coded to allow for a feasible interpretation of the coefficients. The effects of Language, Place and Vowel are simple effects since they are involved in interaction terms. Therefore, differences in mean spectral tilt values will be dealt with in a follow-up post hoc test.

The LMM output in table 8.9, shows that there is no significant effect of Gender (p >0.8) on the mean value of Ahi-A23.

Table 8.9 Output of a linear mixed-effects model for spectral tilt in Traditional delayed test

term	β	S.E	df	t	(p)
(Intercept)	-1.57	2.06	84.33	-0.77	0.4462
LanguageL1	3.44	2.08	87.05	1.65	0.1019
LanguageTL	2.66	1.04	23.87	2.56	**0.0172**
PlaceCoronal	-4.22	0.90	94.16	-4.69	**9.34e-06**
VowelBack-Low	-1.30	2.07	86.02	-0.63	0.5304
VowelFront-High	-0.61	5.82	84.71	-0.10	0.9168
VowelFront-Low	1.56	1.86	57.04	0.84	0.4047
GenderFemale	-0.07	0.45	13.94	-0.15	0.8798
LanguageL1:PlaceCoronal	-0.72	0.88	187.14	-0.82	0.4156
LanguageTL:PlaceCoronal	0.65	0.89	18.21	0.73	0.4747
LanguageL1:VowelBack-Low	-1.21	2.10	118.22	-0.58	0.5663
LanguageTL:VowelBack-Low	-1.87	1.19	16.31	-1.57	0.1355
LanguageL1:VowelFront-High	-0.23	5.86	77.22	-0.04	0.9684
LanguageTL:VowelFront-High	2.65	2.25	21.30	1.18	0.2527
LanguageL1:VowelFront-Low	-0.36	2.01	54.64	-0.18	0.8591
PlaceCoronal:VowelBack-Low	0.18	1.07	99.63	0.16	0.8693
PlaceCoronal:VowelFront-High	0.30	2.37	144.78	0.12	0.9009
PlaceCoronal:VowelFront-Low	-1.40	0.72	19.39	-1.94	0.0677
LanguageL1:PlaceCoronal:VowelBack-Low	1.08	1.18	157.18	0.91	0.3629
LanguageTL:PlaceCoronal:VowelBack-Low	-0.92	1.19	16.30	-0.77	0.4496
LanguageL1:PlaceCoronal:VowelFront-High	-0.27	2.29	217.07	-0.12	0.9050
LanguageTL:PlaceCoronal:VowelFront-High	1.79	2.25	21.27	0.79	0.4365

The results of the post hoc pairwise comparisons as shown in table I.3, Appendix I indicate that within each of the vowel contexts and within either place of articulation, none of the contrasts between L2 and L1 nor L2 and TL are statistically significant with the exception of an overall statistically significant increase (p <.03) by an average of 16 dB in the mean Ahi-A23 value of the L2 coronal plosives compared to their TL counterparts within the front-low vowel context.

8.4 By-Condition results

Model specification

The main goal from this step of the analysis is to explore differences in Ahi-A23 values between training conditions 'Instruction' and time of test 'Test', the language groups L1 and L2. We want to account for potential speaker-related effects of Gender and Age. We also want to control for elicitation task effect. In terms of spectral shape related effects, we want to control for Place of articulation, Vowel and potentially Age. Interaction terms for Instruction, Language, Place and Vowel are added as we want the model to capture variance in Place and Vowel context within each Language within each training condition.

The model added by-speaker and by-item random intercepts. By-speaker random slopes for interaction terms of Language, Place and Vowel are added. By-item random slopes for Language and Task are added. To this end, a linear mixed-effects model with all of the fixed effects and random effects (intercepts and slopes) is built. The following model is considered:

```
Tilt ~ Instruction * Language * Place + Vowel + Gender + Task +
        Test + Age + (1 + Language * Place + Test | speaker) +
        (1 | item)
```

Inferential statistics on Ahi-A23 in all training conditions

Table 8.10 shows the results of a linear mixed-effects model for Ahi-A23 values for all training conditions in both tests. The reference for Instruction is 'Listen and Speak', for

Language ‘L1’, for Place ‘Bilabial’, for Vowel, ‘Back-High’, for Gender ‘Male’, for Task ‘Delayed repetition’, and for Test ‘Post-test’. In this model, the intercept represents mean Ahi-A23 values for bilabial plosives preceding a Back-High vowel, produced by Male participants from the Listen and Speak condition in the Delayed repetition task during the Post-test. All fixed effects are sum coded. The effect of Age is centred.

Table 8.10 Output of a linear mixed-effects model for spectral tilt in Traditional delayed test

term	β	S.E	df	*t*	(p)
(Intercept)	1.25	0.39	39.58	3.16	**0.002981**
InstructionListen Only	0.32	0.37	52.73	0.86	0.393080
InstructionTraditional	0.64	0.36	48.18	1.78	0.081980
LanguageL2	-0.61	0.42	63.77	-1.44	0.153563
PlaceCoronal	-4.61	0.35	29.05	-13.36	**6.25e-14**
VowelBack-Low	-1.61	0.40	29.95	-4.01	**0.000373**
VowelFront-High	-0.35	0.45	181.58	-0.78	0.434188
VowelFront-Low	1.40	0.42	28.01	3.36	**0.002280**
GenderF	-0.34	0.26	46.57	-1.31	0.197798
TaskPicture-naming	-0.35	0.40	2666.83	-0.89	0.374252
TaskRead-aloud	0.20	0.23	2723.66	0.89	0.371704
TestDelayed test	-0.83	0.24	62.98	-3.43	**0.001060**
Age	0.09	0.08	46.60	1.16	0.252105
InstructionListen Only:LanguageL2	0.20	0.22	43.38	0.89	0.377966
InstructionTraditional:LanguageL2	0.30	0.22	40.29	1.34	0.186621
InstructionListen Only:PlaceCoronal	-0.13	0.31	51.57	-0.41	0.683549
InstructionTraditional:PlaceCoronal	-0.02	0.31	48.89	-0.07	0.948095
LanguageL2:PlaceCoronal	0.10	0.29	16.96	0.36	0.726854
InstructionListen Only:LanguageL2:PlaceCoronal	0.08	0.17	162.06	0.45	0.656242
InstructionTraditional:LanguageL2:PlaceCoronal	-0.37	0.17	146.20	-2.16	**0.032091**

The effects of Instruction, Language, and Place are simple effects since they are involved in interaction terms. The two- and three-way interaction terms are also simple effects. Therefore, differences in mean Ahi-A23 values for these effects will be dealt with in a follow-up post hoc test.

In terms of the effects of Vowel, the LMM output in table 8.10 shows that there is an overall statistically significant decrease (p <.001) by an average of 2 dB in the mean Ahi-A23 value between plosives followed by a Back-Low vowels compared to the Back-high context. There is also an overall statistically significant increase (p <.01) by an average of 1 dB in the mean Ahi-A23 value between plosives in the Front-Low context compared to the Back-high context. However, the difference in Ahi-A23 between the Front-High and Back-High contexts is not statistical. As for Gender, the table shows no statistical difference in the mean spectral tilt values between males and females (p >0.1). The table also shows no statistical difference in the mean Ahi-A23 values of plosives produced in the Picture-naming task compared to those in the delayed repetition task (p >0.3). There is no statistical difference in the mean Ahi-A23 values of plosives produced in the read aloud task compared to those in the delayed repetition task (p >0.3) either. However, the model coefficients indicate that there is an overall statistically significant decrease (p <.01) albeit by an average of less than 1 dB in the mean Ahi-A23 values of plosives in the delayed test compared to those in the post-test. For Age, the table shows that there is no statistical difference in mean Ahi-A23 value as a function of Age.

To find out whether mean spectral tilt values for each Instruction were significantly different, the function `emmeans` (Lenth, 2019) in `R` is applied. The following code is used:

```
emmeans(mdl, pairwise ~ Language * Instruction | Place,
                                    lmer.df = 'satterthwaite')
```

Results are averaged over the levels of Vowel, Gender, Task, and Test. Degrees-of-freedom are calculated using the Satterthwaite method using confidence intervals of 0.95. The pairwise function allows for pairwise comparisons specified by the formula, that is pairwise comparisons by Language and Instruction[49] for each place of articulation. The Tukey adjustment method for P values is used for comparing a family of six estimates.

Figure 8.1 shows model-driven predictions for Ahi-A23 for both L1 and L2 in bilabial and coronal plosives in each training condition. Traditional condition illustrates the most noticeable difference between L1 and L2 values for either place of articulation, but most notably within the coronal plosives albeit such differences are not statistical.

---

[49] See table I.4 for results of pairwise comparisons.

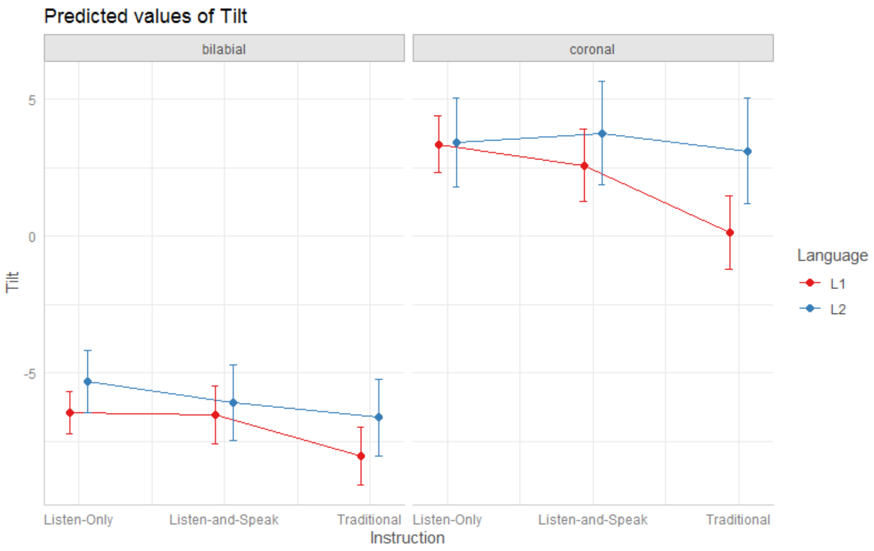


Figure 8.1 Predicted Estimated Marginal Means and marginal error for Ahi-A23 by Instruction, Place of articulation and Language

The figure shows an interaction between experimental instruction and place of articulation whereby, Listen Only show a bigger difference in spectral tilt within bilabial plosives but minimal difference in coronal plosives. Listen and Speak on the other hand demonstrate the opposite pattern. Nonetheless, the results of post hoc pairwise comparisons in table I.4 reveal that none of the differences are statistically significant.[50]

8.5 Summary and discussion

The Ahi-A23 data does not lend support to HYPOTHESIS ONE, which states that the Listen and Speak condition will result in the most target-like pronunciations/least L1 like productions followed by the Listen Only, leaving the Traditional Teaching condition with the relatively least target-like pronunciations/ most L1-like productions. The results show a lack of statistical contrast between conditions within either the native language or the interlanguage Ahi-A23 data. None of the training conditions managed to establish an L2 category that is distinct from that of their L1 for either place of articulation within each vowel context. However, the L2 Ahi-A23 values for either place of articulation within each vowel context in each of the training conditions does not prove to be different from that of their training input.

---

[50] See Appendix I, table I.4.

Despite the lack of such statistical contrast, the spectral shape for bilabials shows that Listen and Speak was the condition that mostly assimilated L1 and L2 values compared to the other two training conditions, which illustrate a rather small difference between the native and target values. As for the spectral shape of coronal plosives, although the Traditional condition illustrates the largest difference between L1 and L2 mean values, their L2 values are not notably different compared to those by the experimental conditions. It seems that the Traditional condition learners, despite sharing the same L1 language background with learners from the experimental conditions, their L1 values seem to behave rather differently (though not statistically) from the experimental conditions. This could be due to several possibilities, of which the fact that they are the only homogeneous group of learners in terms of residential area[51]. The participants from the experimental groups are of mixed residential areas though mainly from CENTRAL MISRATA. The lack of studies concerning within-region phonetic variation does not allow us to make conclusive comments with this regard. One rather remote possibility is that the English input received by the experimental groups may have affected the experimental learners' L1 categories. It has been argued that a merged L2 category affects the perception and production of L1 speech sounds (Flege, 1987d; Flege, Munro, and MacKay, 1995b; Flege, MacKay, and Meador, 1999; MacKay et al., 2001). Such a merged category is proposed to be the result of *equivalence classification* proposed by Flege's (1995) SLM model. The effect of equivalence classification blocks the formation of a new L2 phonetic category, albeit it does not prevent phonetic learning. It has been proposed to have a bidirectional effect impacting both L2 – in that target sounds will never be perceived and/or produced as monolingual target natives – and L1 – in that it will not be perceived and/or produced as monolingual natives of the learners even with increased experience of the target language (Flege, Munro, et al., 1995b). Although the training period is rather short for such an assumption, it cannot be ruled out entirely. No study thus far instrumentally investigated the rate and/or time required before a merged category is formed after exposure to a second language. In terms of the Traditional condition learners having comparable L2 values to those of the experimental conditions, several aspects should be kept in mind. First, there is a chance that the Traditional teacher's L2 English

---

[51] An area called QASER AHMAD four kilometres away from Central Misrata.

values are not statistically different from the native English values displayed by DigLin[52]. There is also a chance that the Traditional condition learners were relying on their phonological memory in the delayed repetition task despite the presence of an intervening phrase. This is presumed based on the fact that three different assistant researchers carried out the data collection for each training condition and that it is possible they had differences in the way they handled this task. On a relevant note, there is a slim chance that participants' from the Traditional condition were able to recognise the material in the delayed repetition task as native compared to the accented input they received from their teacher during the training. However, recall that the results (in table 8.10 indicate no significant effect of task.

The results of the within- and cross-condition Ahi-A23 differences in L2 children's speech confirm that Ahi-A23 for coronals consistently exceeds that for bilabial plosives (in all language groups) (Stevens et al., 1999; Suchato, 2004a). The effect of the subsequent vowel also seems to have an effect on Ahi-A23 amongst children data but against the predictions made in Chapter 5. It was predicted that Back vowels contribute to a longer frontal cavity (especially back-high vowels since they further involve lip-rounding), whereas front vowels lead to a shorter frontal cavity. But the Ahi-A23 data show the opposite trend. It further shows a statistical difference between Back-High and both Front vowels. See table 8.11 and figure 8.2 below.

Table 8.11 Post hoc pairwise comparisons for spectral tilt by Vowel

contrast	estimate	SE	df	t.ratio	p.value
Back-High - Back-Low	-1.258	0.710	75.2	-1.772	0.2949
Back-High - Front-Low	-2.181	0.660	33.1	-3.306	0.0116
Back-High - Front-High	-3.007	0.632	24.6	-4.759	0.0004
Back-Low - Front-Low	-0.923	0.667	329.6	-1.383	0.5107
Back-Low - Front-High	-1.750	0.747	56.8	-2.342	0.1007
Front-Low - Front-High	-0.826	0.653	34.9	-1.267	0.5898

---

[52] A model to test this assumption comparing DigLin with the Traditional teacher's spectral tilt refused to converge despite attempts to use different optimizers/ changing specifications of interaction terms. The model specified is the following: `Tilt Ĩnstruction * Place + Vowel + Gender + (1 | speaker) + (1 | item)`

Running the model output yields the following warning message:

```
In as_lmerModLT(model, devfun) :
Model may not have converged with 1 eigenvalue close to zero:
-8.8e-11.
```

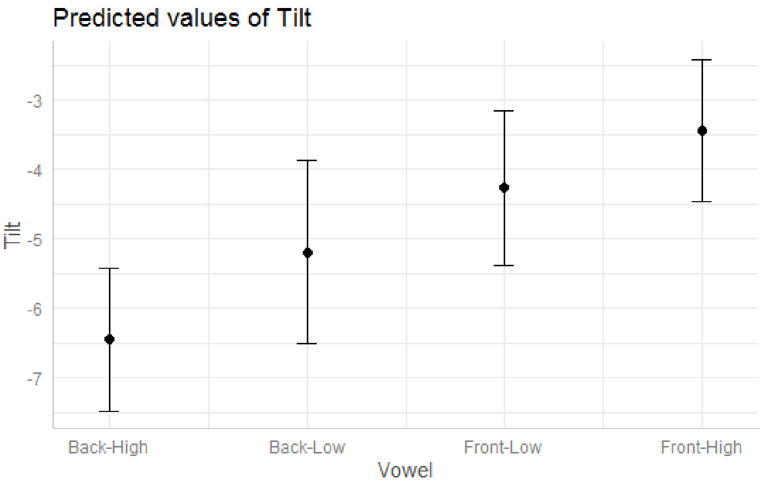


Figure 8.2 Predicted Ahi-A23 values for the four vowel contexts

The lack of statistical difference between males and females at this age (in the by-condition analysis) indicates that such differences may not start to appear until a later age as children continue their anatomical and neuromuscular development of the vocal tract (Vorperian and Kent, 2007). Gender differences in children are controversial (Lee et al., 1999) and their speech components seem to develop at varying stages depending on the acoustic correlate measured. For instance, Eguchi and Hirsh (1969) examined fundamental frequency in prepubescent children aged from five to eleven years and their findings demonstrate a lack of gender-related differences in all of the age groups examined. Lee et al. (1999) also examined several acoustic attributes in children and adolescents aged between five and seventeen alongside adults. Their findings suggest that sex differences in fundamental frequency and formant frequency start to appear around the age of 11 and do not become fully established until the age of 14 to 15, when they resemble canonical adult norms. However, Vorperian and Kent (2007) conclude that gender-related differences in formant frequency start to emerge by the age of four and become more prominent by the age of eight. In terms of temporal measurements, Whiteside, Henry, et al. (2004) examined VOT in prepubescent children aged from 5;8 to 13;2 years. Their findings show a lack of gender differences in younger age groups and that gender-related differences start to appear as children approach the age of 13.

The learners in each condition continued to behave in the same way during the delayed post-test despite the statistical difference between the data form the two times of testing. It can be argued that the difference observed in this regard may be the result of change in oral cavity size, not a deterioration in phonetic learning.

Chapter 9: Phonological Processes

9.1 Introduction

This chapter reports on the results of the analysis of phonological processes exhibited in the seven-year-old Libyan learners' interlanguage after three weeks of English language training. In Chapter 2, models of L2 speech learning and effects of literacy and orthography were discussed. In Chapter 3, we have presented the phonemic inventory of the participants' native language and that of the target language of the programme used in the training. In Chapter 4, it was stated that other factors influencing language learning or acquisition include implicational markedness (Eckman, 1991; Eckman, 2008), input frequency effects (Beckman and Edwards, 2000; Bernhardt and Stemberger, 2017; Edwards et al., 2004; Levelt et al., 2000; Munson et al., 2011; Roark and Demuth, 2000; Storkel, 2004; Zamuner, 2003; Zamuner et al., 2004), functional load (Stokes and Surendran, 2005) and developmental factors. Functional constraints such as frequency of input and functional load were kept constant across the training conditions. We asked however, whether the error patterns observed in L2 phonology were typical of the target language development or reflective of L1 transfer. It may be difficult to decide if the processes observed in the participants' interlanguage phonology are also found in both L1 acquisition and adult L2 acquisition. However, if the avoidance and repair strategies differ between child L1 and adult L2 acquisition, a more reliable decision can be made.[53] One of the aims of the impressionistic transcription was to establish these phonological processes. The other aim was to help make target-likeness judgements. Once the phonological processes are identified, they can be compared to a) the respective input types received in each training condition, b) the developmental processes, usually displayed by typically-developing child learners

---

[53] See more in section 5.7.

of English and Arabic, and c) universal language tendencies. This process was adopted to answer the research question posed in Chapter 5:

RQ2: What developmental processes do beginner Arabic child learners of L2 English exhibit in each condition?

It was hypothesised that the CAPT training conditions – Listen and Speak, and Listen Only – will resemble *relatively more* of the English child phonological developmental stages and *relatively less* L1 interference in comparison with the Traditional training condition.

Furthermore, unlike the Target-likeness and Match ratings, Chapter 6, which dealt with each problematic sound holistically, this chapter deals with problematic sounds in different phonological environments and/or subcategories, that is voiceless vs. voiced, onset vs. coda, specific coda cluster, or specific diphthong. The chapter is divided into the following sections. Section 9.2 deals with processes relating to the class of affricates. It is divided into three sections. Section 9.2.1 deals with word-initial voiceless affricates, section 9.2.2 deals with word-initial voiced affricates, and section 9.2.3 deals with word-final voiceless affricates.[54] Section 9.3 presents results relating to the class of final CC clusters. It is divided into the following sections. Section 9.3.1 deals with word final /-ld/ clusters, section 9.3.2 deals with word final /-lt/ clusters, section 9.3.3 deals with word final /-lk/ clusters, section 9.3.4 deals with word final /-ft/ clusters, section 9.3.5 deals with word final /-st/ clusters, section 9.3.6 deals with word final /-nt/ clusters, section 9.3.7 deals with word final /-nd/ clusters, section 9.3.8 deals with word final /-nz/ clusters, and finally section 9.3.9 deals with word final /-mp/ clusters. Section 9.4 deals with processes found in the class of voiceless dental fricatives.[55] It is divided into the following sections. Section 9.4.1 deals with word-initial singleton dental fricatives, and section 9.4.2 deals with word-final singleton dental fricatives. Section 9.5 deals with processes found in the class of diphthongs. It is divided into three main sections. Section 9.5.1 deals with the process of monophthongisation, section 9.5.2 deals with the process of substitution, and section 9.5.3 deals with the process of gliding. Finally, section 9.6

---

[54] The training software *DigLin* did not have enough word-final voiced affricate tokens for them to be included in the analysis

[55] Again, the training software *DigLin* did not have any tokens of word-initial voiced dental fricatives for them to be included in the analysis. This is because voiced dental fricatives rarely occur in monosyllabic content words in English.

presents results for rhotic approximants as a class of problematic sounds. It is further divided into section 9.6.1 which deals word-initial rhotic approximants and section 9.6.2 deals with word-final rhotic approximants. Each section of the main problematic sound class commences with presenting results for the Traditional Teacher. Each subsection presents results for the post-test first followed by a section for the delayed test. Results in the delayed test reflect two aspects. One is change reflected by ten weeks of no training and the other is change reflected by language development as a function of age/maturity.

9.2 Affricates

Studies of English first language acquisition (e.g. Dodd, Holm, Hua, et al., 2003; Dodd, Holm, Crosbie, et al., 2013; Smit et al., 1990) indicate that English affricates are the first to be acquired in the set of sound classes examined in the present study (excluding plosives) and are categorised as relatively *middle* consonants, that is they are neither early nor late to be acquired.[56] Voiceless affricates are usually mastered by the age of 3;6 to 3;11, that is earlier than their voiced cognates which are usually mastered by the age of 4;0 to 4;5. Deaffrication is a commonly observed substitution process amongst English-speaking children (Dodd, Holm, Crosbie, et al., 2013). Other substitution patterns also observed (Ingram et al., 1980) for English-speaking children are [t], [s] and [ts] for /ʧ/ and [d] (most frequently), [ts] and [dz] for /ʤ/. Ingram et al. (1980) did not observe deaffrication into /ʃ/ or /ʒ/ for /ʧ/ and /ʤ/ respectively in their study. For Arabic adult L2 learners of English, findings show that learners confuse /ʃ/ and /ʧ/ (Altaha, 1995; Kharma and Hajjaj, 1989) and /ʒ/ and /ʤ/ (Barros, 2003; Kharma and Hajjaj, 1989). According to Maddieson (1984), voiceless affricates are more common (50% of world languages included in the UPSID database) compared to their voiced cognates (28% of world languages within the UPSID database). In Jordanian Arabic, Amayreh (2003) argues that children have a higher accuracy rate for affricates in onsets than codas.

---

[56] See Chapter 4, section 4.2.

9.2.1 *Voiceless affricates word-initially*

Test items in this category include the words ‘chest’, ‘chin’, ‘chop’, ‘chip’, and ‘chair’. The Listen and Speak and Listen Only conditions received native British RP input. However, the Traditional Teaching condition received input from the Traditional Teacher. Productions of the female Traditional Teacher are shown in table 9.1. Whilst she exhibited non-targetlike pronunciations of the token as a whole, her production of word-initial voiceless affricates were consistently target-like.

Table 9.1 Traditional Teacher’s realisations of /ʧ-/ tokens

item	task	process	target-likeness	realisation
chair	read aloud	correct	non-targetlike	[ʧeəɾ]
chair	picture-naming	correct	non-targetlike	[ʧeːɾ]
chest	read aloud	correct	targetlike	[ʧɛst^{h}]
chest	picture-naming	correct	targetlike	[ʧɛst^{h}]
chin	read aloud	correct	non-targetlike	[ʧïn]
chip	read aloud	correct	targetlike	[ʧɪp]
chip	picture-naming	correct	targetlike	[ʧɪp]
chop	read aloud	correct	non-targetlike	[ʧop]
chop	picture-naming	correct	targetlike	[ʧɔp]

Post-test

The overall rate of correct realisations for /ʧ-/ was high especially for Listen and Speak, followed by Traditional Teaching and finally Listen Only. See figure 9.1.

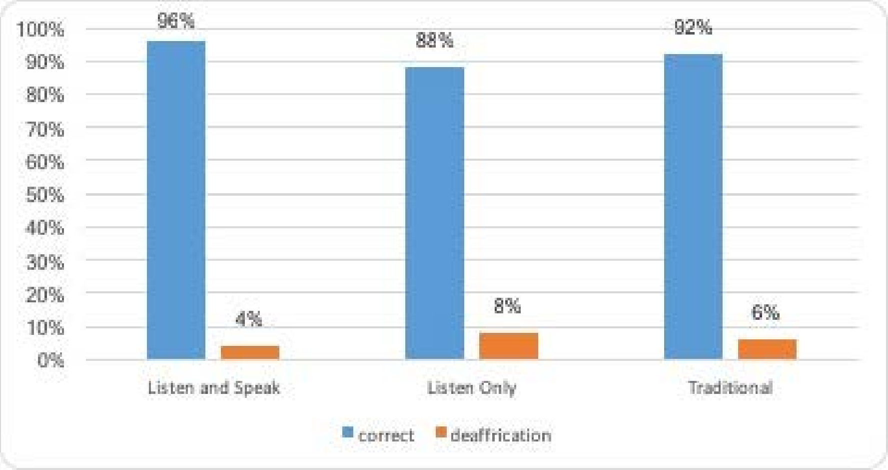


Figure 9.1 Percentages of processes for /ʧ-/ according to instruction type during the post-test

Overall, only 6% of the /ʧ-/ tokens were deaffricated and Listen Only had a relatively higher deaffrication rate than the other groups. Deaffrication is a substitution process English-speaking children (Dodd, Holm, Crosbie, et al., 2013)[57] as well as Arabic-speaking children (Alqattan, 2015) exhibit during their first language acquisition. In English-speaking children, deaffrication persists until after the age of four (Roberts et al., 1990). In Kuwait Arabic children, deaffrication [ʃ] is a relatively frequent (32%) developmental error.[58] However, in Amayreh (2003), this process has been shown to persist in the voiced affricate (as the variety does not employ the voiceless affricate) until the age of 8;4. It is worth-mentioning that in Libyan Arabic, this sound is not part of the phonemic inventory. Therefore developmental patterns for this sound class in this variety are non-applicable. It is also a phonological process Arabic adult L2 learners of English exhibit in their L2 phonology (Altaha, 1995; Kharma and Hajjaj, 1989).

In Listen and Speak instruction, deaffrication was demonstrated in ‘chip’ (for the same speaker in the three tasks) and ‘chop’. In Listen Only, participants who exhibited deaffrication were all literate and mostly females. All but one deaffricated one of the five test items. In Traditional Teaching, it was illustrated in all test items but only during the delayed repetition task. ‘Chin’ was only deaffricated in this group by the only illiterate amongst the rest. An example of consonant harmony is also exhibited in ‘chest’. Consider the following examples in 9.1.

	instruction	item	task	realisation
9.1	Listen and Speak	chip	picture-naming	[ʃɪp]
	Listen Only	chest	delayed repetition	[ʃɛʃt]
	Traditional	chin	delayed repetition	[ʃˡïn]

Less frequent processes include fronting. Consider the following examples in 9.2.

	instruction	item	task	realisation
9.2	Traditional	chest	delayed repetition	[t͡sɛs]
	Listen Only	chop	delayed repetition	[t͡sɔˀp]

---

[57] Dodd, Holm, Crosbie, et al. (2013: 637) define deaffrication as “change of a feature of the affricate”.

[58] The age range of the children in the study by Alqattan (2015) was between 1;4 and 3;4. In Ayyad (2011) and Ayyad et al. (2016), the Kuwaiti Arabic children were older ranging in age between 3;10 and 5;2 and no such error was reported.

Metathesis was also present though infrequent. It occurred on the level of the segment and the syllable. It shows that the participants in these instances considered the affricate as two phonemes instead of one. Consider the following examples in 9.3.

	instruction	item	task	realisation
9.3	Traditional	chest	delayed repetition	[t͡sɛsʃtʰ]
	Listen Only	chest	delayed repetition	[stɛʃt]
	Listen Only	chest	delayed repetition	[sɛʧ]

A very small proportion of the tokens exhibited other less frequent processes including metathesis (Listen Only and Traditional), fronting (Listen Only and Traditional) and vowel epenthesis (Listen Only).

Delayed test

Figure 9.2 shows that the overall rate of correct realisations of /ʧ-/ was high during the delayed test as well. Listen and Speak instruction had a slight rise (+1%) and had the highest rate of correct realisations (97%), whilst Listen Only and Traditional had a slight drop (-2% and -3% respectively) and Traditional having a higher rate of correct tokens (89%) than Listen Only (86%).

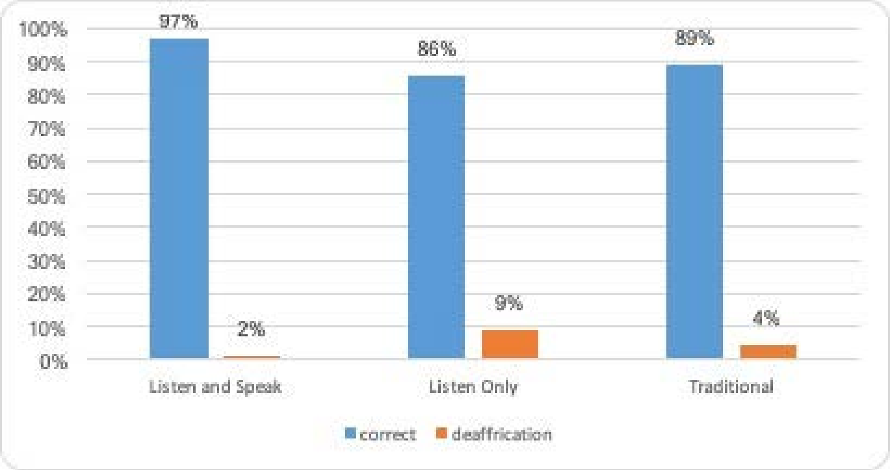


Figure 9.2 Percentages of processes for /ʧ-/ according to instruction type during the delayed test

Only 5.3% of the /ʧ-/ tokens exhibited deaffrication, which was most evident in Listen Only (9%). However, two of the participants exhibited the process in multiple items. It still appeared in other groups but less frequently. The process almost always took place during the delayed repetition task. Consider the following examples in 9.4.

9.4

instruction	item	task	realisation
Listen and Speak	chest	delayed repetition	[ʃɛst]
Listen Only	chip	delayed repetition	[ʃɪp]
Traditional	chop	delayed repetition	[ʃopʰ]

Other less frequent processes include deaffrication and fronting (combined), metathesis, and vowel and syllable epenthesis. Consider the following examples in 9.5.

9.5

instruction	item	task	realisation
Listen and Speak	chest	delayed repetition	[ʃːɛst]
Listen Only	chest	delayed repetition	[t͡sɛʃtʰ]
Listen Only	chest	delayed repetition	[ʃtɛʃ]
Listen Only	chair	delayed repetition	[ʔəʧe̞ː]
Traditional	chest	picture-naming	[sɛʃt]
Traditional	chest	delayed repetition	[təʃɛ̈ʃtʰ]

9.2.2 *Voiced affricates word-initially*

Test items in this category include the words ‘jam’, ‘jet’, ‘jug’, ‘jog’, ‘jar’, ‘jeans’ and ‘jump’. Productions of the female Traditional Teacher show that none were target-like due to various reasons. These include non-targetlike voicing implementation, whether it be in the onset or coda, the realisation of post-vocalic /ɹ/, non-targetlike vowel quality or length, and prosthesis. The latter process might suggest that the phoneme is perceived as a cluster that requires repairing, hence regarded as two independent segments. Consider the Traditional Teacher’ realisations in table 9.2

Table 9.2 Traditional Teacher’s realisations of /ʤ-/ tokens

item	task	process	target-likeness	realisation
jam	read aloud	correct	non-targetlike	[ʤaːm]
	picture-naming	correct	non-targetlike	[ʤaːm]
jar	read aloud	correct	non-targetlike	[ʤɑːɾ]
	picture-naming	correct	non-targetlike	[ʤɑːɾ]
jeans	read aloud	correct	non-targetlike	[ʤ̥ɛnz]
	picture-naming	correct	non-targetlike	[ʤ̥ɛnz̥]
jet	read aloud	correct	non-targetlike	[ʤɛtʰ]

	picture-naming	correct	non-targetlike	[ʤɛtʰ]
jog	read aloud	prosthesis	non-targetlike	[ʔəʤɔg]
	picture-naming	correct	non-targetlike	[ʤɔg]
jug	read aloud	correct	non-targetlike	[ʤʌg]
	picture-naming	prosthesis	non-targetlike	[əʤʌg]
jump	read aloud	correct	non-targetlike	[ʤʌmpʰ]
	picture-naming	correct	non-targetlike	[ʤʌˑmp̬]

Post-test

Figure 9.3 indicates that the rate of correct realisations were generally high for this sound. Listen and Speak instruction had the highest correct score, followed by Listen Only. Traditional had the lowest but still high (82%).

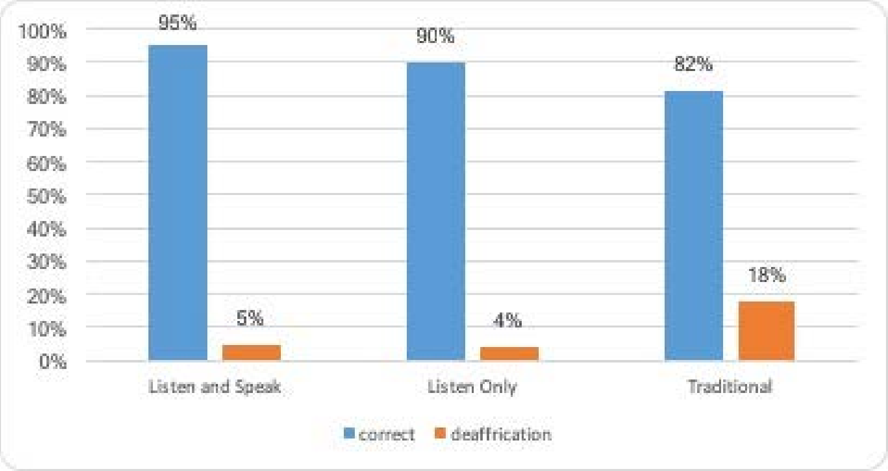


Figure 9.3 Percentages of processes for /ʤ-/ according to instruction type during the post-test

A recurring process exhibited by the data tokens during the post-test was deaffrication. It was not high for the experimental groups. However, a fifth of the tokens containing /ʤ-/ from Traditional demonstrated deaffrication. In Listen and Speak instruction, four out of the six tokens demonstrating deaffrication were produced by the same participant during the delayed repetition task, whilst only one token was produced in the picture-naming task. In Listen Only two out of the five tokens were produced by the same participant. The process appeared in the delayed repetition task only for this group. In Traditional, the twenty seven tokens illustrating this process were elicited in both picture-naming and delayed repetition. These were produced by twelve of the eighteen participants, the majority of which were males. Consider the following examples:

9.6

instruction	item	task	realisation
Listen and Speak	jug	delayed repetition	[ʒʌg]
Listen Only	jeans	delayed repetition	[ʒ̥iːnd̥]
Traditional	jet	picture-naming	[ʒɛt]

Other less frequent processes that took place during the post-test (mostly within Listen Only) include syllable epenthesis, and vowel epenthesis. Syllable epenthesis appeared only in Listen Only and all tokes were by a single male speaker, who also exhibited vowel epenthesis. This process was seen in Traditional (just once) and another Listen Only participant, who produced more than one token illustrating such a process. Not all of the test items underwent these processes. It was confined to the items ‘jet’, ‘jog’, and ‘jug’. Consider the following examples:

9.7

instruction	item	task	realisation
Listen Only	jog	delayed repetition	[ʔəʤɔːg̊ʰ]
Traditional	jet	picture-naming	[əʤɛ̈tˢ]

Delayed test

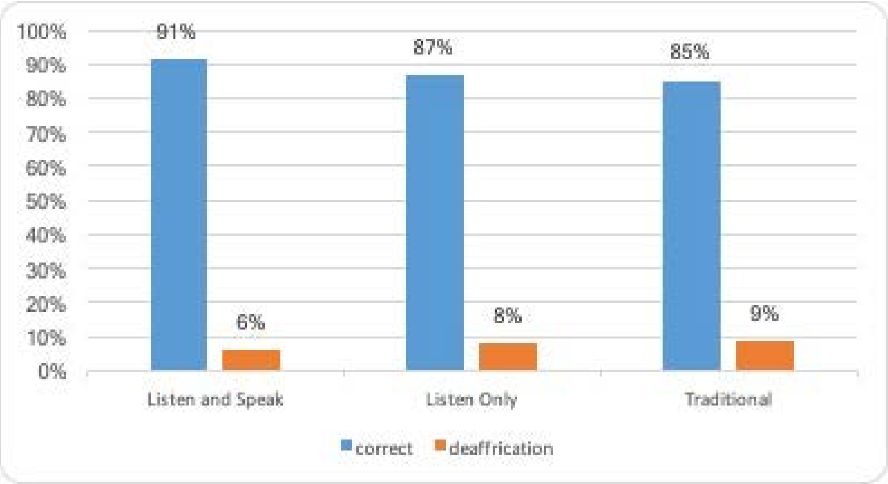


Figure 9.4 Percentages of processes for /ʤ-/ according to instruction type during the delayed test

The figure shows that the overall rate of correct realisations of /ʤ-/ during the delayed test was still high, though it decreased slightly for the experimental groups and increased slightly for Traditional. Consequently, the rate of deaffrication increased slightly for the experimental groups but for Traditional, it decreased by half the initial rate. Consider the following examples:

9.8

instruction	item	task	realisation
Listen and Speak	jam	delayed-repetition	[ʒaːm]
Listen Only	jog	delayed repetition	[ʒɔg]
Traditional	jet	picture-naming	[ʒɛt]

Other less frequent processes during the delayed test include vowel epenthesis (next in place after deaffrication), metathesis (both in Listen Only and Traditional only), and syllable epenthesis (in Listen Only and Listen and Speak). For metathesis, the data implies that the participants seem to consider the affricate as two phonemes as compared to one. Consider the following examples:

9.9

instruction	item	task	realisation
Listen Only	jar	delayed-repetition	[əʤɑ̥ː]
Traditional	jam	delayed-repetition	[əʤaˑm]
Traditional	jeans	delayed-repetition	[d͡siːnʃ]
Listen Only	jeans	delayed-repetition	[ʒ̥d̥iːnz]
Listen and Speak	jug	delayed-repetition	[ʔəʤ̥ʌːg̥]

9.2.3 *Voiceless affricates word-finally*

Test items in this category include the word ‘march’ only. Realisations of the female Traditional Teacher are shown in table 9.3.

Table 9.3 Traditional Teacher’s realisations of /-ʧ/ tokens

item	task	process	target-likeness	realisation
march	read aloud	deaffrication	non-targetlike	[mɑːɾʃ]
march	picture-naming	deaffrication	non-targetlike	[mɑːɾʃ]

Post-test

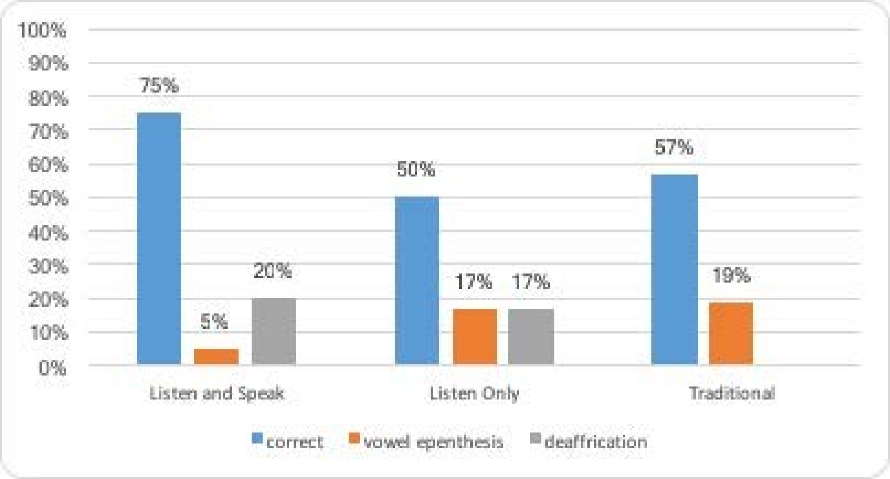


Figure 9.5 Percentages of processes for /-ʧ/ according to instruction type during the post-test

The overall rate of correct realisations for /-ʧ/ in the post-test was high especially for Listen and Speak instruction, followed by Traditional and finally Listen Only. However, figures were all lower in comparison with /ʧ-/.

Two rather recurrent processes took place during the post-test; vowel epenthesis and deaffrication.

Overall, nearly 14% of /-ʧ/ tokens demonstrated vowel epenthesis. The participants seemed to regard the target sound as two independent segments and together were treated as a coda cluster. This clustering was repaired by vowel epenthesis. The nature of the inserted vowel is similar to an excrescent one. The group with the highest occurrence of vowel epenthesis was Traditional, followed closely by Listen Only. Listen and Speak instruction had the lowest rate with only one token demonstrating it. Consider these examples:

	instruction	item	task	realisation
9.10	Listen and Speak	march	delayed repetition	[mɑtɨ̆ʃ]
	Listen Only	march	delayed repetition	[mɑˑtɨʃ]
	Traditional	march	picture-naming	[mɑɾtɨʃ]

Overall, nearly 12% of /-ʧ/ tokens were deaffricated. The group with the highest deaffrication rate was Listen and Speak. Listen Only had a slightly lower rate, whereas Traditional exhibited no such process for this context. Consider the following examples:

9.11

instruction	item	task	realisation
Listen and Speak	march	delayed-repetition	[mɑːəʃ]
Listen Only	march	delayed-repetition	[mɑːʃ]

Other less frequent processes that appeared in Listen Only and Traditional only include deletion (when post-vocalic /ɹ/ was realised, together with the affricate a coda cluster was created), fronting (depalatalisation), stopping, backing and vowel epenthesis and finally backing only. Consider the following examples:

9.12

instruction	item	task	realisation
Traditional	march	delayed-repetition	[mɑːɾ]
Listen Only	march	delayed-repetition	[mɑːt͡s]
Traditional	march	delayed-repetition	[mɑ̹ːt̚]
Listen Only	march	delayed-repetition	[mɑːkɨʃ]
Listen Only	march	delayed-repetition	[moːkʃ]

Delayed test

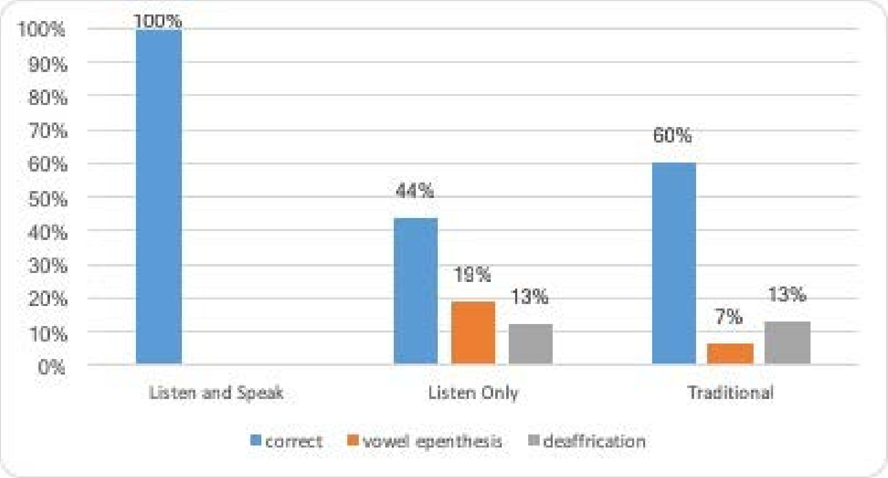


Figure 9.6 Percentages of processes for /-ʧ/ according to instruction type during the delayed test

The figure shows an overall increase in the rate of correct realisations of the word-final affricate in the delayed test. A cross-instruction group examination reveals that Listen and Speak instruction had the highest rate of increase in correct production at 100% (a third times more).

Traditional learner's performance also improved with an additional 3%. However, Listen Only learner's performance declined slightly (-6%) and is still the lowest across instruction types.

Processes appearing in the delayed test were the same as those in the post-test but with some changes in occurrence rate. Given that Listen and Speak had 100% rate of correct realisations of the sound, it illustrated no such processes. However, for Listen Only vowel epenthesis and deaffrication appeared once again at 19% (an increase - only three tokens) and 13% (a decline - only two tokens) respectively. Traditional instruction learners did not show any deaffrication in the post-test but in the delayed test, this process was evident in 13% (only two tokens) of the group's total amount of tokens. Conversely, vowel epenthesis dropped (-12% - only one token) for this instruction type in this test. Other various processes include deletion, backing, stopping, backing and vowel epenthesis (simultaneously), and substitution. Consider the following examples:

9.13

instruction	item	task	realisation
Traditional	march	delayed-repetition	[mɑ̹ːtɨʃ]
Listen Only	march	delayed-repetition	[mɑːtɨʃ]
Listen Only	march	delayed-repetition	[mɑːʃ]
Traditional	march	delayed-repetition	[mawɨʃ]
Traditional	march	delayed-repetition	[maːw]
Listen Only	march	delayed-repetition	[mɑːɾtʰ]
Listen Only	march	delayed-repetition	[mɑːʔʃ]
Listen Only	march	delayed-repetition	[mɑːkʃ]
Traditional	march	delayed-repetition	[mɑ̹ːkəʃ]

9.3 Coda Clusters

Percentages are obtained by dividing the raw count by the total number of produced tokens per coda category and not by stimulus count. Coda categories include /-ld/, /-lt/, /-lk/, /-nd/, /-nt/, /-nz/, /-mp/, /-st/ and /-ft/.

Mismatches in voicing between the target cluster and child's production were not coded as errors. This is because according to Stoel-Gammon and Buder (1999), reliable voicing distinction in codas develops rather late.

In the literature reduction is a term used to refer to the elision of one or two members of the cluster (e.g. McLeod et al., 2001b).[59] I deal with each separately. Percentages are

---

[59] See Chapter 4, sections 4.3.4 and 4.2.1.

calculated on the basis of actual data produced not stimulus items.

9.3.1 */-ld/*

Test items in this category include the word ‘cold’ only. Realisations of the female Traditional Teacher are shown in table 9.4.

Table 9.4 Traditional Teachers’ realisations of /-ld/ tokens

item	task	process	target-likeness	realisation
cold	read aloud	correct	non-targetlike	[kɔːɫd]
	picture-naming	correct	non-targetlike	[kɔːɫd]

Post-test

The figure in 9.7 shows that the rate of correct /-ld/ realisations in the post-test was 30%. Listen and Speak had the highest rate of correct realisations of the /-ld/ cluster at around 43% followed by Traditional 40%. Listen Only had a considerably low rate of correct realisations at about only 5%.

Generally, for /-ld/, the most common processes within the post-test were reduction (50%) deletion (8.3%) substitution (8.3%) and finally reduction and substitution combined (3.3%).

The rates of deletion and substitution were the same in the post-test. Listen Only had the highest rate in both processes 10.5% and about 16% respectively. Listen and Speak had a higher rate for deletion 9.5% and substitution around 5%. Traditional had a rather low rate for either process at 5% each.

Finally, a few cases had two processes occurring simultaneously these are reduction and substitution at roughly 5% of /-ld/ tokens for Listen Only and Traditional. Meanwhile, Listen and Speak exhibited no such processes during the post-test.

No read aloud data for /-ld/ was produced during the post-test. There were two items for picture-naming; one from Listen and Speak and another from Traditional. In the delayed test, there were two from Traditional.

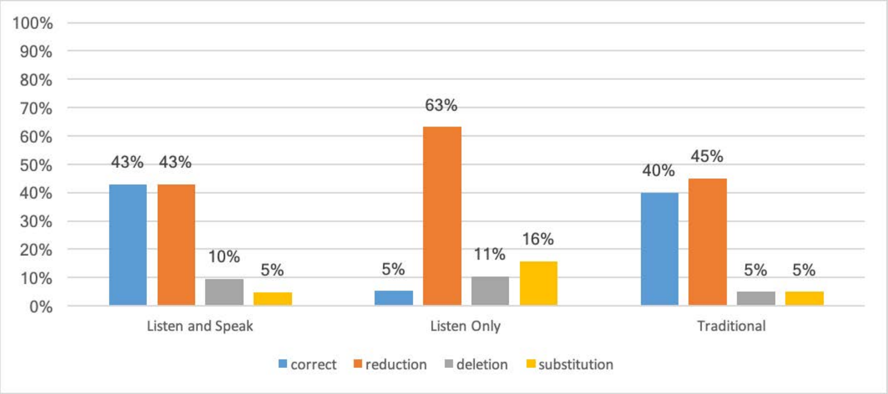


Figure 9.7 Percentages of processes for /-ld/ according to instruction type during the post-test

Reduction

Half the analysable tokens containing /-ld/ in the post-test were reduced. Listen Only has the highest rate at about 63%. Followed by this are Traditional at 45% and Listen and Speak having the lowest rate, though not far behind Traditional, at about 43%.

Reduction involves the elision of C1 or C2. Only a few data items exhibit the elision of C2.

	instruction	item	task	realisation
9.14	Listen and Speak	cold	delayed repetition	[kɔʊd̥]
	Listen Only	cold	delayed repetition	[kɔːl]
	Traditional	cold	delayed repetition	[kɔːd̥]
	Traditional	cold	delayed repetition	[kʰɔuɨd̥]

For a subset of the data coded as exhibiting reduction in the cluster /-ld/, it is a possibility that in this process, the C1 /l/ is vocalised instead of being elided. However, given that the previous segment is a diphthong of which the second member is /ʊ/ that can also be /w/, it was hard to decide whether it belonged to the diphthong or coda.

Deletion

This process involves the elision of both members of the cluster. Because the preceding segment is a diphthong, it is possible that the second vowel either belongs to the diphthong or C1 glided into /w/ for the two cases from Listen and Speak.

9.15

instruction	item	task	realisation
Listen and Speak	cold	delayed repetition	[kʰɔʊ]
Listen Only	cold	delayed repetition	[kʰɔː]
Traditional	cold	delayed repetition	[koː]

Substitution

Five cases of substitution occurring on its own (as opposed to accompanying reduction) took place during the post-test. In all five cases, C1 /l/ was substituted with a homorganic nasal /n/. C2 /d/ remained intact in all cases.

9.16

instruction	item	task	realisation
Listen and Speak	cold	delayed repetition	[kɔnt]
Listen Only	cold	delayed repetition	[kɔnd]
Traditional	cold	delayed repetition	[kɔˑnd]

Reduction and substitution

Similarly, C1 /l/ was replaced with /n/ and C2 was elided.

9.17

instruction	item	task	realisation
Listen Only	cold	delayed repetition	[kʰoːn]
Traditional	cold	delayed repetition	[kʰɔn]

Delayed test

The rate of correct /-ld/ realisations in the delayed test dropped a further 7%. Listen and Speak had a considerable drop in rate of correct realisations of /-ld/ in the delayed test (from 43% to 23%). Traditional did not change much (from 40% to 37%). Listen Only had a slight increase but still performed poorly in the delayed test in comparison to the other instruction types.

A new process appeared in one of the tokens in the delayed test for Traditional, namely vowel epenthesis. Meanwhile it had no instances of a combined reduction and substitution as it had in the post-test.

Overall reduction rate increased to 54.2%. The most evident increase was seen in Listen and Speak (77%). However, this was the only process noted for this group in the delayed test. Listen Only also had an increase but only an additional 5%. Traditional on the other hand had a sharp decline (from 45% to 26%).

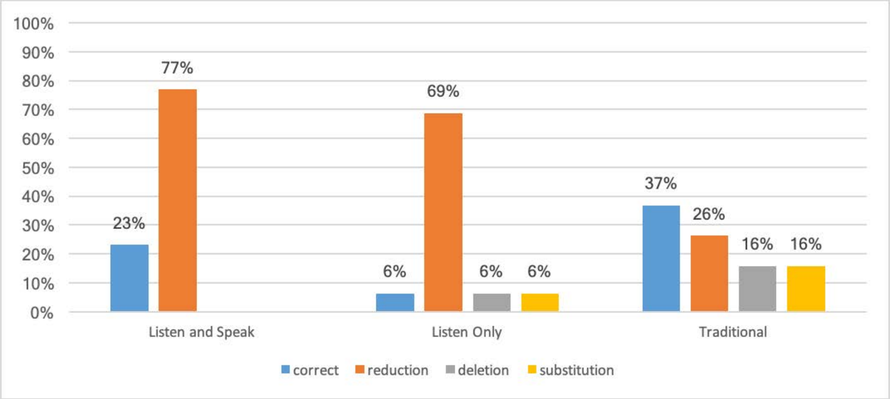


Figure 9.8 Percentages of processes for /-ld/ according to instruction type during the delayed test

Reduction

A total of twenty-six tokens containing /-ld/ were reduced in the delayed test. C1 is elided and C2 is retained with the exception of three cases where C2 was elided. Two of these cases were found in Traditional whereas only one in Listen and Speak and none in Listen Only. Fifteen realisations had a back high rounded vowel that can possibly be C1 /l/ glided.

9.18

instruction	item	task	realisation
Listen and Speak	cold	delayed repetition	[kʰɔʊd̥]
Listen Only	cold	delayed repetition	[kʰɔːd̚]
Traditional	cold	delayed repetition	[kʰɔ̰ːɫ]

Deletion

In all four cases, the presence of the rounded back vowel suggests the possibility of /l/ gliding.

9.19

instruction	item	task	realisation
Listen Only	cold	delayed repetition	[kʰʌʊː]
Traditional	cold	delayed repetition	[kɔːʊ]

Substitution

Again, C1 is the substituted segment just like in the post-test and similarly, it is substituted with its homorganic nasal counterpart.

9.20

instruction	item	task	realisation
Listen Only	cold	delayed repetition	[kʰə̟ʊnd̥]
Traditional	cold	delayed repetition	[kʰɔːnd]̥

Reduction and substitution

In the first example, C1 is elided and C2 is substituted with a homorganic fricative /θ/. There also appears to be an epenthetic vowel before C2. In the second example, however, C2 is elided and C1 is replaced with a homorganic nasal /n/.

9.21

instruction	item	task	realisation
Listen Only	cold	delayed repetition	[kʰɔʊɨθ]
Listen Only	cold	delayed repetition	[kʰɔŋ]

Vowel epenthesis

An excrescent high central vowel is added between the two members of the target /-ld/ cluster by a participant from the Traditional Teaching condition. This vowel is used in Libyan Arabic internally for epenthesis within coda clusters. For example, in [galɨb] ‘heart’, [ɨ] is an excrescent vowel (Ehbara, 2015).

9.22

instruction	item	task	realisation
Traditional	cold	delayed repetition	[kʰɔːlɨd̥]

9.3.2 */-lt/*

Test items in this category include the words ‘belt’, ‘quilt’ and ‘salt’. Realisations of the female Traditional Teacher are shown in table 9.5.

Table 9.5 Traditional Teachers' realisations of /-lt/ tokens

item	task	process	target-likeness	realisation
belt	read aloud	correct	non-targetlike	[bɛltʰ]
	picture-naming	correct	non-targetlike	[bɛltʰ]
salt	read aloud	correct	non-targetlike	[sɔltʰ]
	picture-naming	correct	non-targetlike	[sˤɔɫt]
quilt	read aloud	wrong	wrong	[kwaɪtʰ]
	picture-naming	wrong	wrong	[kwaɪtʰ]

In the first four realisations, the presence of both members of the cluster unchanged is considered a correct realisation regardless of voicing or whether the lateral is clear or velarised. In the latter two realisations for the test item 'quilt', the Traditional Teacher seems to have mispronounced the item or has it confused with the word 'quiet'.

In the previous cluster type /-ld/, it was unclear whether the segment /u/ or /ʊ/ belonged to the diphthong preceding the cluster or the first part of the cluster underwent /l/ vocalisation. In the current test items, there were no diphthongs. Thus, any such cases were considered /l/ vocalisation.

Post-test

The overall rate of correct /-lt/ realisations in the post-test is about 42%. Traditional had the highest rate of correct realisations of the /-lt/ cluster with half the tokens containing /-lt/ realised correctly that is not undergoing any of the phonological processes indicated in the table. The instruction group with the next highest rate of correct /-lt/ realisations is Listen and Speak at just above 39%. Listen Only had the lowest rate of correct realisations at about 35%. All but one token in the memory recall tasks were correct for all groups.

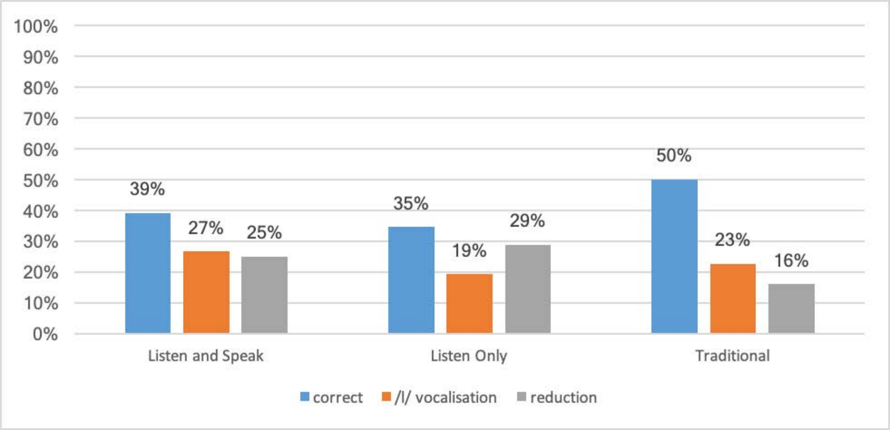


Figure 9.9 Percentages of processes for /-lt/ according to instruction type during the post-test

Generally, for /-lt/, two main processes emerge in the post-test. These are /l/ vocalisation and reduction (23% each). These almost always appear in the delayed repetition task. Another less frequent process is substitution (5.6%).

/l/ vocalisation

For /l/ vocalisation in /-lt/, Listen and Speak has the highest rate at approximately 27%. Followed by this are Traditional at around 23% and Listen Only having a relatively lower rate at 19%. It only appeared in the delayed repetition task. It is most recurrent in the test item 'belt'. It is considerably less evident in 'salt' and 'quilt'. /l/ vocalisation in the former appeared only once in Traditional and twice in Listen and Speak but never in Listen Only. In the latter, it appeared only once in Listen Only and twice in Listen and Speak but never in Traditional.

Reduction

As for reduction, Listen Only has the highest rate amongst instruction groups (29%). Listen and Speak has the next highest rate (25%). Traditional has the lowest rate (16%). By examining the data table, it seems that 'belt' had the least reduction rate across groups, appearing only once in Listen Only and twice only in Traditional, but never reduced in Listen and Speak. Moreover, the lateral in /-lt/ seems to be cluster member affected in this process as it is elided more often than C2. In the experimental groups, C1 always undergoes elision in the reduction process, whereas in Traditional, there seems to be exceptions.

Substitution

Substitution is the next process in place. There is not a noticeable difference in the rate of substitution across groups. Although Traditional has double the amount of Listen and Speak, it is still low. Also, all substitutions for /-lt/ in the post-test took place during the delayed repetition task. Substitution in /-lt/ was exhibited in the three test items ‘quilt’, ‘salt’ and ‘belt’. In each case, it was C1 /l/ that was replaced with either /ɹ/ or /n/. /ɹ/ and /n/ share the place of articulation and voicing with /l/. /ɹ/ also belongs to the same class as /l/, that is liquids but neither share manner. However, both target /l/ and [ɹ] are highly sonorous and this is predictable given C1 is adjacent to the syllable nucleus. C2 remained intact in this process.

There was a single case for consonant epenthesis in Listen Only, ‘quilt’, [kɔlts], which did not appear during the delayed test. Thus, it is not considered a process. The same is true for gliding and vowel epenthesis combined, ‘quilt’, [kwïjɨtʰ] where /l/ was glided and an epenthetic vowel was inserted between members of the cluster. The latter process took place in Traditional.

There were several instances that could not be assigned to a specific process, especially for ‘quilt’ as they seem to be recurring. C1 seems to be moved to the onset to replace the glide /w/ in order to simplify the coda. As for ‘salt’ [sauft˺] and ‘belt’ [pʰɛtʰ], the participants may have confused each item for another test word such as ‘left’ and ‘paint’ respectively.

Delayed test

The overall rate of correct /-lt/ realisations in the delayed test increased a further 2%. The rate increased a further 5% for Listen and Speak, remained the same for Listen Only and slightly increased (a further 2%) for Traditional. The increase was evident in the delayed repetition task only. There were no analysable tokens for memory recall tasks for the experimental groups.

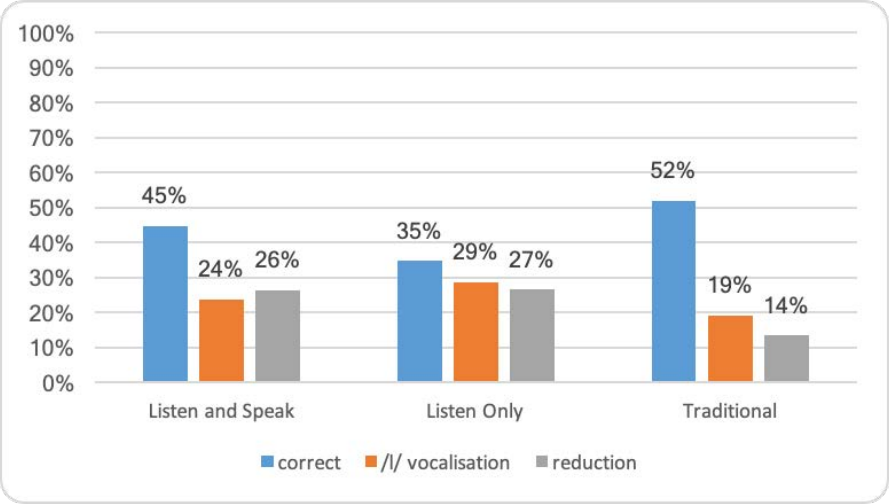


Figure 9.10 Percentages of processes for /-lt/ according to instruction type during the delayed test

Processes appearing in the post-test reappeared in the delayed test. A new process appeared as vowel epenthesis though less frequent, it occurred separately as well as accompanying /l/ vocalisation in one of the tokens. /l/ vocalisation was also accompanied by reduction also in one token. Both of these tokens were from Traditional during the delayed repetition task.

/l/ vocalisation (23.7%)

There is not a large difference in the rate of /l/ vocalisation across instruction types especially with the lack of comparative data across memory recall tasks. It was highest amongst Listen Only increasing a further 10%, whilst it is lower for Listen and Speak (23.7%) and Traditional (19.2%) and has decreased in comparison with the post-test (3% and 3.4% less respectively). It seems to be happening more with the literate than the illiterate.

Reduction (21%)

The rate of reduction in the delayed test decreased slightly (2% less). It increased slightly for Listen and Speak but increased also slightly for Listen Only. In the delayed test, these two instruction groups have the same rate of reduction.

The rate of reduction in /-lt/ also decreased in Traditional (13.5%). Again, in the majority of tokens, C1 is elided and C2 is retained. Two exceptions are ‘quilt, [kwɪl] from Listen and Speak and ‘salt’, [sɔ̜l] from Traditional in the picture-naming task. No compensatory lengthening seems to take place for the latter cases. There was a case of orthographic influence affecting test item ‘belt’, [dɛ̢ˑtʰ].

Substitution (4.3%)

Just like the post-test, C1 underwent substitution and C2 remained intact. C1 /l/ was replaced with /ɹ/ and /n/. One token showed it was replaced with [ʔ].

Vowel epenthesis (3.6%)

This process did not appear in the post-test for this cluster. However, in the delayed test, it appeared in all instruction groups for the test items, ‘quilt’ mainly and ‘salt’. In Listen Only, there seems to be a case of vowel harmony.

/l/ vocalisation and epenthesis

One participant from Traditional exhibited this process in the test item ‘belt’ during the delayed repetition task. Having mispronounced the vowel as well makes it unlikely that this word was learned.

/l/ vocalisation and reduction

This process appeared once for this cluster type. The token was produced by a participant from Traditional during the delayed repetition task for the test item ‘quilt’.

There is a group of items that do not seem to fit any of the above processes. For ‘quilt’ realised as [kɬaɪɬ], the participant seems to be confused between what he learned from his Traditional Teacher, [kwaɪ̠tʰ] and input from the delayed repetition task produced by DigLin. Consider example 9.23 in the following table:

	instruction	item	task	realisation
9.23	Listen Only	quilt	delayed repetition	[kʰɔːwɨt]
	Traditional	quilt	delayed repetition	[kɬaɪɬ]
	Listen Only	quilt	delayed repetition	[kɔɾɔˑ]
	Traditional	salt	delayed repetition	[sˡɔʔsˡ]
	Listen Only	salt	delayed repetition	[t͡sɔːˤ]
	Listen Only	salt	delayed repetition	[sːʌwɨtʰ]

9.3.3 */-lk/*

Test items in this category include the word 'milk' only. Realisations of the female Traditional Teacher are shown in table 9.6.

Table 9.6 Traditional Teachers' realisations of /-lk/ tokens

item	task	process	target-likeness	realisation
milk	picture-naming	correct	target	[mɪɫk]
	read aloud	correct	target	[mɪɫk^h]

Post-test

The rate of correct /-lk/ realisations in the post-test is 47%. Traditional has the highest rate of correct realisations of the /-lk/ cluster at 64% followed by Listen Only and Listen and Speak scoring closely at 37% and 36% respectively.

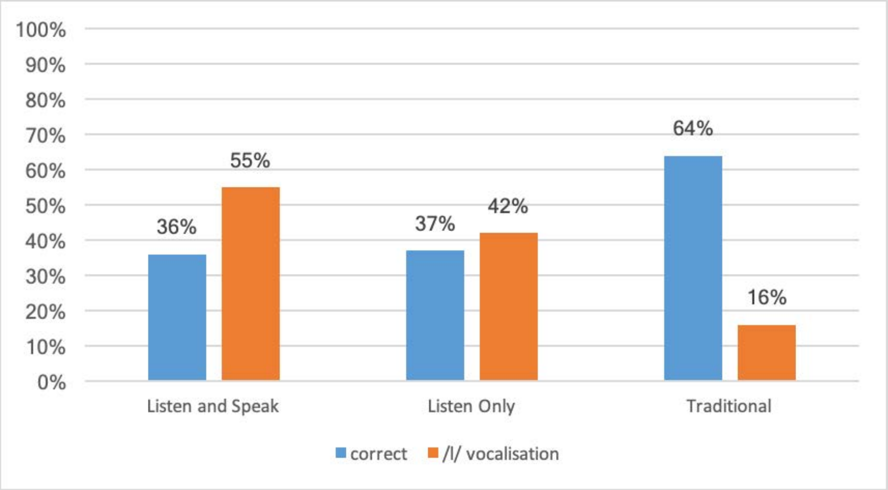


Figure 9.11 Percentages of processes for /-lk/ according to instruction type during the post-test

Generally, for /-lk/, the most common process within the post-test is /l/ vocalisation (36%). Other less frequent processes are /l/ vocalisation and reduction simultaneously (3%), /l/ vocalisation and substitution simultaneously (2%) and finally vowel epenthesis (2%).

/l/ vocalisation (36%)

This process is most evident in Listen and Speak with over half of the tokens containing /-lk/ having C1 /l/ vocalised. Listen Only also had a relatively high rate at 42%. However, Traditional had a rather lower rate with only 16% of C1 in /-lk/ tokens vocalised. Memory recall tokens from experimental groups, despite their scarcity, all exhibited /l/ vocalisation.

Reduction

In all cases, C1 is elided and C2 is retained. This process appeared in tokens from experimental groups only. It occurred twice as much in Listen Only than Listen and Speak.

/l/ vocalisation and reduction

C1 /l/ is vocalised and C2 /k/ is elided in two tokens from Traditional during the delayed repetition task. This does not appear in either of the experimental groups.

/l/ vocalisation and substitution

One token from Traditional during the delayed repetition task exhibited a process whereby C1 is vocalised and C2 is replaced with /f/.

Vowel epenthesis

Two tokens containing /-lk/ showed epenthesis, both in Traditional but by two different speakers during two different tasks; picture-naming and delayed repetition. For the former task, the epenthetic vowel appeared between the cluster members, whereas the latter appeared after the cluster.

Delayed test

The overall rate of correct /-lk/ realisations in the delayed test decreased (-3%). Across instruction groups, increase is exclusive to Listen and Speak (+10%). Listen Only's rate of correct realisations of /-lk/ decreased (-15%) making it maintain its lowest rate also in the delayed test. Similarly, the rate of correct realisations for Traditional decreased (-6%). Despite Traditional's decrease, it still has the highest rate of correct realisations of /-lk/ in the delayed test (58%).

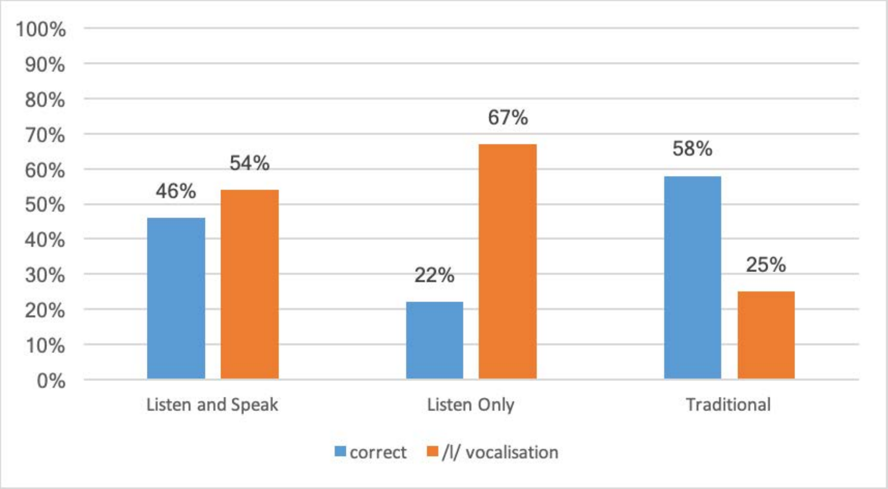


Figure 9.12 Percentages of processes for /-lk/ according to instruction type during the delayed test

/l/ vocalisation (45%)

The overall rate of /l/ vocalisation in the delayed test increased (+9%). Listen Only has the highest rate (67%) followed by Listen and Speak (54%). Traditional still has the lowest rate (25%). For three tokens in Traditional, there seems to be deviations in vowel production accompanying /l/ vocalisation. For this group, the process appeared in delayed repetition task. The Traditional Teacher's realisations were considered correct for this test item.

Reduction

This process reappeared in data tokens from Listen Only by two of the four speakers found in the post-test (04, a literate male and 19, an illiterate female). The other two were participants 11 and 17. In the delayed test, 11 produced /-lk/ correctly and 17 could not be recruited for recording. The other two tokens were produced by two speakers from Traditional. This process emerged only in the delayed test for Traditional.

/l/ vocalisation and substitution

A single token exhibits the two processes of C1 /l/ vocalisation and C2 backing at the same time. This is produced by a female speaker from Traditional.

Vowel epenthesis

An individual case of vowel epenthesis reappeared in the delayed test by the same participant. Similar to his token in the post-test, the epenthetic vowel is inserted after the cluster.

9.3.4 */-ft/*

Test items in this category include the word 'left' only. Realisations of the female Traditional Teacher are shown in table 9.7.

Table 9.7 Traditional Teachers' realisations of /-ft/ tokens

item	task	process	target-likeness	realisation
left	read aloud	correct	target	[lɛft^h]
	picture-naming	correct	target	[lɛft^h]

Post-test

The overall rate of correct /-ft/ realisations in the post-test is about 82%. That is the highest found in all coda clusters. Listen and Speak had the highest rate of correct realisations of the /-ft/ cluster at around 91% followed by Listen Only at a close 89%. Traditional has a lower rate of correct realisations at over 70% but still considered high in comparison with other cluster types.

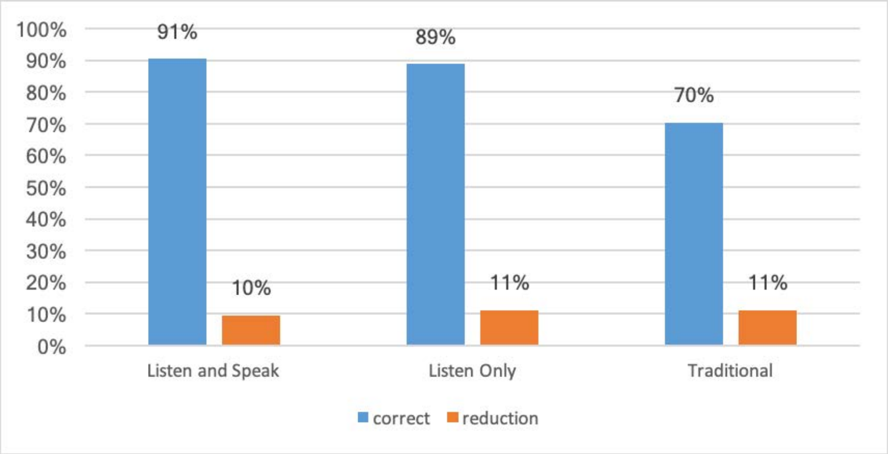


Figure 9.13 Percentages of processes for /-ft/ according to instruction type during the post-test

Generally, for /-ft/, a common process within the post-test is reduction (10.6%).

Reduction (10.6%)

Only a few tokens (seven) where the elision of either segment from the cluster took place. In all but one of the cases, C1 is retained. Four of these, exhibited compensatory lengthening. All of the participants reducing /-ft/ clusters have some degree of literacy.

Other processes that took place less frequently and exclusively within Traditional are vowel epenthesis (3%) and substitution and syllable epenthesis (1.5% each).

Vowel epenthesis

Only two tokens exhibited vowel epenthesis and they are produced by the same speaker in the picture-naming and delayed repetition tasks.

Substitution

This speaker substituted C2 /t/ with /s/.

Syllable epenthesis

This speaker added an extra segment /s/ and used an epenthetic vowel to avoid the new three-consonant cluster.

This following case could not be assigned to any of the above processes. However, this realisation was exhibited in data tokens relevant to the test item ‘list’. It is possible that the participant confused this test item with ‘list’.

Delayed test

The overall rate of correct /-ft/ realisations in the delayed test increased about a further 13% maintaining its highest rate of correct realisations of /-ft/. However, in the delayed test, Traditional had the highest rate of correct realisations (96%), followed by Listen Only (94%) and finally Listen and Speak (92.3%).

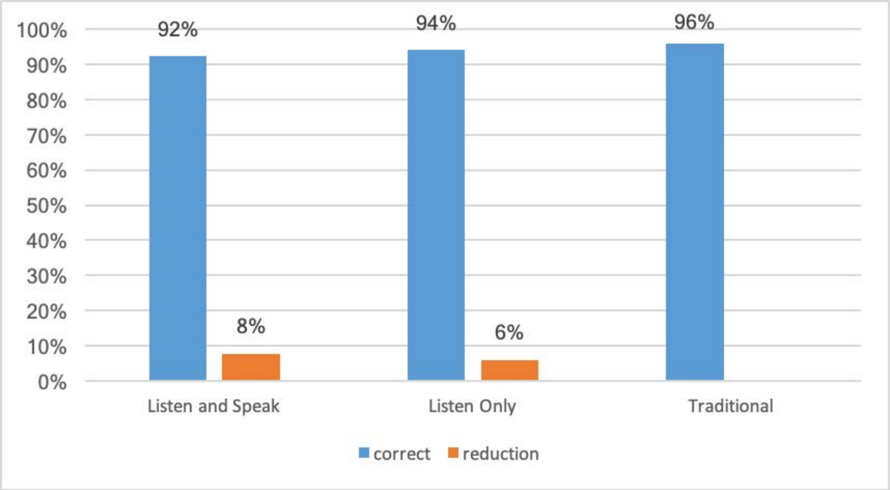


Figure 9.14 Percentages of processes for /-ft/ according to instruction type during the delayed test

Only three of the /-ft/ clusters were produced incorrectly in the delayed test, one from each instruction group. The experimental groups produced a reduced token each, whereas Traditional produced a cluster with vowel epenthesis.

Reduction

In the only two cases witnessed, C2 was elided. C1 showed compensatory lengthening by speaker from Listen Only.

Vowel epenthesis

Only one case of vowel epenthesis appeared for /-ft/ during the delayed test. It was realised as such by a male participant from Traditional in the picture-naming task.

9.3.5 */-st/*

Test items in this category include the words ‘breast’, ‘list’ and ‘vest’. Realisations of the female Traditional Teacher are shown in table 9.8.

Table 9.8 Traditional Teachers' realisations of /-st/ tokens

item	task	process	target-likeness	realisation
breast	read aloud	correct	non-targetlike	[b̥ɾɛstʰ]
	picture-naming	correct	non- target	[bɾɛstʰ]
list	read aloud	correct	target	[lɪstʰ]
	picture-naming	correct	targetlike	[lɪstʰ]
vest	read aloud	correct	non-targetlike	[vɛ̹stʰ]
	picture-naming	correct	non-targetlike	[vɛ̹s̬t]

Post-test

The rate of correct /-st/ realisations in the post-test is about 80%. Listen and Speak has the highest rate of correct realisations of the /-st/ cluster at above 88% followed by Listen Only at about 83%. Traditional has the lowest rate of correct realisations at 70%.

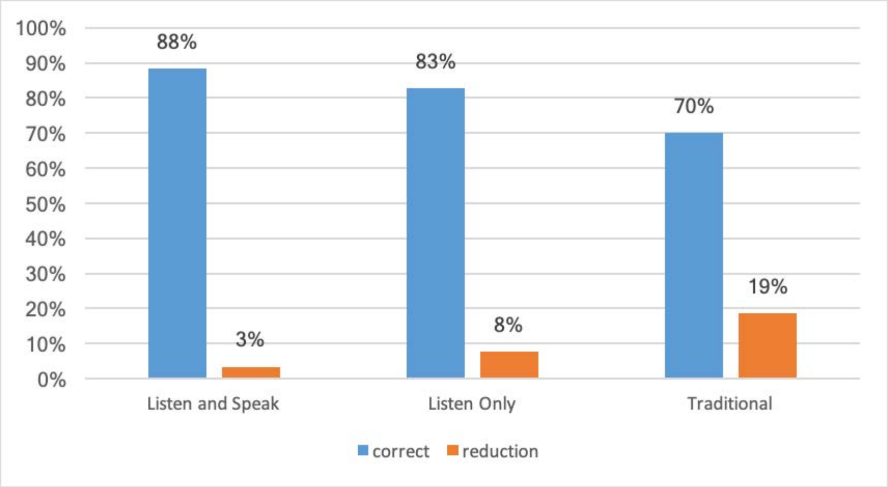


Figure 9.15 Percentages of processes for /-st/ according to instruction type during the post-test

Generally, for /-st/, the most common process within the post-test is reduction (10.4%).

Reduction (10.4%)

In a few cases, the elision of C2 results in compensatory lengthening in C1. Participants having some level of literacy made such lengthening. Reduction was most common amongst Traditional (18.6%). The next relatively high rate was amongst Listen Only (7.7%) and Listen and Speak had the lowest rate (3.3%).

Other less frequent processes are epenthesis (vowel 3.3% and consonant 1.1%)

substitution and metathesis simultaneously (2.7%), metathesis (1.6%), and substitution (0.5%).

These processes do not appear in every instruction group.

Vowel Epenthesis (3.3%)

This process appeared exclusively in Traditional mostly in the picture-naming task by participants with some degree of literacy.

Metathesis and substitution (2.7%)

Aside from metathesis, C1 remained unsubstituted in all examples. The alveolar plosive /t/ was replaced with /k/ in all cases except 'vest' [vːɛʔs] when it was substituted with a glottal stop. This process appeared in data from experimental groups during the delayed repetition task only.

Metathesis (1.6%)

Very few instances of metathesis appeared in data from experimental groups during the delayed repetition task for the test items 'vest' and 'list'. The two tokens from Listen Only come from the same speaker.

Consonant epenthesis (1.1%)

The data items in 9.24 appear to have consonant epenthesis. However, there could be more to it than mere epenthesis.

	instruction	item	task	realisation
9.24	Listen and Speak	breast	delayed repetition	[b̥ɹɛʔst]
	Traditional	list	delayed repetition	[lɪkst]

Substitution (0.5%)

The alveolar plosive in 'list' is substituted for /k/. It is similar to the substitutions when metathesis took place in the previous examples in **Metathesis and substitution**.

There was a case, which could not be assigned to any of the above processes produced by a male speaker from Traditional 'vest', [sɪt].

Delayed test

The rate of correct /-st/ realisations in the delayed test was still high and increased over a further 4%.

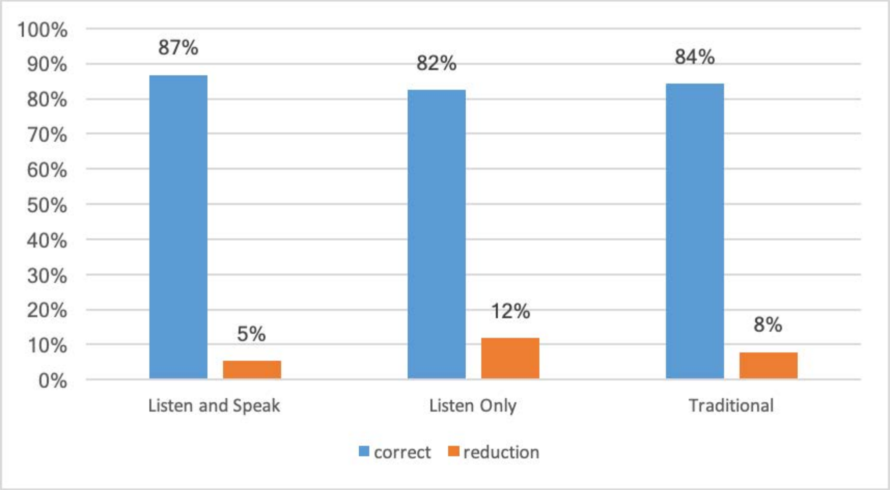


Figure 9.16 Percentages of processes for /-st/ according to instruction type during the delayed test

Reduction (8.6%)

Reduction was still the most common process during the delayed test at an overall rate of 8.6%. However, it occurred more in Listen Only (11.8%) than Traditional (7.8%) and Listen and Speak (5.3%).

No reduction was recorded for the memory recall tasks. In fact, there were only four analysable tokens produced in the picture-naming task, three of which had a /-st/ cluster that was marked correct and one having vowel epenthesis. All memory recall data for this cluster were produced by Traditional only.

Vowel epenthesis

Again, vowel epenthesis was even less frequent with a single occurrence in the delayed test, also within Traditional only in the picture-naming task vest [vɛs̬ɨt].

Other less frequent processes include consonant epenthesis, metathesis and substitution combined and metathesis only.

Consonant epenthesis

This process appeared in the post-test and reappeared again in the delayed test.

Only this time, different participants exhibited such behaviour.

Metathesis and substitution (2.1%)

Again, metathesis and substitution combined had a low rate of occurrence. Only this time, the process was restricted to the test item ‘list’ but also occurred exclusively within the experimental groups. Alongside the process of metathesis, C1 remained unsubstituted in all examples. The alveolar plosive /t/ was replaced with /k/ in all cases. Once again, this process was evident in the delayed repetition task only. The two speakers (07 and 08) from Listen and Speak are twins.

Metathesis (1.4%)

Metathesis still had a low rate during the delayed test. The same participant (19) produced metathesis again in the delayed test only this time it is ‘vest’ only. A participant from Traditional produced a metathesised coda cluster and no participants from Listen and Speak produced any metathesised coda clusters.

A token produced by a female participant from Traditional during the delayed repetition could not be assigned to any of the above processes, ‘vest’, [vḁɾɪst˺]. It could be confused with wrist.

9.3.6 */-nt/*

Test items in this category include the words ‘paint’, ‘point’ and ‘mount’. Realisations of the female Traditional Teacher are shown in table 9.9.

Table 9.9 Traditional Teachers' realisations of /-nt/ tokens

item	task	process	target-likeness	realisation
mount	picture-naming	correct	non-targetlike	[ma̠wənt^{h}]
	read aloud	correct	non-targetlike	[ma̠wɨnt^{h}]
point	picture-naming	correct	non-targetlike	[pɔɪnt^{h}]
	read aloud	correct	non-targetlike	[pɔjɨnt^{h}]
paint	picture-naming	wrong	non-targetlike	[poɪnt]
	read aloud	wrong	non-targetlike	[poˑɪnt^{h}]

Post-test

The overall rate of correct /-nt/ realisations in the post-test is 42%. Traditional has the highest rate of correct realisations of the /-nt/ cluster at around 56% followed by Listen and Speak (41%). Listen Only has the lowest rate of correct realisations at only 27%.

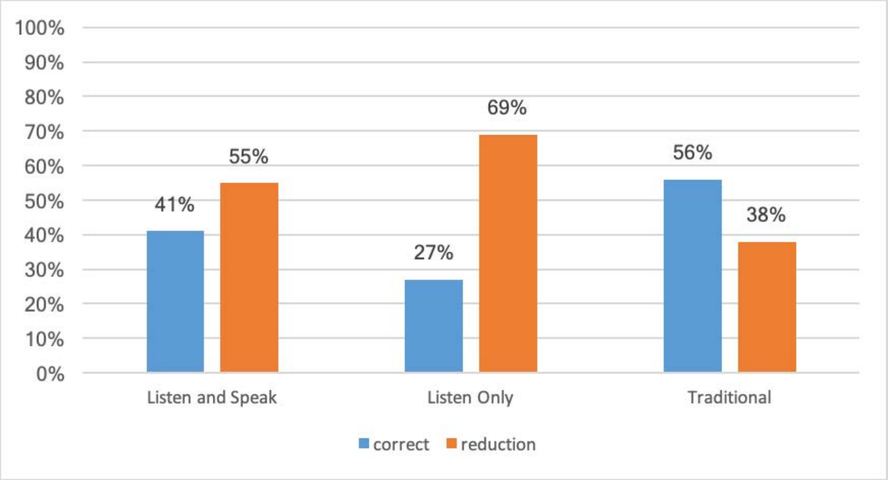


Figure 9.17 Percentages of processes for /-nt/ according to instruction type during the post-test

Generally, for /-nt/, the most common process within the post-test is reduction (54%).

Reduction (54%)

Reduction is most evident in Listen Only instruction (69%). Followed by that in Listen and Speak instruction (55%). The group with the lowest reduction rate is Traditional (38%). In the majority of cases, C1 was the segment undergoing elision. Three exceptions exist; mount [maʊn], paint [p^{h}en], and point [pɔɪn].

Other less frequent processes are substitution (2% only three tokens altogether occurring in the delayed repetition task; two from Listen and Speak and one from Traditional), deletion (1% a single token from Traditional occurring during the delayed repetition task), and finally consonant epenthesis (1% two tokens appearing in Listen Only only).

Substitution (2%)

In the first example in 9.25, there seems to be a case of consonant harmony. The second example from Listen and Speak is possibly confused with test item 'belt'.

	instruction	item	task	realisation
9.25	Listen and Speak	mount	delayed repetition	[mɑmtʰ]
	Listen and Speak	paint	delayed repetition	[pʰeltʰ]
	Traditional	paint	delayed repetition	[penʧ]

Consonant epenthesis (1%)

The first example in 9.26 shows that C1 is elided, C2 is maintained and an additional consonant /s/ is inserted after the underlying cluster. The second realisation also shows that /s/ is inserted after the cluster.

	instruction	item	task	realisation
9.26	Listen Only	mount	delayed repetition	[mɑʊts]
	Listen Only	point	delayed repetition	[pɔwɪnts]

Deletion (1%)

In both of the following realisations in 9.27, the coda cluster /-nt/ is elided for the items 'mount' and 'point'.

	instruction	item	task	realisation
9.27	Traditional	mount	delayed repetition	[mʌu]
	Traditional	point	delayed repetition	[pʰoɪ]

Delayed test

The overall rate of correct /-nt/ realisations in the delayed test increased a further 2%. Across groups, Listen Only's performance improved going from 27% in the post-test to 40% correct realisations rate in the delayed test. Listen and Speak and Traditional's performance deteriorated in the delayed test reaching only 38% and 52% respectively. Nonetheless, Traditional still had the best performance in the realisation of /-nt/.

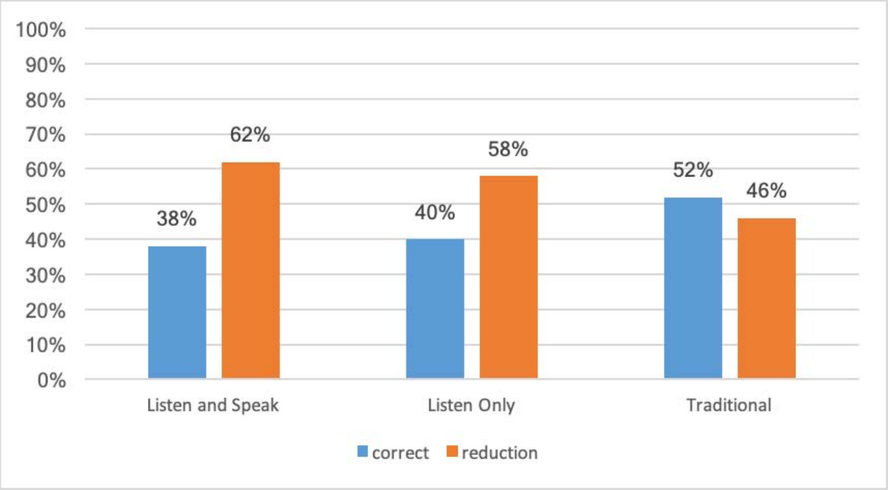


Figure 9.18 Percentages of processes for /-nt/ according to instruction type during the delayed test

Reduction (54%)

Over half the /-nt/ clusters were reduced in the delayed test. Amongst the three groups, Listen and Speak had the highest rate (62%). Listen Only had the next highest rate (58%), whereas Traditional had the lowest reduction rate (46%).

In the majority of the tokens containing /-nt/, C1 /n/ is elided. Two out of seventy-three tokens had C2 elided instead. These are 'point' [p̬h̬ɔn] and [pʰɔɪɨn]. 'Mount' seems to be the test item that is mostly reduced within the three test items.

Substitution

The production made by Listen Only exhibited substitution within C1. /n/ was replaced by a glottal stop. The second case is actually metathesis on a word level rather than the level of the coda cluster. Listen and Speak did not exhibit any substitution.

9.3.7 */-nd/*

Test items in this category include the words 'pound' and 'sound'. Realisations of the female Traditional Teacher are shown in table 9.10.

Table 9.10 Traditional Teachers' realisations of /-nd/ tokens

item	task	process	target-likeness	realisation
pound	read aloud	correct	non-targetlike	[pʰa̱wɨnd]
	picture-naming	correct	non-targetlike	[pa̱wɨndʰ]
sound	read aloud	correct	non-targetlike	[sa̱wənd]
	picture-naming	correct	non-targetlike	[sa̱wɨnd]

Post-test

The overall rate of correct /-nd/ realisations in the post-test is 68% (over two thirds). Listen and Speak has the highest rate of correct realisations of the /-nd/ cluster at 85% followed by Traditional at 65%. Listen Only has the lowest rate of correct realisations at 54%.

Generally, for /-nd/, the most common process within the post-test is reduction (26%) with Traditional displaying higher rates at 35%, followed by Listen Only at 32.4%, whilst Listen and Speak showing only 10% of tokens containing /-nd/ reduced. Other less frequent processes include deletion (3%), which appears only in the data for Listen and Speak and Listen Only at 5% each and substitution (2.6%) that appears in data from Listen Only only.

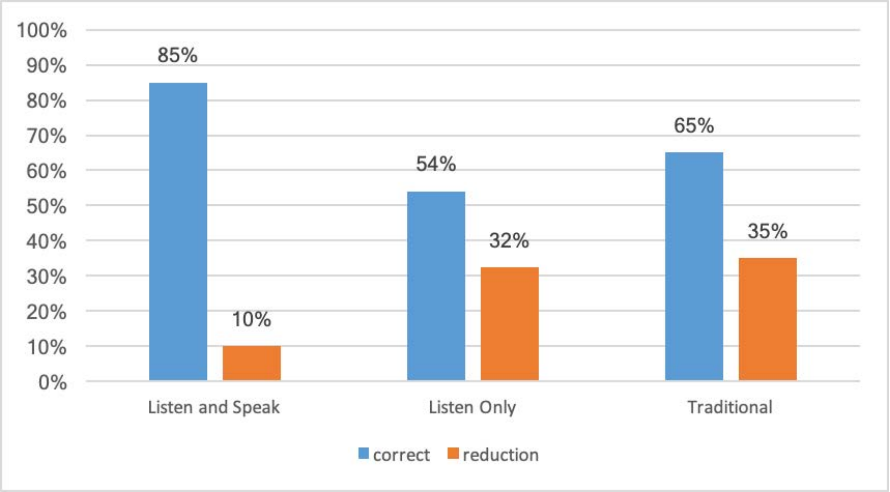


Figure 9.19 Percentages of processes for /-nd/ according to instruction type during the post-test

Reduction

Listen and Speak and Listen Only showed a pattern of C1 elision in all but one token each (three out of four and twelve out of thirteen respectively), whereas Traditional showed a somewhat reverse pattern that is C2 elision in ten out of thirteen /-nd / tokens.

Deletion

The rate of deletion was very low in the post-test for this cluster. It was exclusive to the experimental groups and was restricted to the test item ‘sound’ only. Only one token was yielded by picture-naming; the other two were found in the delayed repletion task.

Substitution (and reduction)

Substitution was seen in Listen Only and Traditional by three different speakers. In each case, C1 was replaced by a velar nasal and C2 by a velar stop. Thus, in both substitutions manner was preserved and change is in place of articulation from alveolar to velar in a process so-called ‘backing’. In the example from Traditional, the coda cluster was also reduced to one segment.

Delayed test

The overall rate of correct /-nd/ realisations in the delayed test increased (+6%). Group comparisons show that it increased for Listen and Speak, Listen Only and

Traditional (+3%, +8% and +11% respectively). The order of highest to lowest correct rate remained the same in the delayed test.

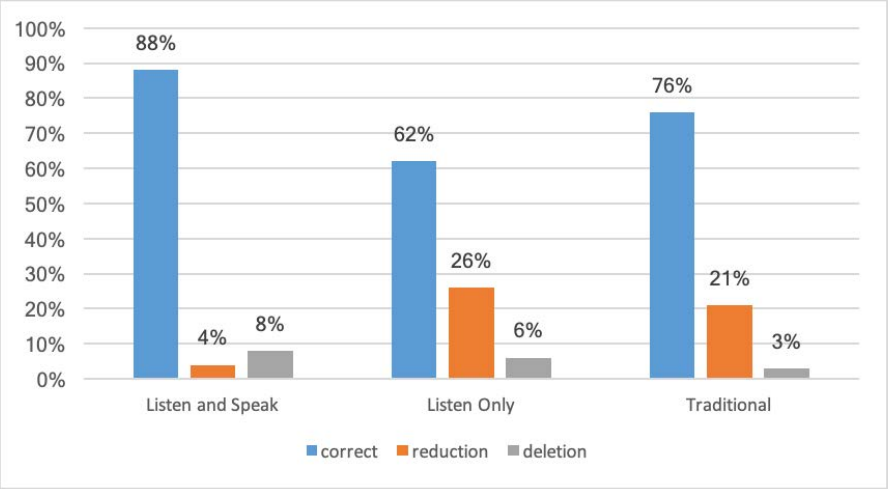


Figure 9.20 Percentages of processes for /-nd/ according to instruction type during the delayed test

Reduction

Increase in rate of correct realisations of /-nd/ resulted in a drop in rate of reduction (-9%) with Listen Only having the highest rate of reduction, followed by Traditional and finally Listen and Speak having the lowest reduction rate with only one /-nd/ token reduced. The elision of C1 /n/ was more common than that for C2, where only three tokens; one from each group exhibiting the elision of C2. All but one participant showing this pattern had some degree of literacy.

Deletion

A less frequent process that appeared in the delayed test for this cluster type was deletion. One speaker from each group seems to demonstrate this process. Speakers (08) and (02) from Listen and Speak and Listen Only respectively seem to repeat this process for the two test items.

Reduction and substitution

A participant from Listen Only deleted a segment from the cluster and substituted the other with [θ]. It could also be the case that the speaker confused ‘pound’ with ‘path’.

Substitution

A participant from Listen Only replaced C1 /n/ with [ŋ] in the delayed repetition task.

9.3.8 */-nz/*

Test items in this category include the word ‘jeans’ only. Realisations of the female Traditional Teacher are shown in table 9.11.

Table 9.11 Traditional Teachers’ realisations of /-nz/ tokens

item	task	process	target-likeness	realisation
jeans	picture-naming	correct	non-targetlike	[ʤ̥ɛnz̥]
	read aloud	correct	non-targetlike	[ʤ̥ɛnz]

Post-test

The overall rate of correct /-nz/ realisations in the post-test was 60%. Listen and Speak had the highest rate of correct realisations of the /-nz/ cluster at 62%. Listen Only and Traditional do not fall far behind at 59% and 58% respectively.

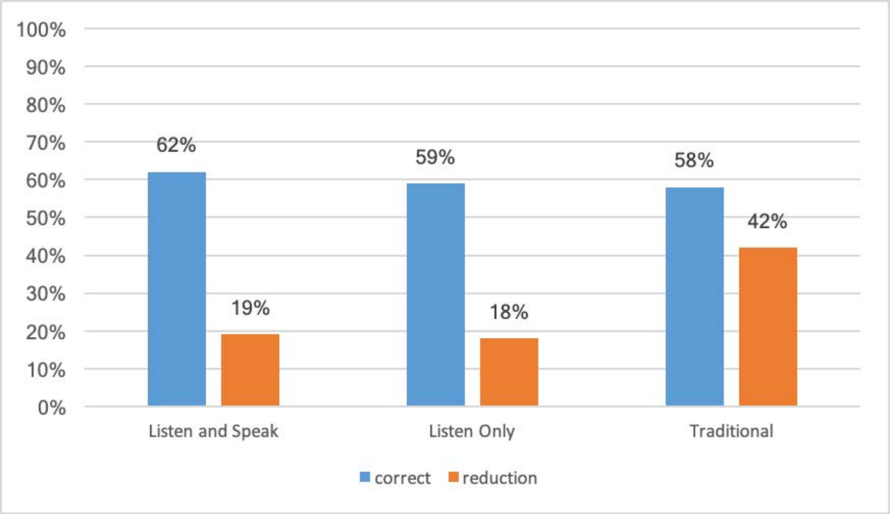


Figure 9.21 Percentages of processes for /-nz/ according to instruction type during the post-test

Generally, for /-nz/, the most common process within the post-test is reduction (27%). Other less frequent processes include substitution (6%), consonant epenthesis, consonant epenthesis and substitution simultaneously; reduction and substitution simultaneously and finally syllable epenthesis at (2%) each.

Reduction

Traditional had the highest rate of reduction in /-nz/ clusters. Listen and Speak and Listen Only had similar rates to each other (19% and 18% respectively). For Listen and Speak and Listen Only, C2 /z/ was elided. In Traditional, there were two exceptions to this. Two out of ten tokens showed elision in C1.

Substitution

C2 is replaced with a stop in a process called ‘fricative stopping’. One token is an exception, showing what seems to be more like consonant harmony.

Miscellaneous

Other less frequent processes took place within the experimental groups only. For example, ‘jeans’ [ʤiːnts], a speaker from Listen and Speak inserted an additional consonant [t] between members of the cluster, not to mention the final devoicing. It could be the final devoicing which elicited the excrescent [t]. Similarly, with [ʤ̥iːmps], a speaker from Listen Only inserted [p] and a process of regressive assimilation of place of articulation can be seen for C1. As with [ʤ̥iːl], it is an example of reduction and substitution that was also present in the delayed test by a different participant from a different group. C2 is elided and C1 /n/ is substituted with a homorganic lateral.

Delayed test

The overall rate of correct /-nz/ realisations in the delayed test increased (+15%) compared to the post-test. Listen and Speak is still taking the lead (92%), followed by Listen Only (82%) and finally Traditional (56%).

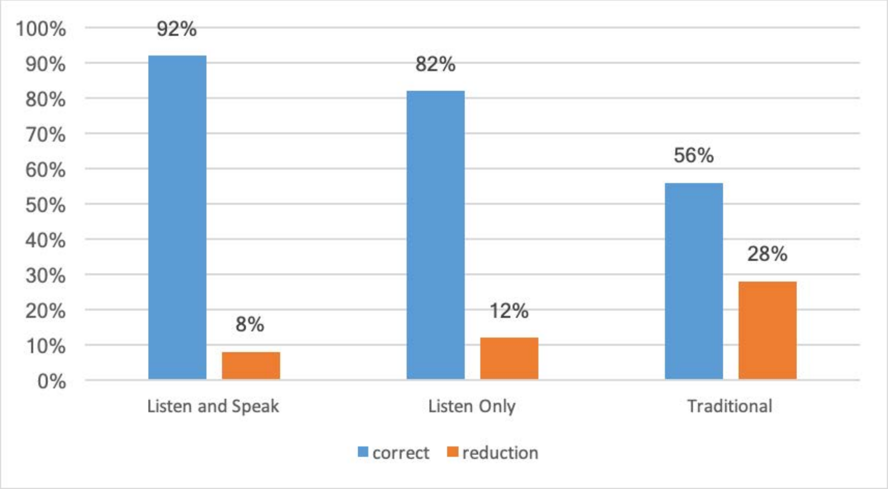


Figure 9.22 Percentages of processes for /-nz/ according to instruction type during the delayed test

Reduction

Reduction was not very high for this cluster in the delayed test. It had a decrease (17%) compared to the post-test (27%) that negatively correlates with the rise in correct realisations. For Traditional, only two thirds of the tokens reduced in the post-test were also reduced in the delayed test. The same for Listen Only; only two thirds of the tokens reduced in the post-test were also reduced in the delayed test. As for Listen and Speak, only half the tokens reduced in the post-test were also reduced in the delayed test. The pattern of elision shows the reverse for the experimental groups. This time C1 /n/ is elided and C2 /z/ is retained. However, Traditional shows the reverse. C1 is retained instead.

Miscellaneous

Other less frequent processes took place mostly within Traditional. For example, ‘jeans’ [ʤeːnð], the speaker substituted C2 /z/ for [ð] in what seems to be fronting. Another example, [ʤiːnʃ], shows a case of consonant harmony. In [d͡siːnʃ], the speaker seems to exhibit metathesis but across syllable constituents rather than segments of the cluster. One realisation from Listen Only showed an example of reduction and substitution in the token [ʤ̥iːl], whereby C2 is elided and C1 /n/ is substituted with a homorganic lateral. A different speaker in Listen and Speak produced a similar realisation in the post-test.

9.3.9 */-mp/*

Test items in this category include the words ‘jump’ only. Realisations of the female Traditional Teacher are shown in table 9.12.

Table 9.12 Traditional Teachers’ realisations of /-mp/ tokens

item	task	process	target-likeness	realisation
jump	picture-naming	correct	non-targetlike	[ʤʌmp̬]
	read aloud	correct	target	[ʤʌmpʰ]

Post-test

The overall rate of correct /-mp/ realisations in the post-test is 46%. Traditional has the highest rate of correct realisations of the /-mp/ cluster at 67% followed by Listen Only at 44%. Listen and Speak has the lowest rate of correct realisations at 27%.

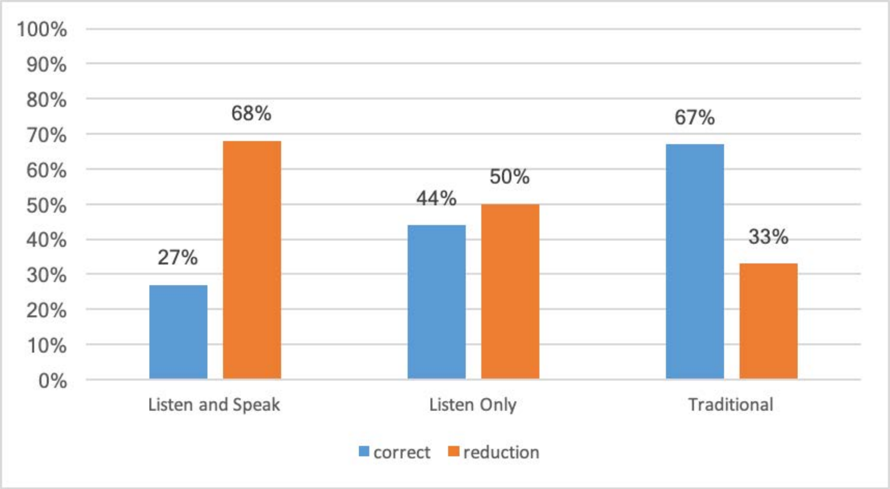


Figure 9.23 Percentages of processes for /-mp/ according to instruction type during the post-test

Generally, for /-mp/, the most common process within the post-test is reduction (51%). Another less frequent process is substitution (3%).

Reduction

The group with the highest rate of reduction was Listen and Speak with over two thirds of /-mp/ analysable tokens reduced (68%). Followed by this is Listen Only with half the /-mp/ tokens reduced. Traditional had the lowest reduction rate (33%). Elision alternated between C1 and C2. In Listen and Speak, eight out of fifteen tokens showed elision of C1. In Listen Only all tokens exhibited elision of C1. In Traditional, all tokens exhibited elision of C2.

Substitution

Substitution was much less frequent for /-mp/. It appeared for two tokens only one from each experimental group; none from Traditional. In Listen and Speak, the participant replaced C2 /p/ with [k]. In Listen Only, the participant replaced both C1 /m/ and C2 /p/ with [ŋ] and [k] respectively.

Delayed test

The overall rate of correct /-mp/ realisations in the delayed test decreased (-8%). By examining groups individually, the drop is most evident in Listen Only and Traditional (-14% and -42% respectively). Data are obtained from delayed repetition tasks only as memory recall yielded no tokens for this test item.

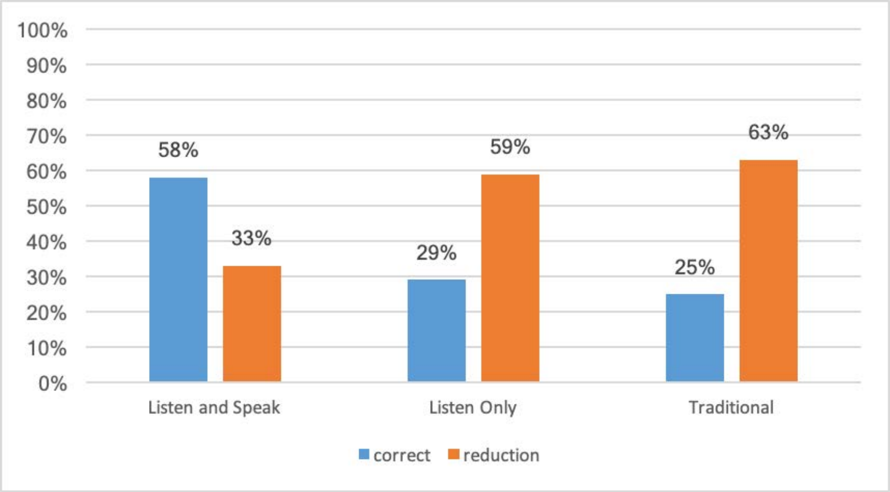


Figure 9.24 Percentages of processes for /-mp/ according to instruction type during the delayed test

Reduction

The group with the highest rate of reduction was Traditional (33%) that is 30% more than the post-test. Followed by this is Listen Only with 59% of the /-mp/ tokens reduced. Listen and Speak had the lowest reduction rate (33%) that is 35% lower than the post-test.

In Listen and Speak, two speakers elided C1 and maintained C2 whilst the other two showed the opposite trend. However, their behaviour in comparison with the post-test did not change with this regard. In Listen Only six out of ten tokens exhibited elision of C1 whereas the remaining four elided C2. In Traditional, contrary to their behaviour in the post-test, all tokens exhibited elision of C1.

Substitution

Two tokens in the delayed test exhibited substitution; one from Listen Only and the other from Traditional. The participant from Listen Only replaced C1 /m/ with /n/. The other speaker substituted both C1 /m/ and C2 /p/ for /n/ and /t/ respectively.

She maintained manner and substituted place of articulation.

Consonant epenthesis

This process appears in one token only for /-mp/. A male participant from Listen Only inserted /k/ after the cluster.

Reduction and substitution

A male participant from Listen and Speak reduced the cluster to one segment and replaced it with the glottal stop. It is assumed that it was C2 that was replaced given /p/ and /ʔ/ share manner of articulation.

9.4 Dental Fricatives

9.4.1 */θ-/*

Test items in this category include the words 'thin', 'thigh', and 'thumb'. Productions of the female Traditional Teacher for the Traditional Teaching are shown in table 9.13.

Table 9.13 Traditional Teachers' realisations of /θ-/ tokens

item	task	process	target-likeness	realisation
thigh	read aloud	reinforcement	non-targetlike	[tθʌf]
thigh	picture-naming	correct	non-targetlike	[θʌf]
thin	read aloud	reinforcement	non-targetlike	[tθɪ̝n]
thin	picture-naming	stopping	non-targetlike	[tɪ̝n]
thumb	read aloud	correct	non-targetlike	[θʌːmb]
thumb	picture-naming	reinforcement	non-targetlike	[tθʌːmb]

Post-test

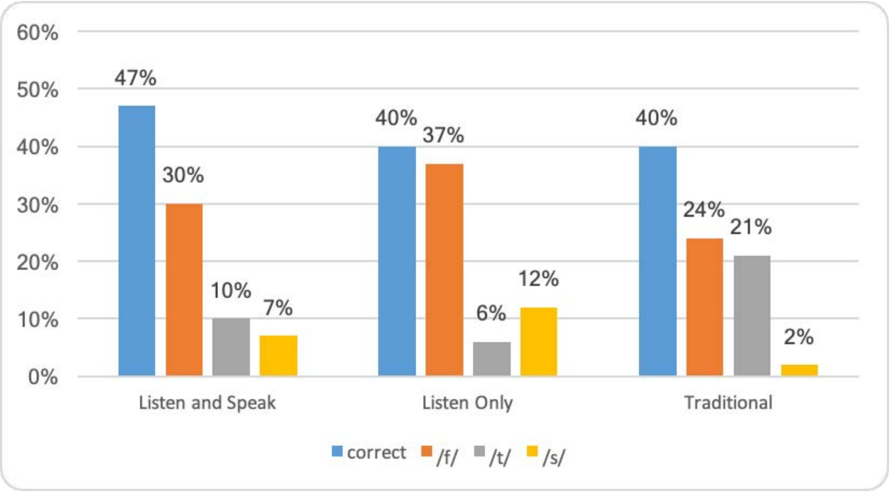


Figure 9.25 Percentages of processes for /θ-/ according to instruction type during the post-test

Overall, the rate of correct realisations of /θ-/ in the post-test was rather low (42%). The chart shows that Listen and Speak had the highest rate of correct realisations, whereas Listen Only and Traditional had lower but matching rates of correct realisation. The target sound underwent processes of various substitutions and consonant reinforcement, though the latter was less common. The most common substitution was /f/ (fronting) (30%), followed by /t/ (stopping) (12%) and /s/ (6%). Fronting was the most frequent in Listen Only and the least frequent in Traditional. Stopping appeared mostly in the Traditional Teaching as per their Traditional Teacher. It was also restricted to the test items ‘thigh’ and ‘thin’ in the post-test. Two tokens from Listen Only exhibited stopping that was accompanied by pharyngealisation. Consider the following examples:

	instruction	item	task	realisation
9.28	Listen and Speak	thin	delayed repetition	[sɪn]
	Listen Only	thumb	delayed repetition	[fʌm]
	Listen Only	thigh	delayed repetition	[tˤɑɪ]
	Traditional	thigh	picture-naming	[tɑɪ]
	Traditional	thumb	delayed repetition	[t͡θʌmb̥]

Delayed test

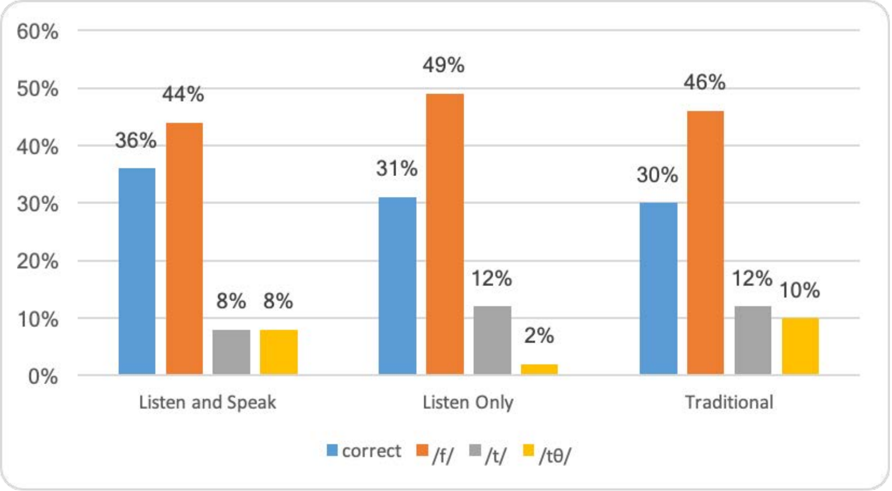


Figure 9.26 Percentages of processes for /θ-/ according to instruction type during the delayed test

Overall, the rate of correct realisations of /θ-/ in the delayed test was again low (32%), lower than that for the post-test (-10%). Figures dropped slightly in Listen and Speak and Traditional (-4% and -7% respectively). However, they increased in Listen Only (+6%). The target sound underwent processes similar to those in the post-test. However, this time, the most common substitution overall was /f/ (fronting) (46%). Listen and Speak, Listen Only, and Traditional had similar rates. Other less frequent substitutions were /t/ (stopping) (11%). Listen and Speak had slightly lower rates than the other groups. /s/ substitutions were infrequent and the quality of the /s/ was dental. Consonant reinforcement was also low (6%) with Listen Only demonstrating the lowest rate. Consider the following examples:

9.29

instruction	item	task	realisation
Listen Only	thumb	delayed-repetition	[fʌm]
Traditional	thigh	delayed-repetition	[taɪ]
Listen and Speak	thin	delayed-repetition	[tθɪn]
Listen and Speak	thin	delayed-repetition	[s̪ɪn]

9.4.2 */-θ/*

Test items in this category include the words ‘bath’, ‘broth’, ‘moth’, ‘mouth’, ‘path’, and ‘tooth’. Realisations of the female Traditional Teacher are shown in table 9.14.

Table 9.14 Traditional Teachers' realisations of /-θ/ tokens

item	task	process	target-likeness	realisation
bath	read aloud	correct	non-targetlike	[baːθː]
bath	picture-naming	correct	non-targetlike	[baːθ]
broth	read aloud	correct	non-targetlike	[bɾoθ]
broth	picture-naming	reinforcement	non-targetlike	[bɹot̪θ]
moth	read aloud	correct	target	[moθ]
moth	picture-naming	correct	non-targetlike	[moθː]
path	read aloud	reinforcement	non-targetlike	[paːtθ]
path	picture-naming	correct/stopping	non-targetlike	[paːθ/paːt̪]
tooth	read aloud	stopping	non-targetlike	[tuːtʰ]
tooth	picture-naming	stopping	non-targetlike	[tuːtʰ]

Post-test

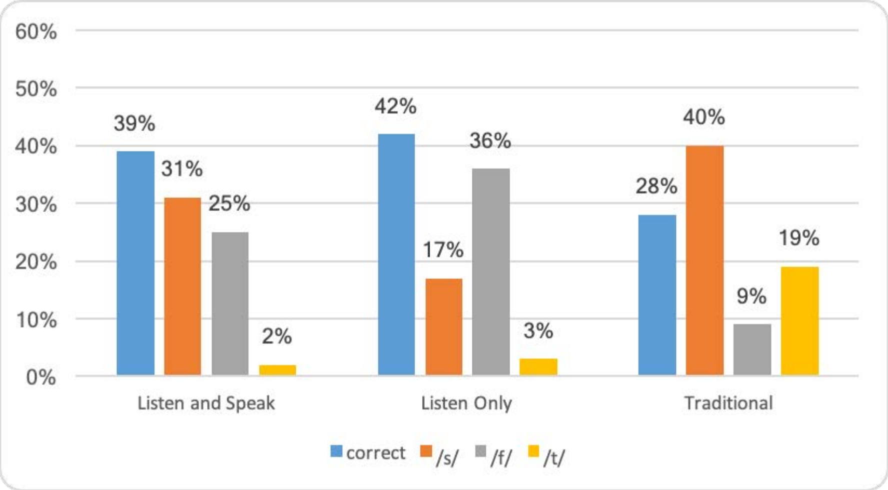


Figure 9.27 Percentages of processes for /-θ/ according to instruction type during the post-test

Overall, the rate of correct realisations of /-θ/ during the post-test was low (36%). Figure 9.27 shows that Listen Only had the highest rate of correct realisations followed by Listen and Speak, whereas Traditional had the lowest rate. The most common substitution was /s/ (31%) with Traditional exhibiting the highest rate followed by Listen and Speak. Listen Only had the lowest rate of /s/ substitution. This is because, for this group, the token was mostly substituted with /f/ (fronting) (36%), which is the next most common substitution. Traditional illustrated the lowest substitution rate for /f/. Stopping was the least frequent substitution. However, for the Traditional

Teaching condition, it was considerably more common that fronting given their Traditional Teacher's demonstration of such a process in her production. Consonant reinforcement was infrequent and was restricted to one token from each group. Deletion was demonstrated in two tokens for this test both from Traditional in the test items 'bath' and 'path' by two different speakers. Consider the examples below:

	instruction	item	task	realisation
9.30	Listen and Speak	broth	delayed repetition	[b̥rɔs̪]
	Listen Only	tooth	delayed repetition	[tʰʊˑf]
	Traditional	bath	read aloud	[baːtʰ]
	Listen and Speak	moth	delayed repetition	[mɔtθ]
	Traditional	bath	delayed repetition	[bɑː]

Delayed test

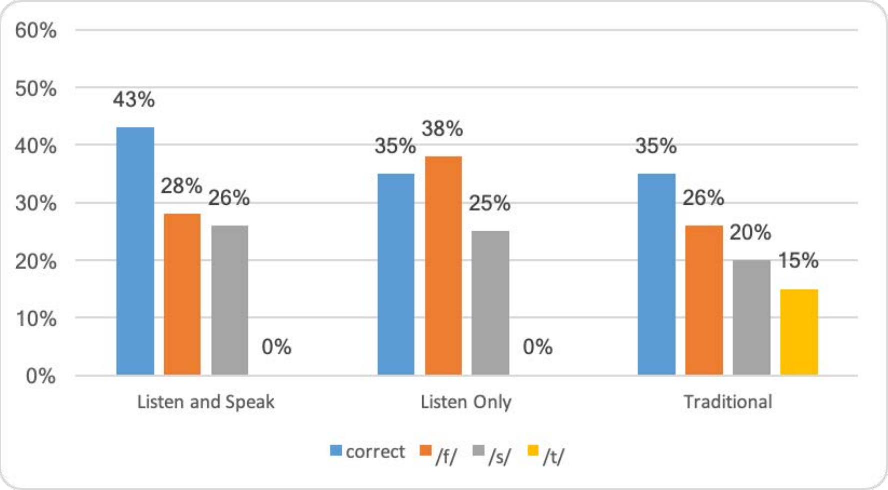


Figure 9.28 Percentages of processes for /-θ/ according to instruction type during the delayed test

Overall, the rate of correct realisations of /-θ/ in the delayed test was rather low (37%). The chart shows that Listen and Speak had the highest rate of correct realisations, whereas Listen Only and Traditional had lower but matching rates of correct realisation. The target sound underwent processes of various substitutions and consonant reinforcement. The most common substitution was /f/ (fronting) (31%) and /s/ (23%). Listen Only had the highest rate of fronting, whereas Listen and Speak and Traditional had similar and lower rates of such a process. Substitution with /s/ appeared a quarter of the total tokens for the experimental groups, whilst for

Traditional, it was demonstrated in only a fifth of the total tokens. It was mostly illustrated in the test item path and was sometimes more dental than alveolar. Additionally, substitution with /t/ (stopping)(15%) appeared exclusively in Traditional as with their Traditional Teacher. It was most evident in the picture-naming task. Consonant reinforcement /tθ/ appeared in one token from Listen and Speak only. Consider the following examples:

9.31

instruction	item	task	realisation
Listen Only	bath	delayed repetition	[b̥ɑːf]
Listen and Speak	broth	delayed repetition	[b̥ɹɔs̪]
Traditional	moth	picture-naming	[motʰ]
Listen and Speak	tooth	delayed repetition	[tuːt͡θ]

9.5 Diphthongs

Diphthong categories include /aɪ/, /ɪə/, /ɔɪ/, /aʊ/, /eɪ/, and /əʊ/. Test items for diphthongs include the words ‘pile’, ‘pine’, ‘rice’, ‘thigh’, ‘mount’, ‘mouth’, ‘pound’, ‘round’, ‘sound’, ‘boil’, ‘coin’, ‘point’, ‘paint’, ‘rain’, ‘tail’, ‘waist’, ‘cold’, ‘roll’, ‘rose’, ‘beard’, ‘fear’, and ‘hear’. Realisations of the female Traditional Teacher are shown in table 9.15.

Table 9.15 Traditional Teachers’ realisations of /aɪ/ tokens

item	task	process	target-likeness	realisation
pile	read aloud	v2-gliding	non-targetlike	[pa̧jɨl]
pile	picture-naming	v2-gliding	non-targetlike	[p̬a̠jɨl]
pine	read aloud	v2-gliding	non-targetlike	[pa̠jɨn]
pine	picture-naming	v2-gliding	non-targetlike	[pa̠jɨn]
rice	read aloud	v2-gliding	non-targetlike	[ɾajɨs̬]
rice	picture-naming	correct	non-targetlike	[ɾaɪs̬]
thigh	read aloud	wrong	non-targetlike	[tθʌf]
thigh	picture-naming	wrong	non-targetlike	[θʌf]

Table 9.16 Traditional Teachers' realisations of /aʊ/ tokens

item	task	process	target-likeness	realisation
mount	read aloud	v2-gliding	non-targetlike	[ma̠wɨntʰ]
mount	picture-naming	v2-gliding	non-targetlike	[ma̠wəntʰ]
mouth	read aloud	v2-gliding	non-targetlike	[ma̠wɨtʰ]
mouth	picture-naming	correct	non-targetlike	[ma̠ʊtʰ]
pound	read aloud	v2-gliding	non-targetlike	[pʰa̠wɨnd]
pound	picture-naming	v2-gliding	non-targetlike	[pa̠wɨndʰ]
round	read aloud	v2-gliding	non-targetlike	[ɾa̠wɨnd]
round	picture-naming	v2-gliding	non-targetlike	[ɾa̠wənd]
sound	read aloud	v2-gliding	non-targetlike	[sa̠wənd]
sound	picture-naming	v2-gliding	non-targetlike	[sa̠wɨnd]

Table 9.17 Traditional Teachers' realisations of /ɔɪ/ tokens

item	task	process	target-likeness	realisation
boil	read aloud	v2-gliding	non-targetlike	[b̥ɔjɨl]
boil	picture-naming	v2-gliding	non-targetlike	[bɔjɨl]
coin	read aloud	v2-gliding	non-targetlike	[kɔjɨn]
coin	picture-naming	v2-gliding	non-targetlike	[kɔjɨn]
point	read aloud	v2-gliding	non-targetlike	[pɔjɨntʰ]
point	picture-naming	correct	non-targetlike	[poɪnt]

Table 9.18 Traditional Teachers' realisations of /eɪ/ tokens

item	task	process	target-likeness	realisation
paint	read aloud	wrong	non-targetlike	[poˑɪnt^{h}]
paint	picture-naming	wrong	non-targetlike	[pɔɪnt^{h}]
rain	read aloud	monophthongisation	non-targetlike	[ɾeːn]
rain	picture-naming	monophthongisation	non-targetlike	[ɾeːn]
tail	picture-naming	monophthongisation	non-targetlike	[teˑl]
waist	read aloud	monophthongisation	non-targetlike	[weˑst^{h}]
waist	picture-naming	monophthongisation	non-targetlike	[weˑst^{h}]

Table 9.19 Traditional Teachers' realisations ofəʊ/ tokens

item	task	process	target-likeness	realisation
cold	read aloud	monophthongisation	non-targetlike	[kɔːɫd]
cold	picture-naming	monophthongisation	non-targetlike	[kɔːɫd]
roll	read aloud	monophthongisation	non-targetlike	[ɾuːlː]
roll	picture-naming	monophthongisation	non-targetlike	[ɾuːlːə]
rose	read aloud	monophthongisation	non-targetlike	[ɾuːz]
rose	picture-naming	monophthongisation	non-targetlike	[ɾoˑz]

Table 9.20 Traditional Teachers' realisations of /ɪə/ tokens

item	task	process	target-likeness	realisation
beard	read aloud	monophthongisation	non-targetlike	[beːɾd]
beard	picture-naming	monophthongisation	non-targetlike	[beːɾd]
fear	read aloud	monophthongisation	non-targetlike	[feːɾ]
fear	picture-naming	monophthongisation	non-targetlike	[feːɾ]
hear	read aloud	correct	non-targetlike	[hiˑəɾ]
hear	picture-naming	correct	non-targetlike	[hiːəɾ]

In considering phonological processes that diphthongs underwent in the data, there were many cases where no process took place, yet words were marked as non-targetlike for segments of the test item other than the diphthong. During the qualitative coding for such data these were labelled correct and thus correct does not imply target-likeness as it only applies to a part of the word and target-likeness refers to the word as a whole. The following sections explore data in terms of the various phonological patterns observed for diphthongs.

Correct realisations

Figures 9.29, 9.30, and 9.31 show that the diphthong /aɪ/ was overall the least challenging amongst all the diphthongs for the learners in both tests, 80% in the post-test and 77% in the delayed test. Followed by this were /ɪə/, /ɔɪ/, /aʊ/, /eɪ/, and finally /əʊ/ was the most challenging having the lowest rate of correct realisations, 30.9% in the post-test and 34.3% in the delayed test.

However, when the rate of correct realisations was examined individually that is across instruction types, Listen and Speak struggled with /ɪə/ more than /ɔɪ/ and /aʊ/ during the post-test. See figure 9.29.

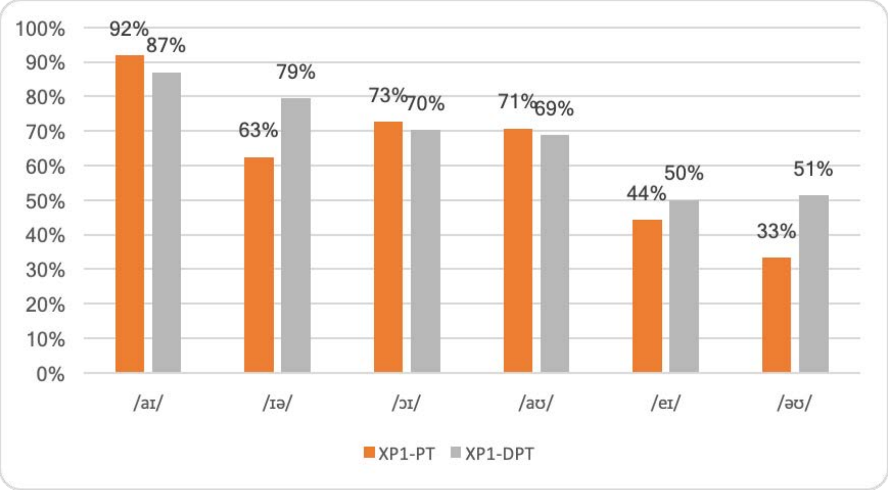


Figure 9.29 Percentages of correct realisations across diphthongs in Listen and Speak during post-test (PT) and delayed test (DT)

Listen Only struggled with /aɪ/ more than /ɪə/ during the delayed test. See figure 9.30.

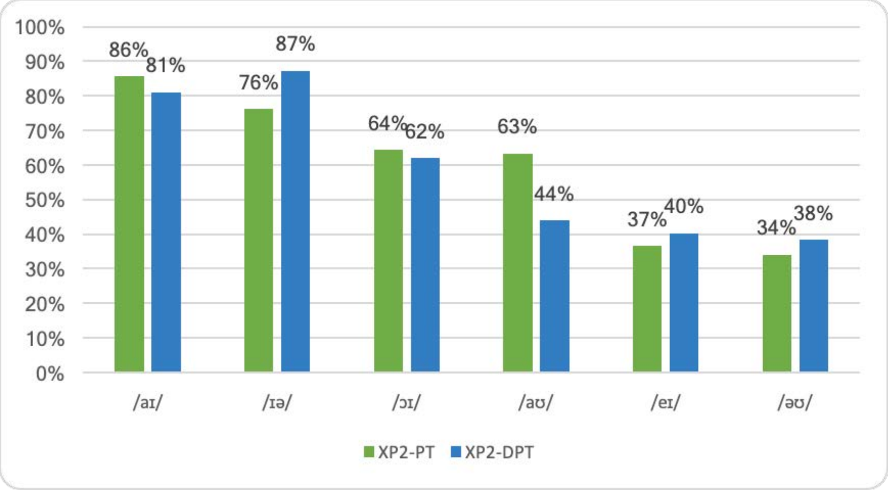


Figure 9.30 Percentages of correct realisations across diphthongs in Listen Only during post-test (PT) and delayed test (DT)

Traditional instruction seems to have struggled more with /ɪə/ than /ɔɪ/. This does not come as a surprise as the Traditional Teacher produced all words containing /eɪ/ with monophthongisation and those containing /ɪə/ exhibited two patterns in her realisations. The test item 'hear' was realised as /hɪˑəɾ/, hence coded as correct whilst the test items 'fear' and 'beard' were realised as /feːɾ/ and /beːɾd/ respectively. The latter two were coded as wrong realisations in the analysis, yet, surprisingly, none of the participants from Traditional matched their Traditional Teacher's production in any of the tasks. See figure 9.31.

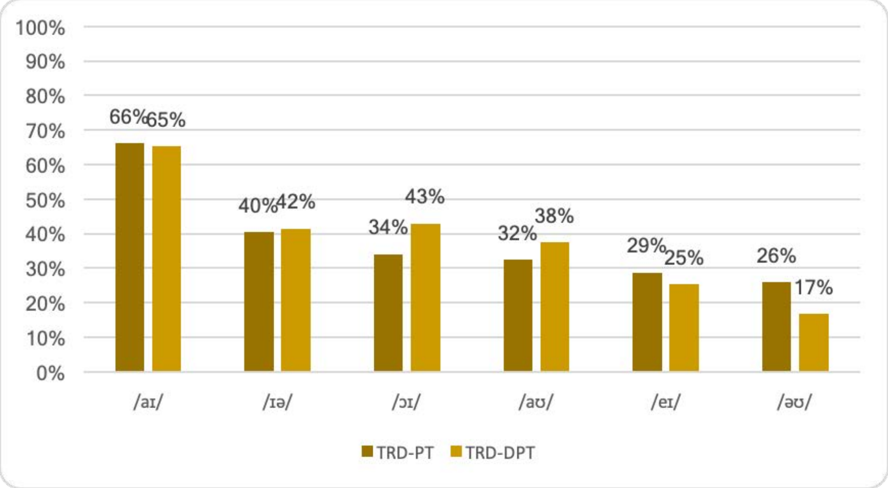


Figure 9.31 Percentages of correct realisations across diphthongs in Traditional during post-test (PT) and delayed test (DT)

Overall, the Listen and Speak seemed to have the highest rate of correct realisations amongst the three instruction types in both tests. Two exceptions to this generalisation were the rate of correct realisations of the diphthong /ɪə/ during both tests. During the post-test, the Listen Only scored considerably higher for this diphthong (76.2%) compared with the Listen and Speak (62.5%) and the same was true for the delayed test (87.2% and 79.4% respectively). Also, the score for the diphthong /əʊ/ was slightly higher for the Listen Only (34%) than the Listen and Speak (33.3%).

A comparison of overall performance within time that is between post-test and delayed test revealed a mixture of improvement and deterioration across the various diphthongs for all instruction types. It is worth-noting that generally speaking, improvement outweighed deterioration, which was only marginal for the two diphthongs /aɪ/ (-3.8%) and /aʊ/ (-6.7%) for Listen and Speak and Traditional. However, for Listen Only, there was a consistent trend of deterioration in the rate of correct realisations with the exception of the diphthong /əʊ/, which seemed to increase (+3.2%) during the delayed test. Listen and Speak exhibited a rise in rate of correct realisations in the delayed test in the diphthongs /aɪ/, /eɪ/ and /ɪə/ (+0.8%, +7.1% and +12.2% respectively) and a fall in /ɔɪ/ and /aʊ/ (-4.7% and -2.4% respectively).

As for Traditional, there was generally a marginal decline in the rate of correct realisations for the diphthongs /aɪ/, /əʊ/ and /eɪ/ (-3%, -2.2% and -1.5% respectively). For /ɪə/, the case was not straight forward as there was a slight decline in performance in the delayed repetition elicitation task (-1.4%) but a rise in performance during the picture-naming elicitation task (+4.7%) with an overall decrease in performance. Performance for /aʊ/ also showed an increase during the picture-naming task. However, the overall trend for this diphthong was an increase (+1.2%). For /ɔɪ/, again there was an increase mainly if not exclusively during the delayed repetition task (+8.7%) and a decrease in the picture-naming (-33.3%). The latter percentage can be misleading as only three tokens containing /ɔɪ/ were actually made, only one of which was coded correct during the post-test, whilst in the delayed test, only one production was made and it was not correct. For /eɪ/, nineteen tokens were made in the post-test and twelve in the delayed test and neither was coded as correct.

The read aloud task seems to be the scarcest of tasks in terms of correct realisations. During the post-test, only four tokens were produced all by Traditional. In the delayed test, only two tokens were produced by Listen and Speak. Followed by this is the picture naming. The task, which yielded most correct realisations, was the delayed repetition elicitation task. Overall, participants did better in delayed production elicitation tasks than in read aloud or picture naming.

Processes

The main phonological processes observed across the data were monophthongisation, substitution and gliding of the second member of the diphthong.

Substitutions generally entail features such as lowering, raising, rounding, unrounding, fronting and backing. In other cases, though less common, one or both members of the diphthong are transformed in more than one way.

Additionally, because test words in the stimuli are produced in isolation, segments tend to be lengthened. The same applies to production data, as it is not produced in spontaneous speech or carrier phrases. For this reason, in the production data, when either member of the diphthong is lengthened, it is not regarded as a lengthening process. However, it should be noted that the lengthening of the second member of the diphthong in test items such as 'mount' and 'mouth', for example, is most recurrent when the coda is deleted serving as compensatory lengthening across syllable constituents. However, it does not behave in the same way for words containing /ɔɪ/. There does not seem to be a compensatory lengthening effect when the coda is missing or reduced.

9.5.1 *Monophthongisation*

Figure 9.32 and 9.33 show that the diphthongs /əʊ/ and /eɪ/ were overall the ones that underwent monophthongisation the most during both tests, 57% and 55.3% respectively in the post-test and 57.5% and 52.5% respectively in the delayed test. Followed by these are /ɪə/, /ɔɪ/ and /aʊ/. /aɪ/, on the other hand, did not undergo any monophthongisation.

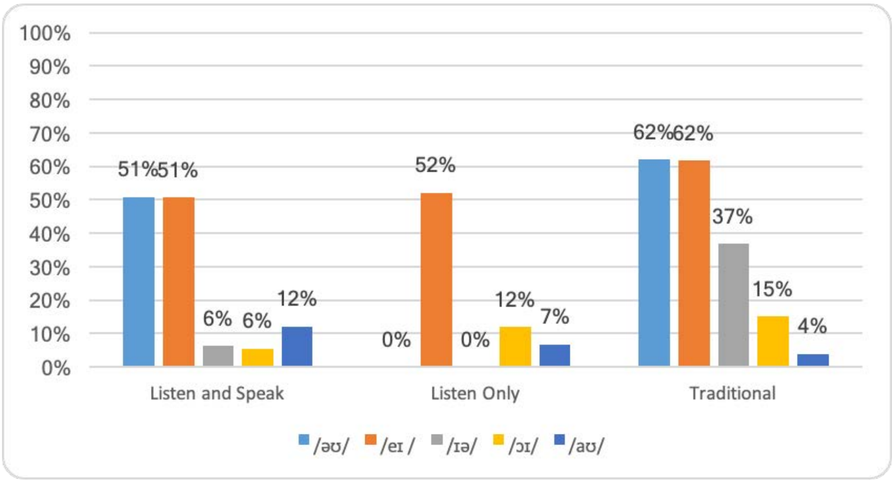


Figure 9.32 Percentages of monophthongisation for diphthongs according to instruction type during the post-test

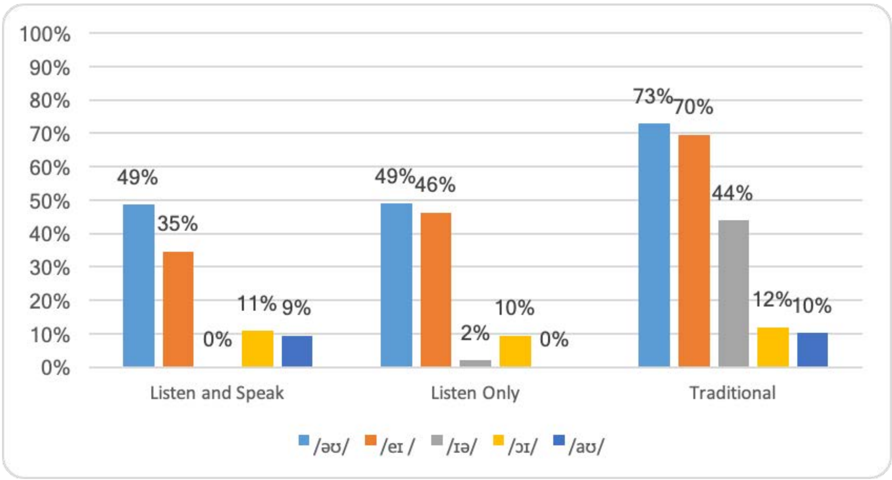


Figure 9.33 Percentages of monophthongisation for diphthongs according to instruction type during the delayed test

Because of the occasional limited tokens especially for the memory recall tasks, percentages for such cases were sometimes disproportionate to those of the delayed repetition task. This has made comparisons meaningless. Therefore, comparisons are restricted to ten or more cases of monophthongisation per task for each diphthong. Traditional instruction learners seems to have the highest rate of monophthongisation.

9.5.2 *Substitution*

Substitution was generally more common for the diphthong /ɪə/ during the post-test. Just over a quarter (25.9%) of the produced test items containing this diphthong were substituted. The majority of substitutions in this diphthong involve the second member being transformed into a front low vowel /a/. This can also be referred to as vowel advancing and lowering. See figure 9.34.

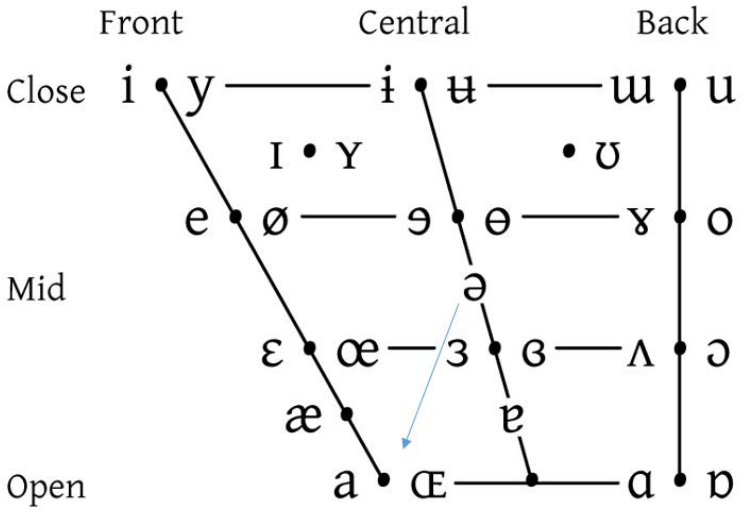


Figure 9.34 Substitution process in /ɪə/

Followed by this are /ɔɪ/ (16.7%), /aʊ/ with roughly a tenth of the realisations substituted (11.3%), /eɪ/ (7.8%), /əʊ/ (6.7%) and finally /aɪ/ (4.2%) that were marginally substituted.

In the delayed test, figures decrease for the diphthongs /ɪə/, /ɔɪ/ and /əʊ/ by 11.3%, 6%, and 4.5% respectively, whilst increasing slightly for /aʊ/, /eɪ/ and /aɪ/ by 2.4%, 1.3% and 5.1% respectively.

However, when the rate of substitution is examined individually that is across instruction types, for each test, figures (see figure 9.35 and 9.36) reveal varying results for each diphthong. For /ɪə/, during the post-test, figures for the Listen and Speak and Listen Only were close (31.3% and 31.9% respectively). For Traditional, this is much lower at 22.8%. During the delayed test, although figures decrease, they do so at varying rates. Listen Only seems to have the lowest rate of substitution 12.5%, followed by Traditional 14.6%, leaving Listen and Speak with comparatively the highest rate at 17.6%.

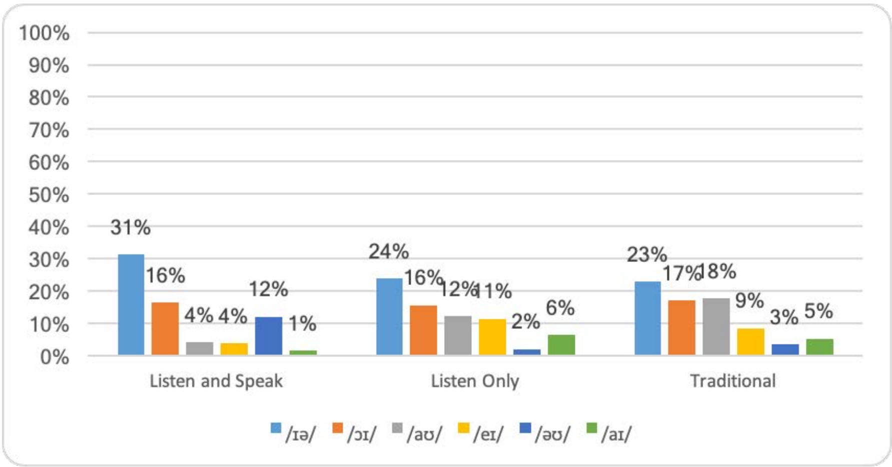


Figure 9.35 Percentages of substitution for diphthongs according to instruction type during the post-test

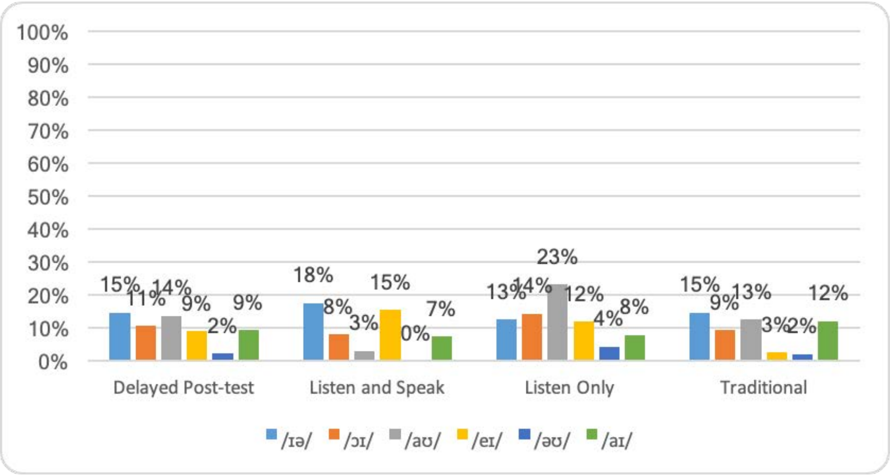


Figure 9.36 Percentages of substitution for diphthongs according to instruction type during the delayed test

For /ɔɪ/, again during the post-test figures are close for all the groups. However, these vary in the rate of decline during the delayed post-test. Whilst Listen and Speak and Traditional scored a close percentage of 8.1% and 9.3% respectively, Listen Only exhibits a higher rate of substitution with 14.3%. For /aʊ/, Listen and Speak seems to have the lowest rate of substitution during both tests with 4% during the post-test and 3.1% during the delayed test. Thus, it is fair to say that the rate remained constant with time. For Listen Only, however, it increased from 12.2% during the post-test to 23.2% during the delayed postest, making it the group with the highest substitution rate for the diphthong /aʊ/ during the delayed test. Traditional on the other hand exhibited a slight decrease from 17.6% to 12.5% in substitution rate during the delayed test.

For /eɪ/, figures 9.35 and 9.36 exhibit mixed results for the instruction types. Despite having the highest rate of substitution during the post-test, it seems to have maintained a steady rate of substitution within time with 11.3% during the post-test and 11.9% during the delayed test. Listen and Speak, conversely, showed a considerable rise from 3.8% to 15.4% in substitution rate across tests. It started with the group yielding the least substitution rate for this diphthong to the one with highest substitution during the delayed test. For Traditional, there was a fall in the rate of substitution from 8.5% to 2.5% making it the group with the least occurrences of substitution for this diphthong.

For /əʊ/, the substitution rate was generally quite low with the exception of Listen and Speak during the post-test at 12%. Consider the examples in 9.32

9.32

item	realisation
cold	[kʰaud̥]
	[kʰʌʊ]
roll	[b̥ɾaːʊl]
	[rɔ̜ːʊ]
	[ɾə̹ʊɫ]
rose	[ˈɔɹuə̹z̥]
	[ɰɔɑz]

However, substitution diminishes in the delayed test as neither of 39 instances were substituted.

For /aɪ/, substitution rates were also rather low especially during the post-test. During the post-test, there is a uniform rise in rates by 7%, 1.6% and 11.9% for Listen and Speak, Listen Only, and Traditional respectively. In other words, Traditional had the highest increase rate followed by Listen and Speak, leaving Listen Only with the least increase in rate of substitution for this diphthong.

As for performance across tasks, it is very difficult to have a meaningful comparison, since data from memory recall tasks is very scarce especially for the experimental groups.

9.5.3 *Gliding*

Gliding here is for cases where the second member of the diphthong transforms into a glide. Usually /ɪ/ transforms into /j/ and /ʊ/ into /w/. It is assumed that as a result, the glide is followed by a vowel acting as a repair mechanism for the resulting cluster of consonants. Consider the examples in 9.33.

9.33

diphthong	item	realisation
/aɪ/	pine	[pʰajɨn]
	rice	[ɾajɨs̬]
/aʊ/	pound	[pʰɑwɨnd̥]
	mouth	[mawɨs]
/ɔɪ/	boil	[bojɨɫ]
	coin	[kʰɔjɨn]
/eɪ/	waist	[wejɨsˡ]
	rain	[ɔreːjɨm]
/əʊ/	rose	[ɹə̠wɨz̥ˡ]
	cold	[kɔwɨl]

Figures 9.37 and 9.38 indicate that gliding seems to be most common in /aʊ/ with a rate of 26.1% during the post-test and 27.8% during the delayed test. The diphthongs with the next highest rate of gliding are /ɔɪ/ and /aɪ/ who come close with a rate of 16% and 15% respectively during the post-test. In the delayed test, figures for the former rose to 20.7% while the latter slightly decreased to 12%. The diphthongs with a marginal rate of gliding are /əʊ/ and /eɪ/ at 5.5% and less than one percent respectively in the post-test. These figures have risen slightly in the delayed test to 6% and 1.5% respectively.

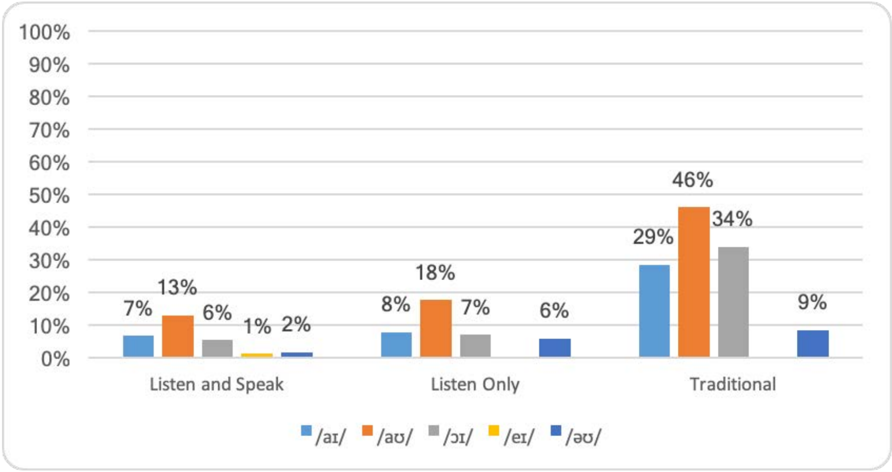


Figure 9.37 Percentages of gliding for diphthongs according to instruction type during the post-test

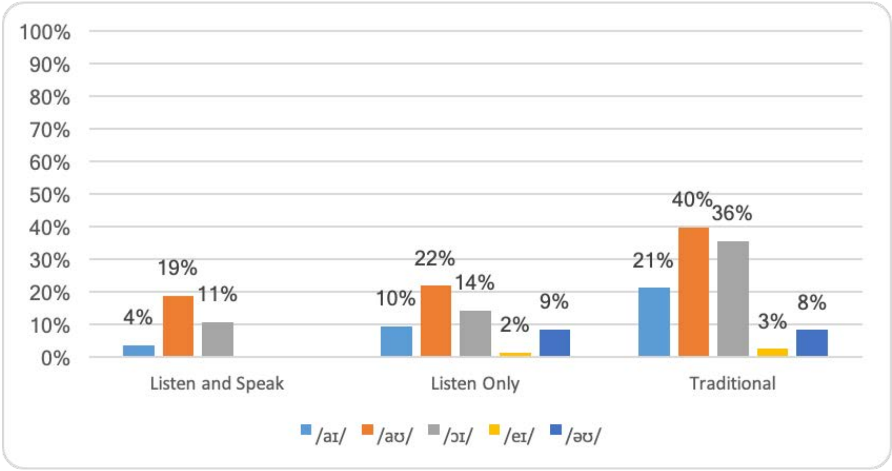


Figure 9.38 Percentages of gliding for diphthongs according to instruction type during the delayed test

9.6 Rhotic Approximants

9.6.1 *Rhotic approximant word-initially*

Test items in this category include the words ‘rain’, ‘read’, ‘red’, ‘rice’, ‘rip’, ‘rob’, ‘roll’, ‘root’, ‘rose’, ‘round’, ‘wrap’, ‘wreck’, ‘wrist’, and ‘wrong’. Realisations of the female Traditional Teacher exhibited a consistent pattern of substitution with the native language counterpart /ɾ/. Consider the Traditional Teacher’s realisations in table 9.21

Table 9.21 Traditional Teacher's realisations of /ɹ-/ tokens

item	task	process	target-likeness	realisation
rain	read aloud	substitution	non-targetlike	[ɾeːn]
rain	picture-naming	substitution	non-targetlike	[ɾeːn]
read	read aloud	substitution	non-targetlike	[ɾiːdːə]
read	picture-naming	substitution	non-targetlike	[ɾiːd]
red	picture-naming	substitution	non-targetlike	[ɾɛd]
rice	read aloud	substitution	non-targetlike	[ɾa̠jɨs̬]
rice	picture-naming	substitution	non-targetlike	[ɾa̠ɪs]
rip	read aloud	substitution	non-targetlike	[ɾɪp]
rip	picture-naming	substitution	non-targetlike	[ɾɪp̬]
rob	read aloud	substitution	non-targetlike	[ɾɔb̌]
rob	picture-naming	substitution	non-targetlike	[ɾɔ̜b]
roll	read aloud	substitution	non-targetlike	[ɾuːlː]
roll	picture-naming	substitution	non-targetlike	[ɾuːlːə]
root	read aloud	substitution	non-targetlike	[ɾuːtʰ]
root	picture-naming	substitution	non-targetlike	[ɾuːtʰ]
rose	read aloud	substitution	non-targetlike	[ɾuːz]
rose	picture-naming	substitution	non-targetlike	[ɾoˑz]
round	read aloud	substitution	non-targetlike	[ɾa̠wɨnd]
round	picture-naming	substitution	non-targetlike	[ɾa̠wənd]
wrap	read aloud	substitution	non-targetlike	[ɾa̠pʰ]
wrap	picture-naming	substitution	non-targetlike	[ɾa̠p]
wreck	read aloud	correct	targetlike	[ɹɛǩʰ]
wreck	picture-naming	substitution	non-targetlike	[ɾɛk]
wrist	read aloud	correct	targetlike	[ɹɪstʰ]
wrist	picture-naming	substitution	non-targetlike	[ɾɪstʰ]
wrong	read aloud	substitution	non-targetlike	[ɾɔŋg]
wrong	picture-naming	substitution	non-targetlike	[ɾɔŋg]

When approaching data tokens from /ɹ-/, four themes emerge; tokens that were realised correctly; tokens that were realised correctly but preceded by a consonant, a vowel or a CV syllable (dummy syllable); tokens that were substituted only; and finally tokens that were substituted and also preceded by a consonant, a vowel or a CV syllable (dummy syllable). For illustration purposes, first, percentages for each of the above themes are presented in 9.39 and 9.40. Later, each of the themes will be delved into separately.

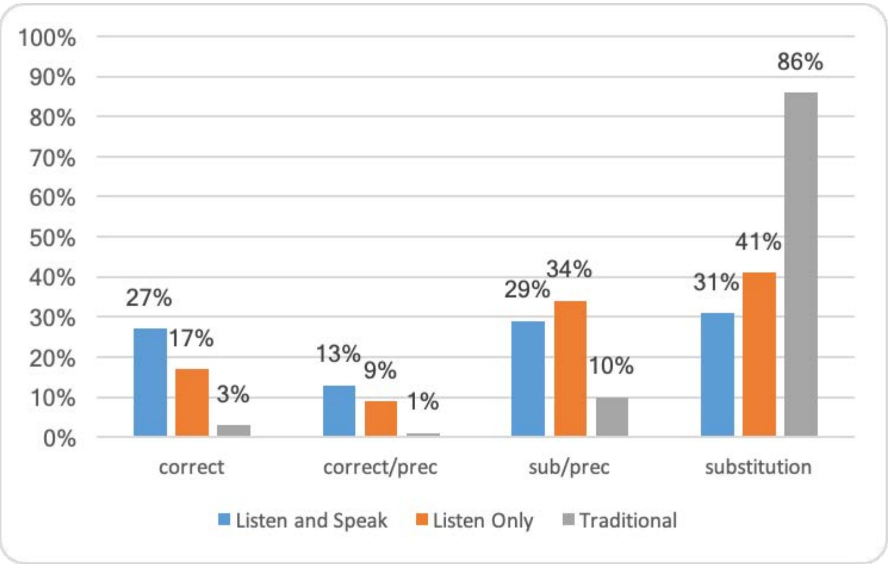


Figure 9.39 Percentages of processes for /ɹ-/ according to instruction type during the post-test

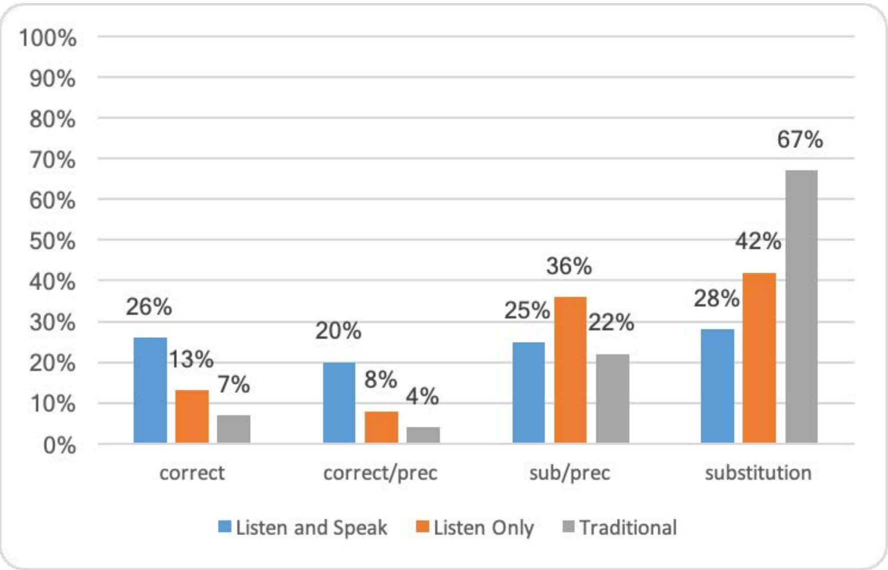


Figure 9.40 Percentages of processes for /ɹ-/ according to instruction type during the delayed test

The figures in 9.39 and 9.40 show that overall, the rate of correct realisations was considerably low in both tests. Listen and Speak, nonetheless, outperformed Listen Only and Traditional. The latter exhibited the lowest rate of correct realisations reflecting their Traditional Teacher's realisations. In the delayed test, the order of performance remains the same. However, rates dropped slightly in the experimental groups, whereas for Traditional, the rate doubled. Quite a similar pattern of performance was observed for tokens that were correct but preceded by either a vowel, a consonant or a CV sequence.

This process category, however, is slightly less frequent than the percentages of correct realisations without preceding segments.

Substitution, on the other hand, negatively correlated with the above two patterns. Overall, substitution was the most frequent process of all during both tests. Traditional had the highest rate of substitution in both tests but it dropped by a quarter in the delayed test (-19%). Listen and Speak had the lowest substitution rate in both tests, though it dropped slightly in the delayed test (-3%). Listen Only was higher by one third of Listen and Speak during the post-test. In the delayed test, substitution increased slightly (-1%).

Finally, substitution accompanied by a preceding segment was the second most frequent process in /ɹ-/. This process was most evident in Listen Only (34% and 36% in the post-test and delayed test respectively). It was also common in Listen and Speak (29% and 25% in the post-test and delayed test respectively), though it was less common than Listen Only. However, unlike Listen Only, its frequency declined in the delayed post-test (-4%). Traditional, on the other hand, demonstrated a rather low occurrence in the post-test. In the delayed-test, the rate rose considerably (+12%) yet did not exceed that of the experimental groups.

The following sections explore variations and the nature of epenthetic segment(s) and alterations within each of the above processes.

Unsubstituted but preceded by a V, C, or CV added

It was illustrated above in 9.39 and 9.40 that a small proportion of the data tokens were realised correctly, that is as a rhotic approximant, but were preceded by a segment that was either a vowel, a consonant, or both. This section deals with the nature of these segments and their rate of occurrence.

A close examination of the data in 9.41 and 9.42 reveals that the process was mainly observed in the experimental groups in the post-test. Traditional only demonstrated this pattern in the delayed post-test. Listen and Speak had a higher occurrence rate than Listen and Speak in both tests. Traditional had the lowest frequency of this pattern. It was also observed that within the sub-patterns, the most frequent preceding segment was the vowel in both tests. The second most frequent was the consonant, also in both tests.

Epenthetic Vowel

Data illustrating an epenthetic vowel shows that the most frequent vowel was a rounded back vowel, mostly /ɔ/ or /o/ occurring at a rate of 86% of the epenthetic vowels. The other vowel was central (a schwa) occurring at a less frequent rate 14%. This pattern was observed in the delayed repetition task only.

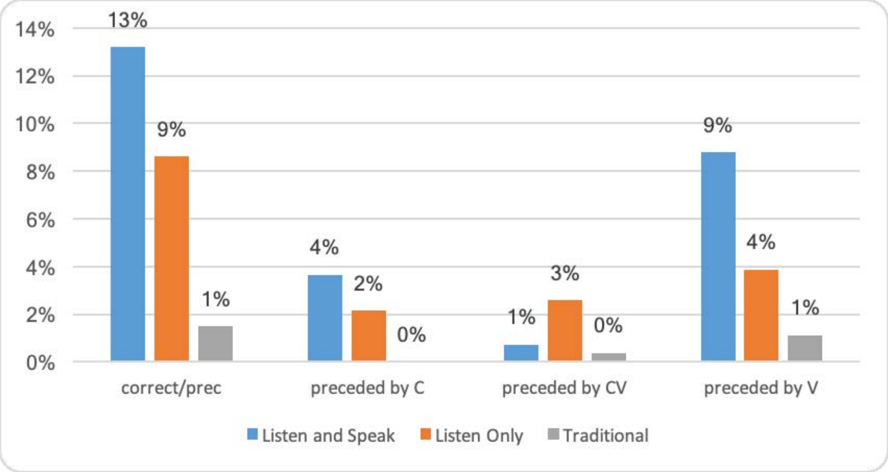


Figure 9.41 Percentages of processes for /ɹ-/ according to instruction type during the post-test

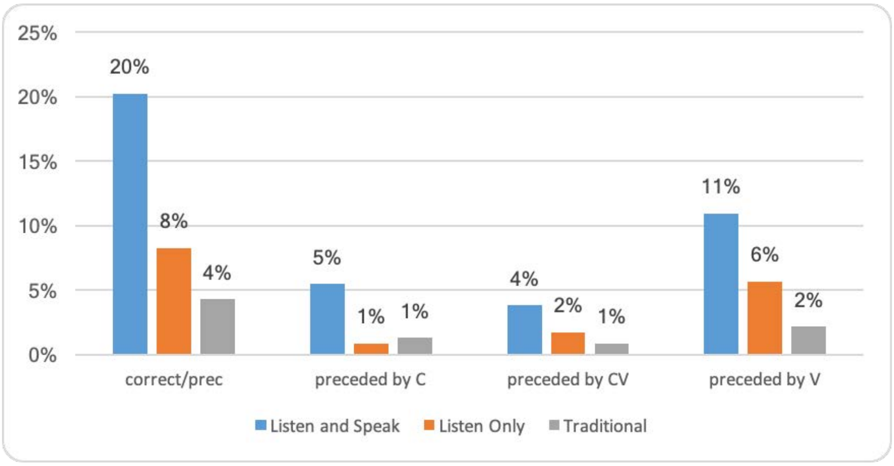


Figure 9.42 Percentages of processes for /ɹ-/ according to instruction type during the delayed test

It was mentioned earlier that Listen and Speak illustrated the highest occurrence rate, whereas Traditional exhibited the lowest. It is worth-noting however, that most speakers, especially those in Listen and Speak produced several tokens illustrating this pattern. Thus, in terms of speakers, in the post-test, 10/20 from Listen and Speak, 5/18 from Listen Only, and 2/20 from Traditional exhibited this pattern. In the delayed test, 7/13 from Listen and Speak, 6/17 from Listen Only, and 4/14 from Traditional. These two observations indicate that the process is more established within speakers from Listen and Speak, given that there were more speakers demonstrating the process and that six out of the ten who did, produced multiple tokens. At the other end of the continuum lies speakers from Traditional, where there were less speakers, whom only produced single tokens each. The process seemed to have increased during the delayed test in all of the groups.

Epenthetic Consonant

A closer examination of epenthetic consonants reveals several points. First, this segment appeared only in the delayed repetition elicitation task. Moreover, the most frequent segment was the bilabial stop (26 out 30 tokens across the three groups). The other two epenthetic consonants were /ʔ/ and /v/. /ʔ/ had three occurrences in the delayed test only within the experimental groups only; two tokens by one male speaker from Listen and Speak and one token also by one male speaker but from Listen Only. This was illustrated in the test items ‘rose’, ‘red’ and ‘rob’. /v/, on the other hand, appeared once in a token from Listen and Speak also by a male participant in the test item ‘round’.

Epenthetic Consonant + Vowel

The majority of the epenthetic CV sequences consist of /ʔ/ followed by a back vowel (12 tokens), for example ‘read’ [ʔɔɹiːd] or followed by a central vowel (one token), for example ‘wrong’ [ʔəɹɔŋ]. This is usually referred to as a dummy syllable. Alternatively, the consonant can be a bilabial plosive (mostly devoiced) followed by a back vowel (4 tokens), for example ‘rip’ [bɔɹɪp] or followed by a central vowel (three tokens), for example ‘rose’ [b̥ɨɹɔʊəz]. There are a few cases, however, when this was not the case. Consider the following examples in 9.34.

	item	realisation
9.34	round	[ɹɑɹɔŋd̥ʰ]
	root	[wəɹʊ̜tʰ]

Substitution only

/b/ appeared in both tests (14 in the post-test and 12 in the delayed test). In the post-test, it was demonstrated by the experimental groups only; nine in Listen and Speak and five in Listen Only even though Listen and Speak had a lower number of participants compared to Listen Only (13 and 17 speakers respectively). The consonant was added by speakers of each gender and many times, speakers would have multiple tokens exhibited /b/ preceding the word-initial rhotic. In the delayed test, Traditional also demonstrated the epenthetic consonant in three of the tokens, two by one speaker and the third by a different speaker. However, this time, for Listen Only, the process was demonstrated in one token only by a female speaker. Listen and Speak, despite having the lowest number of participants in the delayed test, had the highest occurrence of an epenthetic /b/ preceding the initial rhotic. Six tokens were produced by three speakers and two tokens were produced by a single female speaker. Overall, the test items that demonstrated an epenthetic /b/ the most were ‘roll’ (eleven), ‘round’ (eight), and ‘wrap’ (four).

As illustrated above in 9.39 and 9.40, substitution was the most common process in both tests. Within this process, substitution with /w/ was the most common especially for Traditional as this group had the highest rate of substitution with /w/ in both tests (two fifths in the post-test and one fifth in the delayed test). Followed by that was Listen Only, whereas Listen and Speak had the lowest rate of /w/ substitutions (one tenth of the tokens were substituted with /w/ in each test).

The second most common substitution was /ɾ/, which is the participant's native language counterpart.

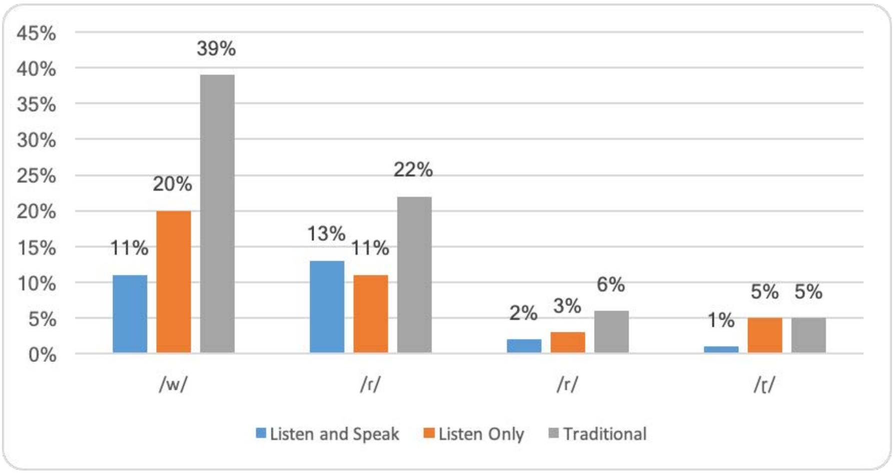


Figure 9.43 Percentages of substitution in /ɹ-/ according to instruction type during the post-test

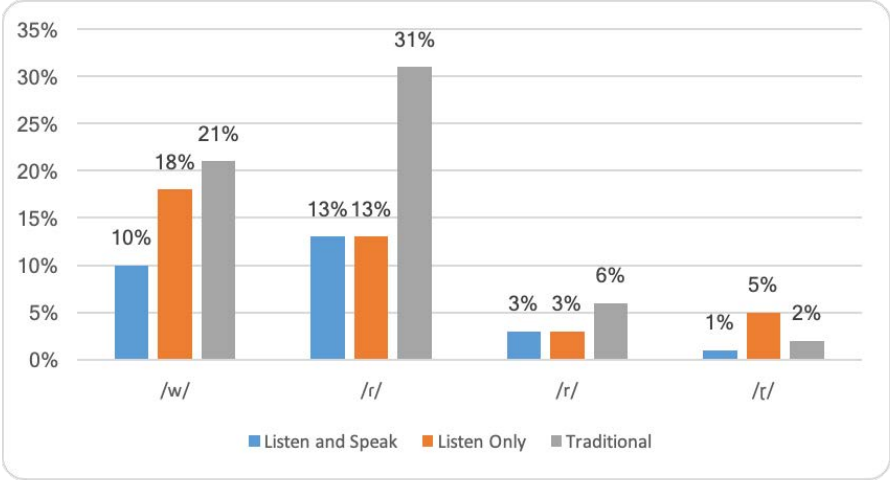


Figure 9.44 Percentages of substitution in /ɹ-/ according to instruction type during the delayed test

Again Traditional demonstrated the highest rate of /ɾ/ substitution. For Traditional in the post-test, it was lower than /w/ substitution (just over half as frequent) but in the delayed test it was higher than /w/ substitution (by 50%). The experimental groups illustrated a somewhat matching rate albeit Listen Only was slightly lower in the post-test. However, it is noted that Listen Only exhibited lower rates in the delayed test in comparison with the post-test, whereas Listen and Speak showed higher rates in the delayed test in comparison with the post-test. Other less frequent substitutions are shown in table 9.22.

Table 9.22 Less frequent substitutions of /ɹ-/

substitution	item	realisation
/r/	red	[rɛd]
/ɽ/	rose	[ɽɔːz]
/ɰ/	wrong	[ɰɔŋg]
/l/	rice	[laɪs]
/ʋ/	wreck	[ʋɛkʰ]
/dˀ/	round	[dˀăʊ̰d]
/z̪/	root	/[z̪uːtʰ]
/ɣ/	red	[ɣɛd]
/h/	roll	[hɔˑl]
/ɽ/	read	[ɽiːdʰ]
/b/	round	[baʊnd̥ʰ]
/ʀ/	rain	[ʀeːn]

For the list of substitution frequencies and how they varied between tests, consider table 9.23 below.

Table 9.23 Sunbstitution frequencies in /ɹ-/ by time of test

		post-test		
	Listen and Speak	Listen Only	Traditional	total
All	85	94	230	409
/w/	30	47	103	180
/ɾ/	35	26	59	120
/r/	5	6	17	28
/ɽ/	3	11	14	28
/l/	2	–	16	18
/ɰ/	5	–	7	12
/ʋ/	2	3	5	10
		delayed test		
	Listen and Speak	Listen Only	Traditional	total
All	52	97	156	305
/w/	18	42	49	109
/ɾ/	24	29	72	125
/r/	5	8	14	27
/ɽ/	2	11	4	17
/l/	–	–	–	–
/ɰ/	–	3	8	11
/ʋ/	1	–	3	4

The figures in the table are organised descending from the most frequent to the least frequent in the post-test. They show that the order of frequency fluctuates between tests and between instruction groups. Moreover, not all substitutions were sustained until the delayed test. Lateralisation, for example, disappears in the delayed test. In the post-test, a subset of the tokens were the result of consonant harmony, for example [łɔł] for ‘roll’. Males produced twice as many lateralised tokens as the females did. All of them were illustrated exclusively in the delayed repetition task. Similarly, a new substitution appeared in the delayed test, pharyngealised /d/ that was not illustrated in the post-test albeit scarce, appearing in Listen Only exclusively and produced by the same speaker for the test items ‘wrap’ and ‘round’.

Substitution preceded by a V, C, or CV added

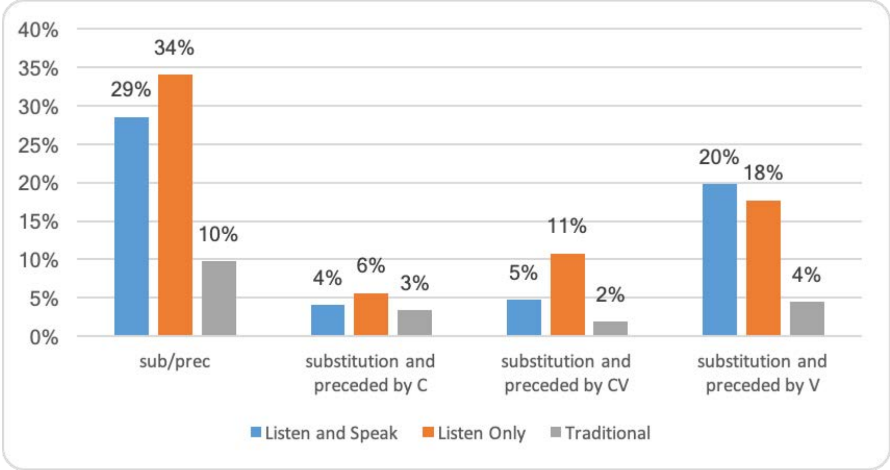


Figure 9.45 Percentages of tokens as per type of segment preceding substitution for /ɹ-/ according to instruction type during the post-test

Figures 9.39 and 9.40 showed that substitution accompanied by an epenthetic segment was the second most frequent process (after substitution only) in word-initial /ɹ-/. Within this process, Listen Only exhibited the highest rate of such a process in both tests, whereas Traditional showed the lowest rate also in both tests. An examination of the sub-patterns that is substitution preceded by a vowel, a consonant or CV sequence, reveals that the epenthetic vowel was the most frequent in both tests. Figures 9.45 and 9.46 demonstrate that within substitution accompanied by an epenthetic consonant, in either test Listen Only had the highest frequency and that Listen and Speak and Traditional had a somewhat matching rate of occurrence. Listen Only also exhibited the highest percentage of substitution accompanied by an epenthetic CV sequence and Traditional had the lowest in the post-test. However, in the delayed test it was Listen and Speak that showed the highest rate and Traditional maintained the lowest rate of occurrence. Finally, substitution accompanied by a preceding vowel was the highest amongst Listen and Speak in the post-test and Listen Only in the delayed test. The group with the lowest rate in the post-test was Traditional, whereas in the delayed test it was Listen and Speak.

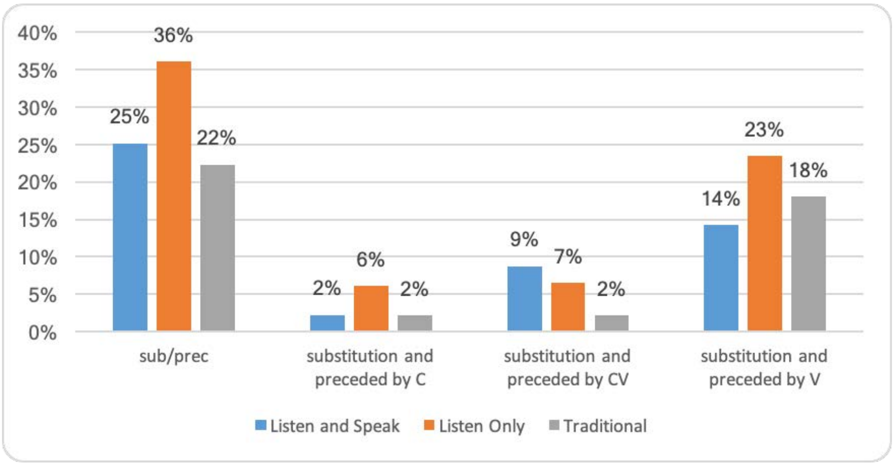


Figure 9.46 Percentages of tokens as per type of segment preceding substitution for /ɹ-/ according to instruction type during the delayed test

Substitution and vowel epenthesis

The most frequent preceding vowel in this pattern was a back vowel in both tests with generally 94% and 96% in the post-test and the delayed test respectively. When there was an epenthetic back vowel, the target /ɹ/ was substituted with various sounds. The most frequent of these was [w] especially for Listen Only. The second most frequent substitution was [ɽ], especially for Listen and Speak. The third most frequent substitution was [ɾ] especially for Listen Only. Other less frequent substitutions included [r] and [j] in Listen Only only, whereas [ɰ], [ʋ], and [ɻ] were seen in Listen and Speak only. Other less frequent epenthetic vowels accompanying substitutions were a central vowel followed by [r] (in Listen and Speak and Traditional) or [ɾ] (in Listen and Speak and Listen Only) or a front-high vowel followed by [j] (in Traditional only).

Table 9.24 Percentages of substitution accompanied by an epenthetic vowel across instruction types in the post-test

preceded by		Listen and Speak	Listen Only	Traditional
back vowel		94%	98%	83%
	[w]	39%	46%	42%
	[ɾ]	7%	34%	17%
	[ɽ]	39%	10%	25%
	[r]	–	5%	–
	[j]	–	2%	–
	[ɰ]	6%	–	–
	[ɻ]	2%	–	–
	[ʋ]	2%	–	–
central vowel		6%	2%	8%
	[r]	4%	–	8%
	[ɾ]	2%	2%	–
front high vowel		–	–	8%
	[j]	–	–	8%

In the delayed test, the back vowel was also the most frequent vowel preceding the substitution process. The highest rate of substitution this vowel was [w]. It was the highest in Listen Only. The next most frequent was [ɽ]. It was highest in Listen and Speak followed closely by Traditional. The next most frequent was [ɾ], followed by [r], especially high in Traditional. [ɰ] was the next most frequent, but it only appeared in the experimental groups, whereas [ʀ] appeared exclusively in Traditional. The other less frequent epenthetic vowel, which appeared in the delayed test was a central one. it was followed by [ɾ] or [n] in Traditional only and [r] in Listen Only and Traditional.

Table 9.25 Percentages of substitution accompanied by an epenthetic vowel across instruction types in the delayed test

preceded by		Listen and Speak	Listen Only	Traditional
back vowel		100%	98%	90%
	/w/	42%	61%	38%
	/ɽ/	31%	11%	24%
	/ɾ/	15%	11%	14%
	/r/	4%	9%	12%
	/ɰ/	8%	6%	–
	/ʀ/	–	–	2%
central vowel		–	2%	10%
	/ɾ/	–	–	5%
	/r/	–	2%	2%
	/n/	–	–	2%

Substitution and consonant epenthesis

The most frequent epenthetic consonant in this pattern was [b] in both tests with generally 79% and 65% in the post-test and the delayed test respectively. In the post-test, when there was an epenthetic [b], /ɹ/ was substituted with various sounds. In both tests, the most frequent of these was [ɾ], which was highest in Listen and Speak, followed by Listen Only and finally Traditional had the lowest frequency. The second most common substitution accompanied by an epenthetic [b] was [ɽ]. During the post-test, it was highest in Traditional. The experimental groups did not fall far behind. During the delayed test, however, it appeared in the experimental tests only, with 25% and 14% in Listen and Speak and Listen Only respectively. [ɰ] was overall as frequent as the latter substitution in the post-test. However, it was the most demonstrated by Traditional (22%). It was evident in Listen and Speak (9%) as well but did not appear in Listen Only. In the delayed test, it only appeared in Listen and Speak (25%). The third most frequent substitution in the post-test was [w], which appeared in Traditional (11%) and Listen Only (8%) only. In the delayed test, it was demonstrated by Traditional only (20%).

Other less frequent substitutions accompanying the epenthetic [b] were [l], [j], and [z], which appeared in Traditional only and exclusively in the post-test. Also, [ʀ] appeared in Listen Only only in the post-test. In the delayed test, it was demonstrated by Traditional only. [r] and [ɻ] appeared exclusively in Listen and Speak during the post-test.

The second most frequent epenthetic consonant was [ʔ] with 18% and 22% in the post-test and the delayed test respectively. It is worth-mentioning that Traditional demonstrated no such epenthetic consonant and that this pattern almost exclusively appears in Listen Only. Listen and Speak only shows this pattern in the post-test. When it did occur it was accompanied by substitution with [ɾ] (9%). As for Listen Only, it was accompanied by three types of substitution, namely [ɽ] with 15% in the post-test and 36% in the delayed test. The other two substitutions occurred only in the post-test; [ɾ] (15%) and [ɻ] (8%).

Other less frequent epenthetic consonants were [v] followed by substitution with [w] appearing in Listen and Speak only during the post-test. Also, [t] followed by substitution with [w] appearing in Traditional and Listen Only only during the delayed test. Finally [z] followed by substitution with [ɾ] appearing in Traditional only during the delayed test.

Table 9.26 Percentages of substitution accompanied by an epenthetic consonant across instruction types in the post-test

preceded by		Listen and Speak	Listen Only	Traditional
[b]		82%	62%	100%
	[j]	–	–	11%
	[l]	–	–	11%
	[ɰ]	9%	–	22%
	[r]	9%	–	–
	[ʀ]	–	8%	–
	[ɽ]	9%	8%	11%
	[ɻ]	9%	–	–
	[ɾ]	45%	38%	22%
	[w]	–	8%	11%
	[z]	–	–	11%
[v]		9%	–	–
	[w]	9%	–	–
[ʔ]		9%	38%	–
	[ɽ]	–	15%	–
	[ɻ]	–	8%	–
	[ɾ]	9%	15%	–

Table 9.27 Percentages of substitution accompanied by an epenthetic consonant across instruction types in the delayed test

preceded by	Listen and Speak	Listen Only	Traditional
[b]	100%	57%	60%
[ɰ̢]	25%	–	–
[ʀ]	–	–	20%
[ɽ]	25%	14%	–
[ɾ]	50%	43%	20%
[w]	–	–	20%
[t]	–	7%	20%
[w]	–	7%	20%
[z]	–	–	20%
[ɾ]	–	–	20%
[ʔ]	–	36%	–
[ɽ]	–	36%	–

Substitution and consonant + vowel epenthesis

The most frequent preceding CV sequence in this pattern was [ʔ] followed by a back vowel in both tests. In the post-test, 25 tokens exhibited this pattern, of which 16 were from Listen Only, seven from Listen and Speak, and only two tokens from Traditional. Substitutions accompanying this epenthetic syllable included [w] being the most frequent overall. This was demonstrated the most in Listen Only. It was also the only substitution following this CV sequence in Traditional. Other less frequent substitutions accompanying this CV combination included [ɾ], [ɰ̢], [ɽ], and [r]. Other CV sequences that preceded the target /ɹ/ included [w], [b] or [l] and a back vowel, [ʔ] or [b] and a central vowel, [wa] and [tɪ]. The group that showed the most variation was Listen Only followed by Listen and Speak. Traditional on the other hand illustrated only a few combinations; 'rose' [loɣɵz̥] and [weːɾɨz], 'wrist' [wɑɾɪs], 'wreck' [ʔɔwɛk^{h}], 'wrist' [ʔɔwɪst].

9.6.2 *Rhotic approximant word-finally*

Test items in this category include the words 'beard', 'chair', 'fear', 'hair', 'hear', 'jar', and 'march'. Productions of the Traditional Teacher are shown in table 9.28.

Table 9.28 Traditional Teacher's realisations of /-ɹ/ tokens

item	task	process	target-likeness	realisation
jar	read aloud	correct	non-targetlike	[ʤɑːɾ]
	picture-naming	correct	non-targetlike	[ʤ̥ɑːɾ]
beard	read aloud	postvocalic /r/	non-targetlike	[beːɾd]
	picture-naming	postvocalic /r/	non-targetlike	[beːɾd]
chair	picture-naming	postvocalic /r/	non-targetlike	[ʧeːɾ]
	read aloud	postvocalic /r/	non-targetlike	[ʧeəɾ]
fear	read aloud	postvocalic /r/	non-targetlike	[feːɾ]
	picture-naming	postvocalic /r/	non-targetlike	[feːɾ]
hair	read aloud	postvocalic /r/	non-targetlike	[heːɾ]
	picture-naming	postvocalic /r/	non-targetlike	[heːɾ]
hear	picture-naming	postvocalic /r/	non-targetlike	[hiːəɾ]
	read aloud	postvocalic /ɹ/	non-targetlike	[hiˑəɾ]
march	read aloud	postvocalic /r/	non-targetlike	[mɑːɾʃ]
	picture-naming	postvocalic /r/	non-targetlike	[mɑːɾʃ]

Overall, the rate of correct realisations of post-vocalic /ɹ/ was high (78% in post-test and 83% in delayed test). The increase in the delayed test suggests an improvement over time. Listen Only, however, showed a slight decrease. Figure 9.47 shows that Listen Only had the highest rate of correct realisations, followed closely by Listen and Speak in both tests. For Traditional, only half the tokens containing post-vocalic /ɹ/ were correct during the post-test. During the delayed test, figures increased; two thirds of the tokens containing post-vocalic /ɹ/ were correct.

Data regarding post-vocalic /ɹ/ exhibit two non-targetlike patterns namely the realisation of post-vocalic /ɹ/ and gliding, the former being more common (overall 21% and 16% in the post-test and delayed test respectively) and the latter being less frequent (0.5% and 1.2% in the post-test and delayed test respectively) and exclusively in the test word 'march' and mostly in Traditional. March is the one of two cases, where post-vocalic /ɹ/ is followed by a consonant. Consider the figures in 9.47. The group that had the most occurrence of post-vocalic /ɹ/ was Traditional in both post-test (46%) and delayed test (33%). Nevertheless, it decreases in the delayed test. Its presence was expected since their Traditional Teacher exhibited this feature in her realisations. Amongst the two experimental groups, percentages were low. Listen and Speak had twice as much as Listen Only during the post-test. This gap tightened in the delayed test and percentages were still low. It seems that such patterns are observed exclusively

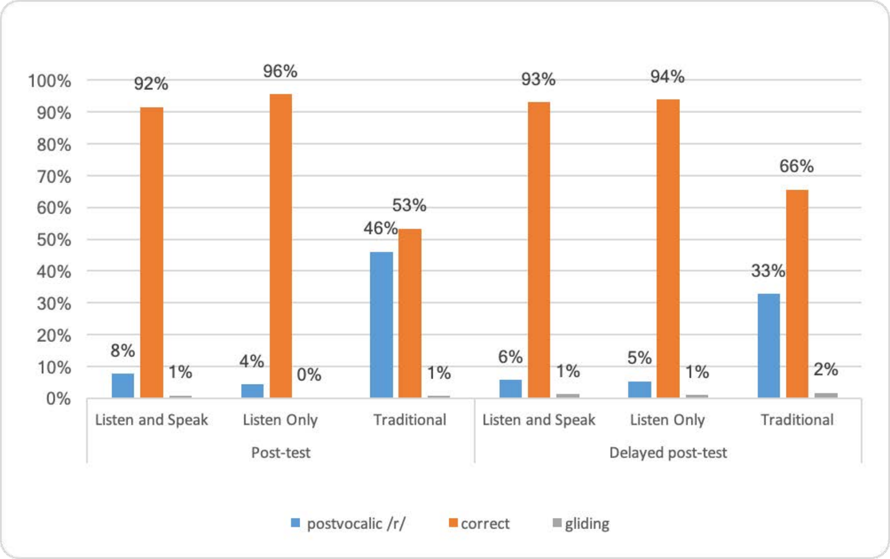


Figure 9.47 Percentages of processes for word-final /ɹ/ according to instruction type in both tests

amongst literate participants within the experimental groups.

For example, ‘fear’, [fɪəɾ], ‘march’, [mɑːɾʧ], ‘beard’, [biːɾd], and ‘jar’, [ʤ̥ɑːɾ].

The quality of post-vocalic /ɹ/ was mostly [ɾ]. However, there were instances of [ɽ], [ɹ], [ʁ], and [ɻ].

Furthermore, there were a few instances from Traditional, where an epenthetic vowel appeared between a post-vocalic /ɹ/ and the final consonant in ‘beard’, for example [b̥ɪəɾɨd̥̚]. Gliding, as mentioned above, was scarce. It was demonstrated in the test item ‘march’ only. It appeared in all of the instruction groups; two tokens from Listen and Speak, one token from Listen Only, and three tokens from Traditional. All tokens were produced by literate or semi-literate speakers and exclusively during the delayed repetition task. Examples include [mɑwt͡s], [mawɨʃ], [mɑːjʧ], [mɑːwʃˀ], [mɑwʧ], and [mau]. It is noticed that in one case that /ɹ/ glided into a palatal and in another case the coda was deleted. Other processes accompanied gliding that affected the coda, affricate.

9.7 Summary and discussion

This chapter presented data relating to phonological processes illustrated by the three training conditions for each segment of interest. The goal was to explore these processes with two aims in mind. One aim was find out whether the processes vary by training. The other aim was to find out whether these processes reflect Arabic L1 transfer, general English child language development or universal language development. The segments of interest were examined by phonological context and for CC coda clusters and diphthongs, by type. This chapter served as a supplementary analysis for the target-likeness results in Chapter 6.

Table 9.29 below summarises the percent correct in each problematic sound class categorised by phonological context/type.

Table 9.29 Summary of correct percent per sound and phonological environment

		Post-test			Delayed post-test		
		Listen and Speak	Listen Only	Traditional	Listen and Speak	Listen Only	Traditional
Affricates	/ʧ-/	96%	88%	92%	97%	86%	89%
	/ʤ-/	95%	90%	82%	91%	87%	85%
	/-ʧ/	75%	50%	57%	1%	44%	60%
Coda clusters	/-ld/	43%	5%	40%	23%	6%	37%
	/-lt/	39%	35%	50%	45%	35%	52%
	/-lk/	36%	37%	64%	46%	22%	58%
	/-ft/	91%	89%	70%	92%	94%	96%
	/-st/	88%	83%	70%	87%	82%	84%
	/-nt/	41%	27%	56%	38%	40%	52%
	/-nd/	85%	54%	65%	88%	62%	76%
	/-nz/	62%	59%	58%	92%	82%	56%
	/-mp/	27%	44%	67%	58%	29%	25%
Dental fricatives	/θ-/	47%	40%	40%	36%	31%	30%
	/-θ/	39%	42%	28%	43%	35%	35%
Diphthongs	/aɪ/	92%	86%	66%	87%	81%	65%
	/ɪə/	63%	76%	40%	79%	87%	42%
	/ɔɪ/	73%	64%	34%	70%	62%	43%
	/aʊ/	71%	63%	32%	69%	44%	38%
	/eɪ/	44%	37%	29%	50%	40%	25%
	/əʊ/	33%	34%	26%	51%	38%	17%
Rhotics	/ɹ-/	27%	17%	3%	26%	13%	7%
	/-ɹ/	92%	96%	53%	93%	94%	66%

highest	lowest	*similar* ± 3

The results for percent correct or target-likeness in this table are different from the results found in Chapter 6 for target-likeness and match ratings.[60] This is because a) these are the results of a micro-analysis of individual sounds or sound structures, in different phonological environments, and b) the data here were not subjected to statistical analysis. We can see here that – contrary to the results in Chapter 6 and table 6.7 – that:

- In the post-test:

    1. Those following the Traditional Teaching condition outperformed those following the CAPT training in the coda clusters exclusively, and in the types /-lt/, /-lk/, /-nt/, and most considerably in /-mp/.

---

[60] See table 6.7 for a summary.

    2. The Traditional Teaching group and the Listen and Speak group had similar levels of sound percent correct in the coda cluster type /-ld/, both of which groups outperformed the Listen Only group.
    3. The Traditional Teaching group and the Listen Only group had similar levels of sound percent correct in the cluster type /-nz/ and word-initial dental fricatives, albeit both groups were outperformed by the Listen and Speak group.
    4. The Listen and Speak group had considerably the lowest percent correct in the coda cluster /-mp/ in the post-test and only marginally the lowest percent correct in /-ft/.

- In the delayed post-test:
    1. Those following the Traditional Teaching condition outperformed those following the CAPT training in the coda cluster types /-ld/,/-lt/, /-lk/, /-nt/, and to a lesser extent /-ft/.
    2. The Traditional Teaching group and the Listen Only group had similar levels of sound percent correct in word-initial voiced affricates and dental fricatives word-initially and finally, and where they were both outperformed by the Listen and Speak group. However, the findings for dental fricatives here are consistent with those in the target-likeness analysis given the consistent performance across both environments.

- Within the various diphthongs, the Traditional Teaching had consistently the lowest percent correct, whereas this varied between the CAPT conditions depending on the individual diphthong. This reflects the input received from the Traditional teacher.

- For the lateral approximant + C coda clusters, the Traditional Teaching group showed the highest overall percent correct due to their production of a clear /l/ reflected in their teacher's realisations. The CAPT training participants however, resorted to /l/ vocalisation which is not only typical in SSBE where the tongue-alveolar ridge contact is weakened (Scobbie and Wrench, 2003; Trudgill, 1984; Tollfree, 1999; Hardcastle and Barry, 1985) but also a prominent feature in child speech (Cruttenden, 2008). This process lowered the CAPT training conditions' learners' percent correct for this cluster type.

- Occasionally – in 4 different sounds/sound combinations: /-nz θ- -θ ʤ-/ – the Traditional Teaching group had a similar percent correct as the Listen Only condition. In all such cases, the Listen and Speak outperformed each.
- In one case, /-ld/ they have a similar percent correct to the Listen and Speak learners where both outperform the Listen Only group.
- In many cases, Traditional Teaching learners did better than those following the Listen Only condition but there was a wide range of variation as this pattern fluctuated between these two groups.

The results presented above also showed processes which learners from the three training conditions exhibited in their interlanguage. A synthesis of these findings alongside inferential results of target-likeness and match ratings will be attended to in Chapter 10.

Chapter 10: Discussion

The aim of the present thesis was to test conflicting hypotheses on the role of delayed (Listen Only) vs. instant (Listen and Speak) production on L2 pronunciation, namely target-likeness and phonetic learning. The aim was also to examine the impact of native vs. accented input on the pronunciation of Arabic child learners of English. It was hypothesised in Chapter 5 that the Listen and Speak condition will result in the most target-like pronunciations/least L1 like productions followed by the Listen Only, leaving the Traditional condition with the relatively least target-like pronunciations/ most L1-like productions. It was also hypothesised that the experimental training conditions will resemble *relatively more* of the English child phonological developmental stages compared to the Traditional teaching condition who received accented input and also *relatively less* L1 interference.

In terms of the impact of comprehension (perception) on L2 phonological learning, the relationship between perception and production is based on the idea that better perception is a prerequisite for better production (Flege, 1995; Wode, 1996) and that there is a correlation between production and perception in adult L2 acquisition (Flege and Schmidt, 1995). It is believed that perception precedes production. Baker et al. (2008) maintain that at the beginning stages of L2 learning it is typical to fail in producing target-like productions due to the incapability of learners to accurately perceive target sounds or structures. Notwithstanding, adequate perception does not guarantee accurate production (Colantoni and Steele, 2008). Several L2 studies have demonstrated that high levels of adequate L2 perception does not necessarily lead to higher levels of production accuracy even in advanced learners (Flege, MacKay, et al., 1999). However, perceptual training studies indicate that training L2 learners to perceive sound contrasts may lead to improved perception and production (Bradlow, Akahane-Yamada, et al., 1999; Lively, Pisoni, Yamada, Tohkura, and Yamada, 1994).

The findings of this study, in a way, support the findings of Herd et al. (2013) and Sakai (2016), who show that combined perception and production training lead to greater gains in the area of production when compared with perception only training. However, this does not mean that delayed production does not have its advantages. It is highly probable that in a perception test, the Listen Only condition will show greater improvements in perception compared with the Listen and Speak condition and the Traditional teaching.

Theories of L2 speech learning vary in their predictions of degree of difficulty. Colantoni and Steele (2008) argue that not all sounds share equal degree of difficulty in acquisition. Additionally, not all L2 difficulty can be based on perceptual difficulties alone. Sounds that are considered *different* from the L1 can have varying degrees of difficulty. Some sounds are articulatorily more demanding than others in terms of voicing, voicing by place, and manner (Diehl and Lindblom, 2004). Difficulty acquiring a non-native sound may also vary according to phonological context, that is a given segment may be harder to acquire in coda position compared to onset position (Colantoni and Steele, 2008). In their hybrid model, Colantoni and Steele (2008) propose that the role of production in L2 learning is enabling learners to reflect upon their outputs by comparing them to their perceived inputs. If differences are discerned, learners will continue to modify their speech until no further differences are noticed. At this point their phonetic learning ceases and their categories become fossilised. Thus the goal of the current study is to determine whether – and to what extent – short-term (15 hours) listening only practice helps develop L2 pronunciation. To this end, comparisons of target-likeness, match rating, lexical learning, and acoustic measures of voicing (VOT and vowel-onset f0) and place of articulation (spectral tilt) were implemented for 7-year-old Libyan Arabic children in three training conditions. Two experimental conditions, one involving aural only practice, one involving listening and speaking practice and the third condition involving traditional teaching all of which conditions were considered using the same materials. The experimental conditions used a CAPT programme in the Digital Literacy Instructor software (Overal, 2014). The Traditional condition were trained by a foreign-accented Libyan teacher of English.

The results varied by analysis type. For the ***Target-likeness*** analysis, the data items were compared to IPA transcription holistically. This analysis showed an *overall* statistical advantage for native input (CAPT conditions) over non-native input (Traditional Teaching) in the post-test. Exceptions here are that a) there was no statistical difference in the class of dental fricatives between any of the training conditions, and b) that the difference in performance between the Listen Only condition and the Traditional Teaching condition in the class of clusters did not reach statistical significance, either. In the delayed post-test, the influence of input type was neutralised a) in coda clusters and dental fricatives only between the Listen and Speak and the Traditional Teaching condition, and b) almost completely between the Listen Only and the Traditional Teaching condition, except in the class of diphthongs. In terms of listening and speaking practice vs. delayed production, there was an advantage for listening and speaking practice in affricates only in the post-test and clusters only in the delayed post-test (see summary table 6.7). For the ***Match*** analysis, only the problematic sound was considered and each was compared to the respective input. This analysis showed that within the CAPT conditions, there was an advantage for listening and speaking practice over delayed production in affricates and rhotic approximants in both tests, but not for any other problematic class. In the remaining problematic sounds, differences did not reach statistical significance. By comparing the performance between the Listen and Speak and the Traditional Teaching condition, who only varied by input type as both practised oral production alongside listening practice, it is found that native speaker input has an advantage over non-native speaker input in all problematic sound(s)/structures, except in plosives in the post-test as the difference did not reach statistical significance. In the delayed post-test, the advantage of native speaker input is only statistically evident in affricates, dental fricatives, and diphthongs. By comparing the performance between the Listen Only and the Traditional Teaching condition, who varied by not only input type but also mode of practice, it is found that in the post-test, the former had an advantage over the latter in the class of dental fricatives, diphthongs, and plosives only. In the delayed post-test, this was restricted to dental fricatives and diphthongs, but not plosives.

These results indicate that:

- for affricates and rhotic approximants the effect of native input and listening/speaking practice combined are the determining factors in pronunciation accuracy, and that this has a lasting effect evident in the consistency of this finding ten weeks after the training finished,

- for dental fricatives and diphthongs, native input is *the only* determining factor for pronunciation success, and that this also has a lasting effect evident in the consistency of this finding ten weeks after the training finished,
- for clusters, native input has a positive impact on pronunciation accuracy but only when combined with listening/speaking practice, but only immediately after the training. Ten weeks later, this effect is lost,
- for plosives, native input once again has a positive impact on pronunciation accuracy but only when combined with *delayed* oral practice, but again only immediately after the training. This effect is lost without further practice.

The results of Match ratings varied from those of the Target-likeness ratings because the rating of the word as a whole was determined by how many problematic sounds there were in a single test item as well as to what they were compared, that is the target language vs. respective input. The results from the phonological processes chapter further varied from either of the above as they dealt with individual sound(s)/structure type and/or phonological context. This analysis showed that within one problematic sound class, there were variations between the CAPT conditions depending on the position of the problematic sound in a word. For example in dental fricatives and rhotic approximants, the Listen and Speak group did better word-initially and the Listen Only group did better word-finally. There were also variations between the CAPT conditions by sound type. For example in diphthongs, the Listen Only group did better in /ɪə/ in both tests, whereas the Listen and Speak group did better in the remaining vowels in both tests. An exception to this was /əʊ/ for which the two CAPT groups showed a similar performance in the post-test. There was further variation in the class of coda clusters depending on cluster type: a) the CAPT groups had similar percent correct for /-lk/, b) the Listen and Speak group had the lowest percent correct for /-mp/ not only compared to the Listen Only group but also the Traditional Teaching group, c) this changed in the delayed post-test where the learners of this condition improved to the point of surpassing their peers from each of the other training conditions, d) the Listen and Speak group exceeded the Listen Only group in the percent correct for the cluster types fricative+obstruent and alveolar nasal+obstruent, e) the Listen Only group showed the poorest performance in the cluster types /-ld/, /-lt/, /-nt/, and /-nd/ in the post-test indicating that for lateral/nasal+ alveolar plosive, production practice is crucial (see summary table 9.29). This analysis also showed that in the post-test, the Traditional Training condition was disadvantaged by the non-native input in the voiced affricates more so than their voiceless counterpart, the fricative+obstruent clusters, word-final dental fricatives, in all diphthongs, and rhotic approximants in either position. For diphthongs and rhotic approximants, the non-native input has a lasting effect evident in their lowest percent correct in the delayed post-test as well. The Traditional Teaching group's performance however, was parallel to that of the Listen and Speak condition within the /-ld/ cluster in the post-test, which they have managed to exceed in the delayed post-test having the highest percent correct for this cluster type. This was thought to be as a result of producing a clear [l] when the CAPT conditions showed a process of /l/ vocalisation typical of SSBE (Cruttenden, 2008; Hardcastle and Barry, 1985; Scobbie and Wrench, 2003; Trudgill, 1984; Tollfree, 1999). The Traditional Teaching group's performance was parallel to that of the Listen Only condition in the /-nz/ cluster and word-initial dental fricatives in the post-test, indicating no advantage for native speaker input without listening and speaking practice for these sounds types. Whilst the Traditional Teaching group managed to maintain a level of performance similar to the Listen Only group for the word-initial dental fricatives in the delayed test, their performance fell behind in the /-nz/ in the delayed post-test. The Listen Only condition demonstrated the worst performance in the voiceless affricates, lateral+alveolar plosive and nasal+alveolar plosive clusters. This is indicative of the negative impact of delayed oral practice on those specific sounds/structures and/or context in the post-test. In the delayed post-test, their percent correct in voiced affricates worsened and was similar to that of the Traditional Teaching condition, they had the lowest percent correct in /-lk/ clusters after being no different from their CAPT peer in the post-test, and they were outperformed by the Traditional Teaching group in the /-st/ clusters where they had the lowest percent correct suggesting once more a disadvantage of delayed oral practice which seems to override the effect of native input.

The results of the phonetic analyses showed that the three training conditions created intermediate L2 categories for the voiceless and voiced plosives which resemble both TL and native L1 values. This does not mean that learning did not take place as the L2 categories were mostly indistinguishable from the TL categories. Moreover, they have all successfully managed to statistically distinguish their voiced from voiceless categories including bilabial plosives. This is surprising given /p/ is absent in Libyan Arabic and considering native English speakers find it difficult to identify it when produced by Arabic speakers (Flege and Port, 1981).

Affricates had the highest probability of target-likeness rating compared to the other sound classes. This is not surprising since studies of English first language acquisition (e.g. Dodd, 2003; 2013 and Smit et al., 1990) and simultaneous Arabic-English bilingual children (e.g. Al-Amer, 2018) indicate that English affricates are the first to be acquired in the set of sound classes examined in the present study.[61] Affricates are not the first in this group for English only, but for many other languages (Mcleod and Crowe, 2018). It seems that mastering the articulatory settings for this sound class is less demanding compared to the other sounds examined in the current study. Moreover, in the examination of affricates by voicing and phonological context, results revealed that learners generally had a higher accuracy rate in the voiceless affricates compared to their voiced counterpart and in onsets compared to codas, albeit stronger effects of position than voice. Following Colantoni and Steele (2008), to explain the asymmetries in the relative accuracy of affricates by voicing, I argue for an explanation based on cross-linguistic typology for affricates. Voiced affricates, it is argued, are more marked in relation to their voiceless counterpart. An explanation of such onset-coda asymmetries can be derived from Lindblom's (1990; 1996) Hyper and Hypo theory (H&H). For clusters, several studies on Germanic languages such as English, German and Dutch revealed the opposite pattern, whereby the acquisition of coda clusters precedes onset clusters (Kirk and Demuth, 2005; Levelt et al., 2000; Lleó and Prinz, 1996; Templin, 1957) with the exception of McLeod et al. (2001b), who found no such differences. In second language acquisition, researchers like Wenk (1979; 1983) demonstrate a tendency amongst L2 learners to master novel articulations word-initially (and intervocalically) prior to those word-finally. Clusters are arguably

---

[61] An exception to this was the class of plosives. This point will be revisited later in the discussion.

mastered later (Wenk, 1979; Wenk, 1983).

In terms of the phonological processes observed in the L2 data for this sound class, the most frequent process was de-affrication, vis-à-vis the replacement with either [ʃ] or [s]. This process is widely attested in English child development (e.g. Dodd, Holm, Hua, et al., 2003; McIntosh and Dodd, 2008). In Arabic child acquisition, deaffrication was observed mostly amongst Kuwaiti Arabic children aged 2;8-2;11 (14%) (Alqattan, 2015). The error resolved around the ages of 3;0-3;3 (6%) and 3;4-3;7 (4%). It was also observed in Amayreh's (2003) study of older Jordanian children aged between 6;6 and 8;4. This is because this phoneme is only introduced in school when children (typically aged around 6;0) receive education in the *High* variety of the language vis-à-vis Educated Spoken Arabic. In studies of other varieties of Arabic (e.g. Ammar and Morsi, 2006; Ayyad, 2011; Dyson and Amayreh, 2002) including Libyan Arabic spoken in Misrata (see Chapter 3), this process is not reported simply because affricates are not part of the children's L1 phonemic inventory. In Misrata Libyan Arabic it is not a sound introduced in Educated Spoken Arabic through formal education either.

In word-initial voiceless affricates, this process was mostly observed in the Listen Only condition and was least occurring in the Listen and Speak condition. Whilst the frequency of this process decreased in the Listen and Speak and Traditional conditions in the delayed test, it seems to have increased for the Listen Only condition. The data gathered from the Traditional teacher indicate that her realisations were target-like for this segment in this position. The Listen Only condition is the only condition that did not practise speaking and was not *forced* to produce outputs. Learners in this condition did not have the opportunity to notice differences between the TL and their produced outputs and modify accordingly. In word-final voiceless affricates, it was solely observed in the experimental conditions and with considerably higher rates in comparison to the word-initial context. This once again supports the onset/coda asymmetry discussed earlier. This was in the post-test. In the delayed post-test, the Listen and Speak learners produced this sound correctly in all of the tokens, but the deaffrication persisted in the Listen Only learners' data. The fact that the Traditional condition did not show deaffrication is surprising since the realisations of the Traditional teacher both showed deaffrication. Instead, this group of learners realised the post-vocalic rhotic [ɾ] or [w],], for the single test item having a word final affricate 'march', which created a /-ɾʧ/ or a /-wʧ/ cluster. They relied on vowel epenthesis as their avoidance strategy. The next most attested process was vowel epenthesis between a stop and an affricate, which was the highest in Listen Only and Traditional conditions but not as frequent in the Listen and Speak condition. This seems to indicate that for these two conditions, the affricate was not perceived as a single segment. Rather, the cognitive representation for it is two sounds from the L1 /t/ and /ʒ/ merged together or in sequence. In the Traditional condition, this would be further exacerbated by the presence of the post-vocalic rhotic which consequently gives rise for a more complex coda cluster. In word-initial voiced affricates, the Traditional teacher exhibited a few instances of epenthesis preceding the affricate that was not shown in the voiceless cognates. Her realisations during the training are likely to have more of these instances given that in the data collected, she is aware of being recorded for analysis. For this context, Traditional condition showed the most cases of deaffrication, which again seems to be the most dominant process. However, in the delayed test, this declined considerably (by 50%). The experimental conditions seem to have similar rate of deaffrication, which can be explained by the nature of the input they received in comparison to that of the Traditional condition.

Dental fricatives – and rhotic approximants alike – on the other hand, had the lowest probabilities of a target-likeness compared to the remainder set of sounds in the present study. Studies of English child language acquisition indicate that these are amongst the later acquired sounds (Mcleod and Crowe, 2018).

For Arabic children, (inter)dental fricatives in Educated Spoken Arabic were also amongst the late sounds to be acquired by Arabic children (Amayreh, 2003). Amayreh (2003) provides two explanations for the late development of the interdental fricative in Arabic children. One reason, he argues, is the lack of input from Educated Spoken Arabic given that this sound is introduced in school when children are six years of age. Amayreh however, does not rule out the potential of Jordanian Arabic children being exposed to the sound from sources other than school. However, in schools, it is highly likely that teachers also use local variants instead of Modern Standard Arabic and this is why he observes substitutions in the children's data that reflect local variants. The other reason he suggests is the relative markedness of (inter)dental fricatives based on cross-linguistic typology. Indeed (inter)dental fricatives are segmentally more marked relative to the set of sounds examined in the current study.

The onset-coda asymmetries were observed for this sound class – amongst the Listen and Speak and Traditional condition but not the Listen Only condition – in the same way as affricates most probably for the same reasons. However, these observations are strictly descriptive as no statistical tests were carried out to make statistical inferences. For the phonological processes observed, the most common substitution for the experimental conditions was [f], most commonly referred to as *fronting*. About 7% of Arabic children substitution aged between 6;6 and 7;4 substituted target /θ/ in Educated Spoken Arabic with [f], [ʔ] and [s] or deleted it. Only 3% of children aged between 7;6 and 8;4 exhibited such processes. One explanation proposed for this observation is that TL /θ/ was perceptually mapped to the nearest acoustically similar L1 sound category, vis-à-vis /f/ as a result of *equivalence classification*. A study by Tabain (1998) examining the perceptual discrimination of fricatives by native Australian English speakers revealed that /f/ and /θ/ are the most likely fricatives to be confused. Her results show that [θ] was identified as /f/ in 29% of the cases and vice versa in 28% of the cases with no significant inter–speaker effects. Spectra below 10 kHz, for both [θ] and [f] was to a certain extent flat with no detectable peaks. Due to the low intensity, [θ] is hardly distinguishable from a non-sibilant fricative like [f]. It is argued that distinction between the two fricatives is based on the consonant onset and offset transitions (Harris, 1958; Schleef and Ramsammy, 2013; Tabain, 1998). Tabain's (1998) findings are supported by the findings of other studies such as Jongman, Wayland, and Wong (2000). A similar finding was found in L2 learners in a study by Syed (2013) who shows that advanced Pakistani learners of English perceived English [θ] as [f]. His production data however, showed that learners produced target [θ] as [t̪ʰ t̪] indicating an asymmetry between perception and production.

The limited ability to distinguish between /f/ and /θ/ is also proposed as a driver for language change. The neutralisation between the two phonemes is referred to as th-fronting in Cockney (Wells, 1982) and in language change studies and is also explained based on perceptual grounds (Blevins, 2004; J. Ohala, 1993b; J. Ohala, 1993a) especially in deprived conditions.[62] Due to this perceptual similarity, it is argued that target /θ/ is perceptually equated to the nearest L1 category /f/ under the influence of equivalence classification in SLM (Flege, 1995) and the magnet effect in NLM (Kuhl, 1991). Such a process is also observed in English child language acquisition (Vihman, 1996).

As for the substitution with [s], there could be several reasons for that. One of the reasons is that [s] is *phonologically* the closest sound to [θ]. The only feature differentiating [θ] from [s] is stridency. Whilst this feature is activated in English, it is not the case for Libyan Arabic. Stridents are produced by forcing air against two surfaces creating a high intensity frication noise. Non-stridents do not require the same level of complexity of constriction according to Halle and Clements (1983: 7). Another possible reason is attributed to the sound quality of DigLin recordings, which are very likely to be sampled at lower frequencies causing a loss of the energy observed in the higher frequencies region (above 10 kHz) (Tabain, 1998). Several test items containing /θ/ sounded like [s] even to native speakers. The sibilant [s] has a sharp low frequency cuttoff according to Ladefoged and Maddieson (1996: 177). Figure 10.1 below shows differences in the spectra of various fricatives as illustrated in Wester, Gilbers, and Lowie (2007).

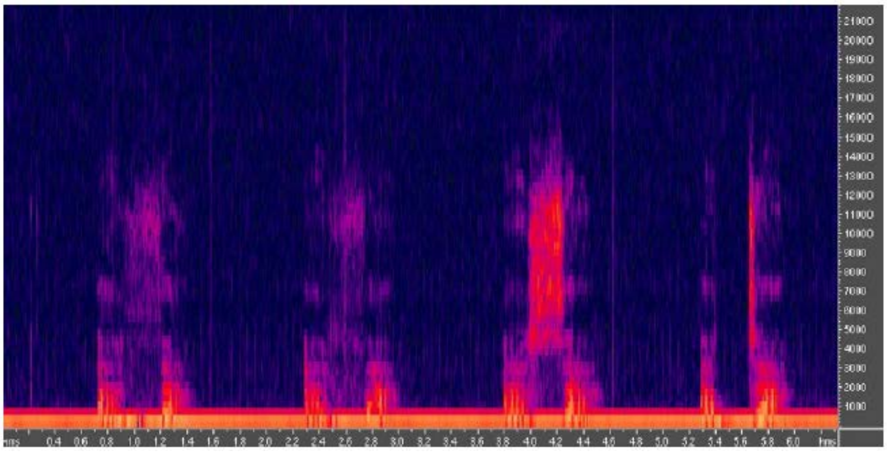


Figure 10.1 Spectrogram of [ɑθɑ], [ɑfɑ], [ɑsɑ], and [ɑtɑ] (x-axis: time(seconds); y-axis: frequency(Hz))

The Listen Only condition seems to reflect this substitution the most indicating that

---

[62] Milroy (2007) proposes that th-fronting is due to markedness of /θ/ cross-linguistically. However, it can be argued that these two accounts do not necessarily clash. The rarity of /θ/ in the world's languages may well be the result of its low perceptual saliency causing a /f/-/θ/ neutralisation.

their productions are mainly explained by their training condition. However, several learners managed to produce the correct output regardless of this technical fault. It is assumed that these learners may have relied on orthography to discern appropriate productions by comparing letters to various realisations from DigLin. Rosenblum (2008) argues that speech perception is multimodal, that is learners not only rely on auditory cues, but also visual cues. A visual cue relevant here is orthographic input. DigLin allows learners to listen to whole words or individual sounds corresponding to sound segments.[63] However, this assumption is made with some degree of discretion. The learners are reported to not have had formal instruction in English prior to the training. However, this does not mean that they have not developed some degree of literacy during training. The DigLin programme is designed for illiterate (and low-literate) users (Cucchiarini, Dawidowicz, et al., 2015) which makes it an ideal candidate for training children with no prior knowledge of the English orthographic system. A recent development to the programme was designing a sound bar at the bottom of the exercise screen. The bar has single graphemes, digraphs and trigraphs. The purpose of this tool is to allow learners to build letter-to-sound associations (Cucchiarini, Dawidowicz, et al., 2015). See figure 10.2 below:

Figure 10.2 English sound bar in DigLin (Overal, 2014)

Another process observed for this sound class is stopping [t] and affrication [tθ]. The former is the most widely attested substitution in Libyan Arabic adult learners of English and was observed for several tokens produced by the Traditional teacher. It is also widely attested in L2 learners from other linguistic backgrounds such as Yoruba speakers (Owolabi, 2012) and in New Zealand Pasifika English (Bell and Gibson, 2008). [t] substitutions in particular are considered the least *segmentally marked* of all substitutions. According to Wester et al. (2007), /t/ is classed a strident and the only feature distinguishing it from /s/ is continuancy. They support their argument based on evidence from a pilot experiment on categorical perception. In this experiment, the

---

[63] In DigLin, users can play whole words by clicking on the larger circle or individual sounds corresponding to sound segments by clicking the smaller circle. They also have access to a sound bar at the bottom of the screen. See figure 5.3 in Chapter 5.

sound [s] in [seː] was shortened in ten steps whilst the high-intensity noise was kept intact. When [s] was shortened to 30 ms, listeners perceived it as a [t]. preserving intensity, they argue that from an acoustic perspective the sole difference distinguishing [s] and [t] is continuancy. Child language development studies show that children prefer plosives (non-continuants) initially and fricatives (continuants) finally (van der Linde, 2001)

These processes are also attested in the speech of first generation Polynesian immigrants in New Zealand. Bell and Gibson (2008) found that the most common processes for target [θ] were stopping and fronting. Affrication was also observed albeit with lower degrees. The interdental fricative does not exist in the phonemic inventory of the Pasifika adstrate language. Surprisingly the affricate variant observed does not exist in the native language either. As for rhotics, the Listen and Speak condition once again surpassed the Listen Only and Traditional conditions in that it had the highest probability of TL-like and matching productions. Although no statistical difference was discerned for TL-likeness between the experimental conditions, a statistical difference was observed in matching productions. What is surprising is the lack of statistical difference between the Listen Only and Traditional conditions in the post-test for the match probabilities. The Traditional condition further outperformed the Listen Only condition in the delayed test, although not statistically. This ultimately indicates the crucial role of output practice in pronunciation challenging the claims of Krashen's (e.g. 1982; 1985; 1994) Comprehension/Input Hypothesis.

Rhotic approximants, like dental fricatives, seemed one of the most challenging sounds for the learners. Based on cross-linguistic typology, rhotic approximants are segmentally more marked relative to rhotic taps/flaps (Maddieson, 1984). Thus, the segmental markedness of /ɹ/ could explain the relative difficulty learners faced with this sound class. They are one of the late developing sounds in many languages. Onset/coda asymmetries are also observed in the acquisition of rhotic approximants albeit for different reasons than those mentioned above. First of all, SSBE is a non-rhotic dialect. Therefore, rhotics cannot occur post-vocalically. At face value, the data in Chapter 9 may indicate that learners mastered the word-final position prior to word-initial position. However, this cannot be accepted for a number of reasons, of which the non-occurrence of rhotic approximants post-vocalically in SSBE and thus markedness constraints on codas or H&H assumptions are not applicable. The other reason is the absence of /r/ does not always indicate correct production. Coda deletion is a developmental process irrespective of the consonant albeit for younger children. For example, Alqattan (2015) found that the tap rhotic was deleted in word-final position 13% of the time and post-vocalic /l/ was also deleted 11% of the time by children in her study of the developmental processes in Kuwaiti Arabic children. The commonality between these two sounds is that they fall under the class of liquids. A great deal of final /l/s were also deleted in the current data supporting the findings of Alqattan (2015) and several other studies on English child language acquisition. Post-vocalic /l/ in SSBE English is velarised. The articulatory complexity associated with the secondary velarisation could well explain this phenomenon. The Traditional teachers realisations mostly showed a substitution with the L1 tap [ɾ]. We are not claiming the teacher's interlanguage is a case of L1 transfer. Rather, we argue for an interlanguage based on accented input given the teaching methodologies followed in Libyan classrooms, which are deprived of exposure to native language input. This confirms the input-based explanation by Young-Scholten (1995). In terms of effect over time, we see that whilst the probability of matching realisations for the experimental conditions dropped with time, that for the Traditional condition improved but does not rise to the level of Listen and Speak condition who continued to statistically outperform Traditional learners and outperform Listen Only learners. This indicates that the effect of listening/speaking practice continues to prove optimal results for pronunciation in comparison to the other training conditions.

A common processes illustrated in the /ɹ/ data was the presence of a rounded back vowel preceding the target sound. This process was divided into two distinct categories; one category involves the insertion of a rounded back vowel preceding a TL-like realisation of [ɹ]; the other one involves the insertion of a rounded back vowel followed by substitution. The first type was most common amongst the Listen and Speak condition followed by the Listen Only condition. This is an attempt to achieve lip-rounding which is usually associated with American English [ɹ] (Alwan, Narayanan, and Haker, 1997; Delattre and Freeman, 1968; Ladefoged and Maddieson, 1996).

A higher proportion of the non-TL-like data also showed a presence of a rounded back vowel that was accompanied with substitution of the target rhotic, yet the biggest proportion of the /r/ data exhibited substitution only. The nature of these substitutions reveals a great deal about the nature of acquisition development. The most widely attested substitution is that of [w]. It seems that learners detected the labiality of SSBE /r/. However, this was mostly evident in the Traditional condition in the post-test. This is a time when this condition was first introduced to such an amount of targeted exposure to native input. One possible explanation for this substitution is that this feature is the first to be acquired in the L2 acquisition of English [ɹ]. The fact that this type of substitution declines and is replaced by a different type of substitution in the delayed test further supports this argument. Surprisingly, the [w] substitution amongst the Traditional condition exceeds that of the closest L1 category [ɾ] and [r] or that of their teacher's realisations. The possibility that their teacher had different realisations during the training is a remote one. The teacher showed her capability of producing a TL-like rhotic in a few examples she produced. It is worth mentioning here that the Traditional Teaching condition group not only differed from the experimental groups in terms of type of input but that the participants also had access to social interaction with the source of that input from the teacher, unlike the experimental condition groups which only had access to prerecorded input but did not have a chance to interact with the speakers. Research demonstrates that infants as young as 9-10 months old benefit from face-to-face exposure indicating that phonetic learning is improved by social interaction, something that was missing from exposure to prerecorded materials, more so than from long-term perceptual training (Kuhl, Tsao, and Liu, 2003). However, it is argued that she produced those tokens knowing that she is being recorded for analysis purposes, unlike in the training sessions, whereby she had a more relaxed environment and more likely to produce L1-like realisations (as evident in the majority of her /r/ tokens). Instead, it is argued that some sounds do not require perceptual training to discern. The acoustic/articulatory difference between L1 and TL /r/ is beyond perceptual confusion. In other words, [ɾ] is categorised as *different* and thus not warped or equated by the magnet effect or equivalence classification. It is perceived instantly. The obstacle for these child learners is an articulatory one and this is why some TL sounds continue to sound non-native-like/non-TL-like even after years of L2 experience. The fact that Traditional learners deviated from the input they received as soon as they heard the auditory stimulus proves this point. Substitution with [ɹ] and [r] seems to reflect L1 transfer in the experimental conditions and this seems to be exacerbated with the type of input received for the Traditional conditions. A small proportion of less frequent substitutions include other variants of various rhotics that do not exist in either Libyan Arabic or English proving further that rhotics are a natural sound class (Boyce et al., 2016; Ladefoged and Maddieson, 1996). Substitution with [l] is also present and especially notable in the Traditional condition. This indicates the consideration of target /ɹ/ as a liquid.

Results from final clusters show that there is great between-cluster type variability in the acquisition of coda CC clusters. Generally, Listen Only seems to have the lowest advantage over the other two training conditions in /-ld -lk -nt -nd/ whilst Listen and Speak seems to be the least advantageous for /-mp/ learning though only in the post-test. The Traditional teaching condition seems to have the lowest TL-like productions in /-ft -st -nz/ whilst this condition had the greatest advantage in /-lk/. The commonality between the first two cluster types in the Traditional condition is that the first member is a fricative and second is a plosive. The poor performance of the Traditional condition however, cannot be explained by the accented input received for these cluster type. The realisations of the Traditional teacher seemed stable for all the cluster types with the exception of the alveolar (*nasal + plosive*) final clusters, /-nt -nd/.

Data from /-lt/ did not show great between-condition variability in accurate realisations. However, in /-lC/ clusters such as /-ld, -lt, -lk/ there was a great cross-cluster type variability which can be in part attributed to the phonological structure of the word as well as the number of target items for each. For example, /-ld/ and /-lk/ each had only one test item, vis-à-vis 'cold' and 'milk' respective, unlike /-lt/ that had three test items and was practised more often than the previous two. Research on the effect of frequency shows that L1 English children are sensitive to frequency of clusters at the phonotactic level (Kirk and Demuth, 2005: 724). Frequency effects seems to explain the observed higher scores for /-lt/ compared to /-ld/ but not compared to /-lk/. As for /-ld/, a great deal of the reduction can be explained in the presence of diphthong preceding the cluster. It is difficult to discern whether the [-ʊ-] is part of the diphthong or a vocalisation of the /ɫ/. Generally, for the lateral + plosive clusters, the issue is [ɫ] post-vocalically which seems to be problematic especially for the experimental learners given that the only form consistently introduced to them is [-ɫ-].[64]

Some of the commonly observed processes for the *lateral + plosive* clusters is /l/ vocalisation. This is most evident in the experimental conditions. This does not indicate that the Traditional condition performed better than the experimental conditions TL-likeness-wise. Instead, it shows how the experimental condition's outputs were faithful to the input they received. This is because it is widely attested in the literature on English child language acquisition that children go through a stage of post-vocalic /l/ vocalisation (Dodd, Holm, Crosbie, et al., 2013). This cluster type also revealed a great deal of reduction, where the first member is elided and the second one is maintained. In the delayed test, generally Listen Only condition again had considerably the lowest accuracy rates. The data for the coda CC cluster confirm HYPOTHESIS TWO.

10.1 Lexical learning

The data show that although Listen and Speak condition yielded the most fully learned words amongst the three training conditions followed by Listen Only, the difference between the experimental conditions exhibited in the delayed repetition and picture-naming tasks was statistically non-significant. Traditional teaching yielded the least amount of fully learned words and the mean number of learned words was statistically significant from either experimental condition in the delayed repetition and picture-naming tasks. In the read aloud tasks, all training conditions performed poorly and there were no statistically significant differences amongst them. However, had the audio clips in the delayed repetition task been based on the Traditional teacher's productions, it is possible we would see the reverse pattern for experimental vs. traditional training conditions. To corroborate this point, in the picture-naming task – which preceded the delayed repetition task, before which the Traditional learners are potentially directly exposed to native input – the Traditional learners outperformed the experimental learners and the difference reached statistical significance (see

---

[64] Even though data was collected from the Traditional teacher shows that she produced [ɫ], we cannot reliably conclude that this was the only form to which Traditional learners were exposed. This is evident in her realisations of the test item 'belt' consistently as a plain lateral and the variability observed for the token 'salt' also realised once as a plain lateral and as a velarised lateral in another elicited token. See Chapter 9, sections 9.3.2 and 9.3.2.

Chapter 6). In the read aloud task, which also preceded the Traditional participants' direct exposure to native input, the difference in performance across the three training conditions does not reach statistical significance. The Traditional teaching learners seem to have picked up from the input of the audio clips during the delayed repetition task. This is also evident in their overall improved performance on target-likeness exhibited in the delayed test as discussed in Chapter 6. It is worth mentioning that even though the performance of participants in the Traditional Teaching condition improved slightly over time, this improvement was not significant. Neither was it consistent for all problematic sounds. It was namely seen in clusters, dental fricatives, and diphthongs. In affricates and plosives, their performance declined. Additionally, as much as the participants in the Traditional Teaching improved in the delayed post-test, they did not statistically outperform the participants in the Listen and Speak training in any of the problematic sounds. They have only outperformed the participants in Listen Only training condition a) in the target-likness analysis in the class of dental fricatives, albeit the difference did not reach statistical significance and b) in the match rating analysis in the class of rhotic approximants, where once again the difference did not reach statistical significance (see summary table 6.7).

As far as the children's slightly better performance in the picture-naming task compared to the read aloud task, it could be due to children's general preference for learning through images than written text. Some studies exploring presentation modes show that presenting learning materials in pictures alongside words seems to improve levels of comprehension and information storage abilities (Tindall-Ford, Chandler, and Sweller, 1997). Images have been argued to facilitate memory recall of items stored in long-term memory as a result of processes of dual coding leading to associative learning (Acha, 2009).

10.2 Conclusion

Overall, results for the segmental, phonotactic, and lexical learning domain show an overall disadvantage for the delayed production practice (Listen Only) in comparison to the Listen and Speak training when it comes to pronunciation. Whilst this was the overall trend, exceptions were found indicating an advantage for delayed oral practice for specific sounds in specific phonological contexts, such as word-final dental fricatives and /ɪə/. This indicates that whilst listening/speaking practice was important for better pronunciation accuracy, it should not be treated uniformly across all problematic sounds. Results for coda cluster types demonstrated that for certain cluster types (/-ld/), listening/speaking practice overrules the impact of native speaker input, whereas for other cluster types, there is an advantage for traditional training practice over CAPT training. One explanation for this is that the Traditional training condition group had the advantage of social interaction over the other training conditions. The findings of Kuhl, Tsao, et al., 2003 indicate that infants who had social interaction during exposure to a foreign language showed perceptual learning advantage over prolonged perceptual training with prerecorded foreign language materials. For other clusters there is an advantage for native speaker input when combined with listening and speaking practice. In phonetic learning, results from VOT, f0, and spectral tilt analyses show that none of the training conditions created statistically independent within-voiceless or within-voiced L2 phonetic categories. Instead, all of the three conditions created merged L2 categories that reflect L1 and TL values. The L2 values were, however, intermediary indicating learning did take place in all training conditions. Moreover, they all successfully managed to statistically distinguish their voiced and voiceless categories including bilabial plosives, of which the voiceless one is absent in the learners' L1 phonemic inventory and especially considering that native English speakers find it difficult to identify it when produced by Arabic adult speakers (Flege and Port, 1981). This also shows no advantage of one training technique over the other, except that with minimal native input, Arabic children show evidence of English VOT learning. Even though the study did not seek to examine differences in L1 VOT durations, interestingly the results for L1 VOT durations show inconsistencies across the training conditions despite having a shared L1 background. The average L1 VOT values for those following the Listen and Speak training lean relatively more towards the positive end of the VOT continuum, whereas those following the Traditional Teaching leaned relatively more to the negative end of the scale, with those following the Listen Only training somewhere in between (see figure 7.1). Despite the lack of data from learners prior to the training and the relatively short period of training, and in the absence of any other explanatory factors, we argue that the experimental learners' L1 are altered as a result of the merger category supporting the findings by Flege (1995) and MacKay et al. (2001) and the notion of *backward transfer* by Kartushina, Hervais-Adelman, Frauenfelder, and Golestani (2016), whose findings demonstrate evidence of mutual influences between L1 and L2 sound categories in production after short-term visual articulatory feedback training.

Colantoni and Steele (2008) assert that not all sounds are equally difficult. For the Traditional condition not having a statistical difference from the experimental conditions, PAM puts forward that some TL sounds are readily perceiveable without training necessary. Furthermore, the majority of theories of speech perception/learning are not enough to account for cross-linguistic difficulties in L2 learning. The varying degrees of difficulty observed for the different problematic classes observed in the current study support the findings of Colantoni, Steele, and Escudero (2015), which demonstrate that articulatory constraints and typological markedness also play a role in L2 acquisition. It also supports the role of phonetic/phonological context on L2 learning.

Indeed as proponents of output practice such as Swain (1995) and Swain (2005) for general L2 learning and Colantoni, Steele, and Escudero (2015) for L2 speech learning postulate, output practice allows learners to reflect on their productions, comparing them to the TL in a continuous refining process. Output also allows learners to develop articulatory muscle memory for the target sounds and structures with more practice comes better tuning of articulatory-motor skills. Differences between child adult L2 learning cannot be explained merely by age as previous studies mostly emphasise but rather on type of input. We have seen how learners of the same age (prepubescent) can have different performances resulting from differing types of input. Not all IL phonology is governed by transfer, some aspects are explained by universal language acquisition tendencies regardless of the age group. Universal tendencies also relate to acoustic salience and articulatory settings. We have also seen that in the delayed test, whereby the Listen and Speak condition maintained its rank in target-likeness/matchness and the Listen Only condition's performance levels out with that of the Traditional condition who received accented input.

Chapter 11: Conclusion

11.1 Limitations

In Chapter 10, it was demonstrated that the two experimental groups outperformed the Traditional group on a number of pronunciation aspects showing that input type plays a role in L2 phonological – albeit not L2 phonetic – learning. Moreover, the Listen and Speak training outperformed the two other conditions on the whole demonstrating the crucial role output practice plays in shaping L2 pronunciation. However, the present study has several limitations worth noting.

11.1.1 *Lack of pretest data in L1 Arabic*

Although we stated that there were no statistically significant differences in the mean L1 VOT across training conditions, there was noticeable cross-condition variability. Specifically, the participants in the experimental conditions generally exhibited 1) shorter mean lead voicing for the Arabic voiced plosives and 2) longer VOT for the Arabic voiceless plosives, compared to the Traditional teaching condition. Between the experimental conditions, we see the Listen and Speak training condition exhibiting a similar pattern. This is puzzling. Because all of the participants come from the same city where they were born and raised, it is unlikely that the reason for this variability stems from their linguistic background. Instead, I speculate that English was influencing their Arabic, at least during the period when the study took place. (Flege, 1995; MacKay et al., 2001). It is, however, not possible to verify this. An attempt was made to collect Arabic data prior to instruction. However, this was not feasible for two reasons. First, data collection at the outset of the training would have risked the participants withdrawing from the training. This was especially the case after considering how keen parents were that their children started learning English from the very first day of the training. The fact that some parents withdrew their children in the final day(s) of the training and right before data collection confirms this anticipation. Second, Arabic data were further collected in the delayed post-test to ensure that L1 and L2 data were comparable in terms of age and to capture L1 development 10 weeks after the initial data collection stage. In this case, the impact of L2 on L1 would have been inevitable given the children had already been exposed to English 10 weeks prior, that is during the training.

11.1.2 *Lack of perception data*

We have seen how the children from the three training conditions varied in their L2 pronunciation performance. However, we did not gather data on their L2 perception. We know that the respective training that participants in each condition received plays a role in their production, but we do not know how this was grounded in their perception. There were also access issues and testing conditions which hindered the collection of perception data that I discuss in section 11.1.4.

11.1.3 *Lack of data from Traditional teacher during training*

The data collected from the Traditional teacher was conducted in a laboratory context. We cannot argue with great certainty that she did not modify her speech as a result of this context or that her performance during data collection reflected that in the classroom (Cucchiarini, Strik, and Boves, 2002).

11.1.4 *Access and the testing conditions*

Access was an issue for the current study. Libya, where the study took place, was considered a risk zone and thus I was not permitted to conduct or supervise the trainings process on-site. Instead, assistant researchers were trained on Skype to carry out the training and data collection on my behalf. Owing to the unstable state of affairs within the country, some of the sessions were interrupted either by blackouts, losing internet access or both.

Due to the difficulties with testing participants remotely, this study could not provide the ideal environment for testing Krashen's Comprehensible Input Hypothesis using the Listen and Speak vs. Listen Only conditions. An ideal environment for such model would require an extended training period of immersion learning. As such the model could not be tested for validity. The findings thus apply solely for the limited training period of three weeks and the testing which took place ten weeks later.

11.1.5 *Data elicitation tasks*

We have seen how the data from the read aloud and picture-naming did not yield enough tokens for analysis. That is because children did not develop adequate literacy in English. Children could not identify most of the words in the picture-naming task either, suggesting that perhaps three weeks of training were not enough for this age group. Deliberations between the assistant researchers, myself and supervisors led to resorting to a delayed repetition task. Whilst this technique for data elicitation is widely used, it comes with its own disadvantages. Despite using an intervening phrase to ensure that participants did not rely on phonetic-sensory memory, the Traditional teaching condition group palpably relied on mimicry to produce the tokens prompted by audio clips used for this task from DigLin. This was also demonstrated in their improved performance in some aspects in the delayed post-test.

11.1.6 *DigLin*

Two main issues were observed in the CAPT software, DigLin. First, the audio files did not seem to have been recorded with an appropriate sampling frequency. This had rendered issues, especially with some of the sounds files containing /θ/ and /p/. It seems the sound files were generally not scrutinised as words beginning with /ɹ/ sounded as though they were preceded by a bilabial plosive transient. On a different but related level, the practice items did not exhibit the full range of distributional possibilities for English. For example, the velar plosives could not be considered in the training since there were not enough comparable tokens in the voiced and voiceless contexts. Similarly, there seems to be an asymmetry with the distribution of sounds across various phonological contexts in the practice material. For example, whilst some sounds such as /θ/ were available in enough practice items word-initially and -finally, others, were predominant (or non-existent) either word-initially or word-finally. Furthermore, whilst it was not possible to consider a comparison in learning between voiceless /θ/ and voiced /ð/ due to the rarity of the latter in English monosyllabic words, examples such as ‘clothes’ or ‘bathe’ may have been sufficient for the purpose.

11.1.7 *Lexical learning*

We noted in Chapter 6 that full coverage of the participants’ lexical learning was beyond the scope of the present thesis. This does not mean it is unimportant.

11.2 Ideas for future research

In the present study, the statistical data were well supported by qualitative data. Whilst the rigorous experimental approach was appropriate, it is worthwhile to examine perception in future studies. This is to ascertain whether the participants’ performance in perception tasks would match that of their production. Further research is also required to examine traditional foreign-accented teachers’ productions during the classroom-based instruction to capture the actual input learners receive during training. Further research is also needed to examine the long-term training effect on perception and production. It is also recommended that CAPT training software packages follow rigorous procedures for creating their audio and visual materials. Furthermore, further research is required to determine the amount of training child learners require to develop decoding and to learn test words efficiently. Moreover, theses on the acquisition of L2 phonology by children are relatively rare compared to other domains of L2 acquisition. Moreover, the ones that do exist rarely cover lexical learning in any depth. Moreover, there are few studies of children’s early learning of vocabulary in the classroom (an exception is Heimbach, 1994, for example). This should not be the case, and needs to be pursued.

11.3 Implications and recommendations

It is well-known that hours dedicated to L2 teaching in school classrooms is insufficient for intensive language training. In Libya – as in many other EFL contexts around the world – class time is quite limited (80-90 minutes per week) and is usually more focused on grammar, other skills and exam-directed practice. When practice of oral/aural skills does take place, it is quite limited, and students are anxious of making mistakes in front of their peers. Students often refrain from class participation in fear of peer derision or judgement. Teachers very rarely make use of native speaker input and rely on their own pronunciation which is non-native in drilling tasks (c.f. section 1.3). Eliminating or reducing accented L2 English in EFL classrooms seems very challenging given the current status of the Libyan education system and Libyan teachers' perspectives on teaching pronunciation. The use of Computer-Assisted Pronunciation Training software can help relieve some of these issues by allowing learners to work independently, at their own pace, and with reduced anxiety from peer derision. Such software can be used as supplemental practice side by side with traditional teaching classrooms. Moreover, second language teachers should focus on the importance of oral practice and supplying lessons with native speaker input. This is especially the case for pronunciation improvement. Gains in perception and gains in production can also be discerned from the findings of this study and teachers/educators should not treat problematic sounds uniformly. Interlanguage phonological processes are to be expected as part of the learning process. They are evident in English-speaking children and thus, do not necessarily reflect a lack of learning.

Appendix A: Consent Form

The Impact of Computer-Assisted Pronunciation Teaching on Libyan Child Learners of English

A Research Project by: Hana Farid Ehbara

Consent Form for Parents

Name of Child:................................	Date of Birth: ....
Name of Parent/Guardian:	

Please, indicate that you have read and agree with the statement by putting your initials in the following boxes:

1. I confirm that I have read and understood the information sheet for the above project and have asked any questions I wanted to.
2. I understand that participation by my child is voluntary and that she/he is free to withdraw at any time, without giving a reason.
3. I understand that all information collected
    - will be labelled with a number and not my child's name to maintain anonymity.
    - will be stored in a secure place.
4. I understand that if my child discloses information that indicates he/she may be at risk it will be followed up through established channels.

Parent/Legal Guardian Signature.....................Date.....................

Researcher: Hana Farid Ehbara, Signature ....

Please give the form to Hana Ehbara or the school office. You will be given a copy of this consent form to keep.

Appendix B: Information Sheet

The Impact of Computer-Assisted Pronunciation Teaching on Libyan Child Learners of English

A Research Project by: Hana Farid Ehbara

Information for Parents

Your child is being invited to take part in a research study. This letter has important information about the reason for doing this study, what I will ask your child to do, and the way I would like to use information about your child if you choose to allow your child to be in the study. Please, take time to read the following information carefully and discuss it with others if you wish.

What is the purpose of the study?

The aim of the study is to explore different training techniques that can help improve pronunciation of children in early school life. The findings of this study will be an important piece of new knowledge about the effectiveness of supporting language classes with computer assisted pronunciation training in early school years. The study is being undertaken by the researcher in completion of her PhD. It will be written up as a PhD thesis and the findings will be submitted for publication in peer reviewed journals and/or conferences.

What will my child be asked to do in this study?

Should you decide to allow your child to take part in this research study, and all your questions have been answered to your satisfaction, you will be asked to sign a consent form. The training will last three weeks, five days a week. Each session will last one hour with a break in the middle. During the sessions, your child will learn new words, how to say them, how to write them and their meanings through picture associations. Then your child will be recorded twice. The first recording will take place immediately after the training. The second recording will be four weeks later. Both recording sessions will last up to 30 minutes. The training and recording will take place during the summer holiday in one of Misrata's language centres.

What are the possible benefits for my child or others?

You will be given a report outlining the training results and what they mean in relation to your child. Your child will learn 300 English words. Learning involves pronunciation, spelling and meaning. The results will have significance for second language pronunciation teaching in the in early school years. This will have implications for national curriculum planning.

How will you protect the information you collect about my child?

We will ensure that your child's results are confidential. Your child will be identified in the research records by a code, for instance, as (participant CG1). All data will be kept confidentially in a password-protected personal external drive. Only the researcher can access the data.

What are my child's rights as a research participant?

Your child's participation is entirely voluntary. She /he is free to choose not to participate. Should you and your child choose to participate, he/she can withdraw at any time without consequences of any kind.

Who is conducting the Research?

The researcher is Hana Ehbara. She is from Libya. The research is also conducted by Fatma Suliman who will assist in the training.

Who can I contact if I have questions or concerns about this study?

If you have questions or concerns during the time of your child's participation in this study, please contact:

Hana Farid Ehbara (researcher)
Email: h.f.o.ehbara1@newcastle.ac.uk
Tel. +44(0)7749191244

Prof. Martha Young-Scholten (supervisor)
martha.young-scholten@newcastle.ac.uk
Tel. +44 (0) 191 208 7751

Dr Jalal Al-Tamimi (supervisor)
jalal.al-tamimi@newcastle.ac.uk
Tel. +44 (0) 191 208 5208

Thank you for taking the time to read this letter. We hope you will give permission for your child to take part in this important and exciting study.

Appendix C: Most Recalled Words

```
   item Instruction       n
 1 dog  Traditional       52
 2 dog  Listen and Speak  50
 3 dog  Listen Only       47
 4 pad  Listen and Speak  42
 5 left Traditional       41
 6 bee  Traditional       40
 7 jam  Traditional       39
 8 jet  Traditional       38
 9 red  Listen and Speak  38
10 dish Traditional       37
11 milk Traditional       37
12 pad  Listen Only       37
13 bat  Listen and Speak  35
14 bee  Listen and Speak  35
15 hair Listen and Speak  35
16 list Traditional       35
17 pea  Traditional       35
18 bat  Listen Only       34
19 dish Listen Only       34
20 jar  Listen and Speak  34
21 moth Listen Only       34
22 pile Listen and Speak  34
23 rip  Listen and Speak  34
24 bat  Traditional       33
```

```
25 bush   Listen Only         33
26 chair  Listen and Speak    33
27 chair  Listen Only         33
28 hair   Listen Only         33
29 jog    Traditional         33
30 path   Listen and Speak    33
31 belt   Listen and Speak    32
32 chin   Listen Only         32
33 chip   Traditional         32
34 dish   Listen and Speak    32
35 path   Listen Only         32
36 pea    Listen and Speak    32
37 read   Listen and Speak    32
38 three  Listen and Speak    32
39 tooth  Listen Only         32
40 bush   Listen and Speak    31
41 chin   Listen and Speak    31
42 chin   Traditional         31
43 chip   Listen and Speak    31
44 jam    Listen Only         31
45 left   Listen Only         31
46 rip    Listen Only         31
47 rob    Listen and Speak    31
48 thin   Listen Only         31
49 waist  Traditional         31
50 bee    Listen Only         30
51 chip   Listen Only         30
52 chop   Traditional         30
53 list   Listen Only         30
54 milk   Listen Only         30
55 salt   Traditional         30
56 tea    Listen and Speak    30
```

57	thread	Listen and Speak	30
58	chop	Listen and Speak	29
59	jam	Listen and Speak	29
60	left	Listen and Speak	29
61	pile	Listen Only	29
62	red	Listen Only	29
63	tail	Listen and Speak	29
64	three	Traditional	29
65	chop	Listen Only	28
66	fear	Listen Only	28
67	jar	Listen Only	28
68	list	Listen and Speak	28
69	milk	Listen and Speak	28
70	rice	Listen and Speak	28
71	tail	Listen Only	28
72	tea	Listen Only	28
73	wreck	Listen and Speak	28
74	chest	Listen and Speak	27
75	march	Listen and Speak	27
76	moth	Traditional	27
77	root	Listen and Speak	27
78	tap	Listen and Speak	27
79	tool	Listen Only	27
80	vest	Listen and Speak	27
81	wrist	Listen and Speak	27
82	wrong	Listen and Speak	27
83	bath	Listen Only	26
84	belt	Listen Only	26
85	boil	Listen Only	26
86	fear	Listen and Speak	26
87	jog	Listen Only	26
88	pea	Listen Only	26

89	tap	Traditional	26
90	thin	Listen and Speak	26
91	thumb	Listen and Speak	26
92	chest	Traditional	25
93	day	Listen Only	25
94	jog	Listen and Speak	25
95	tap	Listen Only	25
96	tea	Traditional	25
97	thumb	Listen Only	25
98	tool	Listen and Speak	25
99	tooth	Listen and Speak	25
100	wreck	Listen Only	25
101	beard	Listen Only	24
102	boil	Listen and Speak	24
103	cold	Listen and Speak	24
104	hear	Listen and Speak	24
105	jet	Listen and Speak	24
106	moth	Listen and Speak	24
107	read	Listen Only	24
108	breast	Listen and Speak	23
109	day	Listen and Speak	23
110	jeans	Listen and Speak	23
111	jug	Listen Only	23
112	hear	Listen Only	22
113	jeans	Listen Only	22
114	jet	Listen Only	22
115	rob	Listen Only	22
116	thigh	Listen Only	22
117	thin	Traditional	22
118	vest	Listen Only	22
119	wrist	Listen Only	22
120	round	Listen and Speak	21

121	thigh	Listen and Speak	21
122	beard	Listen and Speak	20
123	day	Traditional	20
124	push	Traditional	20
125	tooth	Traditional	20
126	vest	Traditional	20
127	coin	Listen and Speak	19
128	jug	Listen and Speak	19
129	pine	Listen Only	19
130	pound	Listen and Speak	19
131	pound	Traditional	19
132	push	Listen Only	19
133	salt	Listen and Speak	19
134	sound	Listen and Speak	19
135	wrist	Traditional	19
136	broth	Listen and Speak	18
137	cold	Traditional	18
138	jug	Traditional	18
139	mouth	Listen and Speak	18
140	roll	Listen and Speak	18
141	root	Listen Only	18
142	coin	Traditional	17
143	mouth	Listen Only	17
144	pine	Listen and Speak	17
145	pine	Traditional	17
146	pool	Listen Only	17
147	rice	Listen Only	17
148	thread	Listen Only	17
149	salt	Listen Only	16
150	broth	Listen Only	15
151	bush	Traditional	15
152	chest	Listen Only	15

153	rose	Listen and Speak	15
154	jump	Traditional	14
155	march	Listen Only	14
156	pad	Traditional	14
157	push	Listen and Speak	14
158	quilt	Listen and Speak	14
159	red	Traditional	14
160	rice	Traditional	14
161	root	Traditional	14
162	sound	Traditional	14
163	wrong	Listen Only	14
164	bath	Listen and Speak	13
165	coin	Listen Only	13
166	jump	Listen and Speak	13
167	pool	Listen and Speak	13
168	rose	Listen Only	13
169	round	Traditional	13
170	tool	Traditional	13
171	waist	Listen and Speak	13
172	waist	Listen Only	13
173	bath	Traditional	12
174	breast	Listen Only	12
175	cold	Listen Only	12
176	jump	Listen Only	12
177	point	Listen and Speak	12
178	rip	Traditional	12
179	wreck	Traditional	12
180	belt	Traditional	11
181	hair	Traditional	11
182	rain	Listen and Speak	11
183	read	Traditional	11
184	sound	Listen Only	11

185	tail	Traditional	11
186	wrap	Listen and Speak	11
187	hear	Traditional	10
188	pound	Listen Only	10
189	quilt	Listen Only	10
190	three	Listen Only	10
191	wrong	Traditional	10
192	mount	Traditional	9
193	rain	Traditional	9
194	roll	Listen Only	9
195	point	Traditional	8
196	rose	Traditional	8
197	round	Listen Only	8
198	breast	Traditional	7
199	broth	Traditional	7
200	mouth	Traditional	7
201	rob	Traditional	7
202	chair	Traditional	6
203	pile	Traditional	6
204	rain	Listen Only	6
205	boil	Traditional	5
206	pool	Traditional	5
207	wrap	Listen Only	5
208	jar	Traditional	4
209	path	Traditional	4
210	point	Listen Only	4
211	paint	Listen Only	3
212	jeans	Traditional	2
213	mount	Listen and Speak	2
214	thumb	Traditional	2
215	beard	Traditional	1
216	paint	Listen and Speak	1

```
217 thread Traditional           1
218 wrap   Traditional           1
```

Appendix D: Post Hoc Tests for Target-likeness Rating

Predicted Probabilities

Table D.1 Predicted probabilities of target-like rating

Instruction	Test	prob	SE	asymp.LCL	asymp.UCL
Affricates					
Listen and Speak	Post-test	0.681	0.092	0.483	0.830
Listen Only	Post-test	0.486	0.104	0.294	0.682
Traditional	Post-test	0.301	0.087	0.160	0.493
Listen and Speak	Delayed test	0.592	0.105	0.382	0.773
Listen Only	Delayed test	0.427	0.102	0.248	0.628
Traditional	Delayed test	0.291	0.086	0.153	0.482
Clusters					
Listen and Speak	Post-test	0.144	0.049	0.072	0.267
Listen Only	Post-test	0.099	0.037	0.047	0.197
Traditional	Post-test	0.066	0.025	0.031	0.137
Listen and Speak	Delayed test	0.151	0.053	0.073	0.286
Listen Only	Delayed test	0.046	0.019	0.020	0.100
Traditional	Delayed test	0.076	0.029	0.035	0.156
Dental fricatives					
Listen and Speak	Post-test	0.130	0.049	0.060	0.259
Listen Only	Post-test	0.101	0.041	0.045	0.213
Traditional	Post-test	0.066	0.028	0.028	0.145
Listen and Speak	Delayed test	0.116	0.047	0.051	0.242
Listen Only	Delayed test	0.068	0.029	0.030	0.151
Traditional	Delayed test	0.083	0.034	0.037	0.180
Diphthongs					
Listen and Speak	Post-test	0.236	0.079	0.115	0.421
Listen Only	Post-test	0.193	0.069	0.091	0.364
Traditional	Post-test	0.025	0.012	0.010	0.063
Listen and Speak	Delayed test	0.228	0.080	0.109	0.418
Listen Only	Delayed test	0.228	0.078	0.110	0.413
Traditional	Delayed test	0.043	0.019	0.018	0.102
Plosives					
Listen and Speak	Post-test	0.435	0.099	0.259	0.628
Listen Only	Post-test	0.370	0.095	0.209	0.567
Traditional	Post-test	0.189	0.063	0.094	0.342
Listen and Speak	Delayed test	0.317	0.090	0.170	0.512
Listen Only	Delayed test	0.246	0.076	0.128	0.420
Traditional	Delayed test	0.151	0.053	0.074	0.286
Rhotic approximants					

Table D.1 Predicted probabilities of target-like rating

Instruction	Test	prob	SE	asymp.LCL	asymp.UCL
Listen and Speak	Post-test	0.083	0.031	0.040	0.168
Listen Only	Post-test	0.052	0.021	0.023	0.112
Traditional	Post-test	0.007	0.004	0.002	0.022
Listen and Speak	Delayed test	0.072	0.028	0.033	0.151
Listen Only	Delayed test	0.037	0.016	0.016	0.082
Traditional	Delayed test	0.010	0.005	0.004	0.029

Post Hoc Results

Table D.2 Post hoc within-test pairwise comparisons for target-like rating

Test	contrast	odds.ratio	SE	z.ratio	p.value
Affricates					
Post-test	Listen and Speak / Listen Only	2.259	0.600	3.065	**0.0265**
	Listen and Speak / Traditional	4.953	1.264	6.269	**<.0001**
	Listen Only / Traditional	2.193	0.531	3.244	**0.0150**
Delayed test	Listen and Speak / Listen Only	1.949	0.563	2.312	0.1892
	Listen and Speak / Traditional	3.539	1.004	4.454	**0.0001**
	Listen Only / Traditional	1.816	0.457	2.369	0.1673
Clusters					
Post-test	Listen and Speak / Listen Only	1.531	0.418	1.558	0.6263
	Listen and Speak / Traditional	2.374	0.636	3.229	**0.0157**
	Listen Only / Traditional	1.551	0.436	1.561	0.6245
Delayed test	Listen and Speak / Listen Only	3.685	1.185	4.055	**0.0007**
	Listen and Speak / Traditional	2.168	0.654	2.567	0.1055
	Listen Only / Traditional	0.588	0.184	-1.701	0.5312
Dental fricatives					
Post-test	Listen and Speak / Listen Only	1.321	0.374	0.981	0.9240
	Listen and Speak / Traditional	2.117	0.594	2.672	0.0809
	Listen Only / Traditional	1.603	0.472	1.603	0.5967
Delayed test	Listen and Speak / Listen Only	1.786	0.581	1.781	0.4779
	Listen and Speak / Traditional	1.438	0.453	1.153	0.8590
	Listen Only / Traditional	0.805	0.250	-0.698	0.9822
Diphthongs					
Post-test	Listen and Speak / Listen Only	1.291	0.374	0.882	0.9508
	Listen and Speak / Traditional	11.830	4.006	7.297	**<.0001**
	Listen Only / Traditional	9.163	3.157	6.430	**<.0001**
Delayed test	Listen and Speak / Listen Only	1.002	0.311	0.007	1.0000
	Listen and Speak / Traditional	6.572	2.271	5.449	**<.0001**
	Listen Only / Traditional	6.558	2.156	5.721	**<.0001**
Plosives					
Post-test	Listen and Speak / Listen Only	1.308	0.339	1.038	0.9053
	Listen and Speak / Traditional	3.307	0.826	4.790	**<.0001**
	Listen Only / Traditional	2.528	0.635	3.693	**0.0030**
Delayed test	Listen and Speak / Listen Only	1.425	0.398	1.268	0.8024
	Listen and Speak / Traditional	2.599	0.726	3.419	**0.0083**
	Listen Only / Traditional	1.824	0.478	2.294	0.1965
Rhotic approximants					
Post-	Listen and Speak / Listen Only	1.668	0.495	1.724	0.5160

Table D.2 Post hoc within-test pairwise comparisons for target-like rating

Test	contrast	odds.ratio	SE	z.ratio	p.value
test	Listen and Speak / Traditional	12.752	6.272	5.176	**<.0001**
	Listen Only / Traditional	7.647	3.886	4.003	**0.0009**
Delayed test	Listen and Speak / Listen Only	2.036	0.684	2.115	0.2793
	Listen and Speak / Traditional	7.579	3.426	4.480	**0.0001**
	Listen Only / Traditional	3.723	1.725	2.838	0.0517

Appendix E: Post Hoc Tests for Match Rating

Predicted Probabilities

Table E.1 Predicted probabilities of match rating

Instruction	Test	prob	SE	asymp.LCL	asymp.UCL
Affricates					
Listen and Speak	Post-test	0.953	0.0182	0.902	0.978
Listen Only	Post-test	0.855	0.0435	0.748	0.922
Traditional	Post-test	0.796	0.0540	0.670	0.882
Listen and Speak	Delayed test	0.944	0.0239	0.874	0.976
Listen Only	Delayed test	0.816	0.0521	0.692	0.897
Traditional	Delayed test	0.780	0.0581	0.646	0.873
Clusters					
Listen and Speak	Post-test	0.609	0.0714	0.464	0.737
Listen Only	Post-test	0.521	0.0769	0.373	0.666
Traditional	Post-test	0.437	0.0740	0.301	0.583
Listen and Speak	Delayed test	0.658	0.0736	0.504	0.785
Listen Only	Delayed test	0.533	0.0766	0.384	0.676
Traditional	Delayed test	0.480	0.0769	0.335	0.628
Dental fricatives					
Listen and Speak	Post-test	0.583	0.0781	0.427	0.724
Listen Only	Post-test	0.540	0.0811	0.382	0.690
Traditional	Post-test	0.326	0.0708	0.205	0.476
Listen and Speak	Delayed test	0.481	0.0845	0.323	0.643
Listen Only	Delayed test	0.475	0.0810	0.324	0.631
Traditional	Delayed test	0.292	0.0677	0.178	0.439
Diphthongs					
Listen and Speak	Post-test	0.670	0.0731	0.515	0.795
Listen Only	Post-test	0.695	0.0716	0.540	0.815
Traditional	Post-test	0.297	0.0683	0.182	0.445
Listen and Speak	Delayed test	0.753	0.0673	0.600	0.861
Listen Only	Delayed test	0.685	0.0730	0.528	0.808
Traditional	Delayed test	0.298	0.0692	0.182	0.448
Plosives					
Listen and Speak	Post-test	0.702	0.0642	0.563	0.811
Listen Only	Post-test	0.758	0.0583	0.627	0.854
Traditional	Post-test	0.567	0.0763	0.416	0.706
Listen and Speak	Delayed test	0.618	0.0762	0.462	0.753
Listen Only	Delayed test	0.658	0.0705	0.510	0.780
Traditional	Delayed test	0.539	0.0776	0.388	0.683
Rhotic approximants					

Table E.1 Predicted probabilities of match rating

Instruction	Test	prob	SE	asymp.LCL	asymp.UCL
Listen and Speak	Post-test	0.326	0.0650	0.213	0.464
Listen Only	Post-test	0.196	0.0488	0.118	0.309
Traditional	Post-test	0.169	0.0431	0.100	0.271
Listen and Speak	Delayed test	0.379	0.0726	0.250	0.528
Listen Only	Delayed test	0.155	0.0408	0.091	0.253
Traditional	Delayed test	0.251	0.0570	0.156	0.378

Post Hoc Pairwise Comparisons

Table E.2 Post hoc within-test pairwise comparisons for Match rating

Test	contrast	odds.ratio	SE	z.ratio	p.value
Affricates					
Post-test	Listen and Speak / Listen Only	3.452	1.2949	3.302	**0.0123**
	Listen and Speak / Traditional	5.233	1.8579	4.662	**<.0001**
	Listen Only / Traditional	1.516	0.4336	1.455	0.6929
Delayed test	Listen and Speak / Listen Only	3.786	1.5894	3.171	**0.0190**
	Listen and Speak / Traditional	4.726	1.9400	3.783	**0.0021**
	Listen Only / Traditional	1.248	0.3672	0.754	0.9750
Clusters					
Post-test	Listen and Speak / Listen Only	1.429	0.3410	1.496	0.6669
	Listen and Speak / Traditional	2.002	0.4512	3.081	**0.0252**
	Listen Only / Traditional	1.401	0.3204	1.475	0.6803
Delayed test	Listen and Speak / Listen Only	1.688	0.4600	1.922	0.3883
	Listen and Speak / Traditional	2.091	0.5623	2.744	0.0668
	Listen Only / Traditional	1.239	0.3015	0.880	0.9515
Dental fricatives					
Post-test	Listen and Speak / Listen Only	1.194	0.2853	0.740	0.9769
	Listen and Speak / Traditional	2.893	0.6613	4.646	**<.0001**
	Listen Only / Traditional	2.423	0.5626	3.813	**0.0019**
Delayed test	Listen and Speak / Listen Only	1.025	0.2699	0.094	1.0000
	Listen and Speak / Traditional	2.249	0.5888	3.095	**0.0241**
	Listen Only / Traditional	2.194	0.5310	3.246	**0.0149**
Diphthongs					
Post-test	Listen and Speak / Listen Only	0.892	0.2382	-0.429	0.9982
	Listen and Speak / Traditional	4.802	1.1831	6.368	**<.0001**
	Listen Only / Traditional	5.384	1.3713	6.610	**<.0001**
Delayed test	Listen and Speak / Listen Only	1.405	0.4329	1.104	0.8798
	Listen and Speak / Traditional	7.199	2.1401	6.641	**<.0001**
	Listen Only / Traditional	5.123	1.3584	6.161	**<.0001**
Plosives					
Post-test	Listen and Speak / Listen Only	0.749	0.1886	-1.148	0.8612
	Listen and Speak / Traditional	1.796	0.4206	2.500	0.1239
	Listen Only / Traditional	2.398	0.5787	3.624	**0.0039**
Delayed test	Listen and Speak / Listen Only	0.841	0.2251	-0.647	0.9873
	Listen and Speak / Traditional	1.383	0.3610	1.243	0.8154
	Listen Only / Traditional	1.645	0.4060	2.017	0.3325
Rhotic approximants					

Table E.2 Post hoc within-test pairwise comparisons for Match rating

Test	contrast	odds.ratio	SE	z.ratio	p.value
Post-test	Listen and Speak / Listen Only	1.983	0.4597	2.951	**0.0373**
	Listen and Speak / Traditional	2.382	0.5300	3.901	**0.0013**
	Listen Only / Traditional	1.202	0.2842	0.776	0.9716
Delayed test	Listen and Speak / Listen Only	3.316	0.8456	4.702	**<.0001**
	Listen and Speak / Traditional	1.824	0.4386	2.499	0.1243
	Listen Only / Traditional	0.550	0.1324	-2.483	0.1290

Appendix F: Post Hoc Tests for VOT

Listen and Speak

Table F.1 Results of post hoc pairwise comparisons for VOT durations by Language and Voice for Listen and Speak

Voice	contrast	estimate	SE	df	t.ratio	p.value
Post-test						
Voiceless	L2 - L1	33.9	11.6	13.6	2.919	0.0961
	L2 - TL	-35.8	24.5	46.8	-1.466	0.6870
Voiced	L2 - L1	18.6	10.2	16.5	1.813	0.4843
	L2 - TL	-13.7	29.3	43.7	-0.469	0.9970
Delayed Post-test						
Voiceless	L2 - L1	20.98	14.5	14.2	1.451	0.6980
	L2 - TL	-49.95	22.3	30.9	-2.243	0.2477
Voiced	L2 - L1	-4.03	12.7	17.6	-0.318	0.9995
	L2 - TL	-41.63	30.0	26.7	-1.386	0.7347

Listen Only

Table F.2 Results of post hoc pairwise comparisons for VOT durations by Language and Voice for Listen Only

Voice	contrast	estimate	SE	df	t.ratio	p.value
Post-test						
voiceless	L2 - L1	21.0	9.79	13.4	2.146	0.3220
	L2 - TL	-53.6	23.49	38.2	-2.280	0.2271
voiced	L2 - L1	18.3	8.62	17.0	2.121	0.3228
	L2 - TL	-21.2	27.61	35.9	-0.769	0.9709
Delayed Post-test						
voiceless	L2 - L1	24.5	13.4	14.2	1.830	0.4785
	L2 - TL	-57.3	30.0	29.2	-1.912	0.4155
voiced	L2 - L1	18.9	11.6	17.4	1.628	0.5922
	L2 - TL	-38.5	28.4	36.6	-1.358	0.7512

Traditional

Table F.3 Results of post hoc pairwise comparisons for VOT durations by Language and Voice for Traditional

Voice	contrast	estimate	SE	df	t.ratio	p.value
		Post-test				
voiceless	L2 - L1	14.9	14.0	18.32	1.062	0.8899
	L2 - TL	31.7	23.8	7.64	1.336	0.7604
voiced	L2 - L1	26.3	10.8	19.61	2.440	0.1907
	L2 - TL	61.4	30.2	9.06	2.032	0.3945
		Delayed Post-test				
voiceless	L2 - L1	-0.997	13.7	14.7	-0.073	1.0000
	L2 - TL	17.087	28.5	13.9	0.599	0.9894
voiced	L2 - L1	32.955	11.6	17.6	2.836	0.0976
	L2 - TL	59.090	46.0	17.8	1.285	0.7889

By-condition

Table F.4 Results of post hoc pairwise comparisons for VOT durations by Instruction and Voice for L2

Voice	contrast	estimate	SE	df	t.ratio	p.value
voiceless	Traditional - Listen and Speak	-28.639	9.07	61.7	-3.157	0.0922
	Traditional - Listen Only	-15.550	8.95	55.9	-1.738	0.8425
	Listen and Speak - Listen Only	13.089	8.97	55.0	1.459	0.9451
voiced	Traditional - Listen and Speak	7.341	8.30	64.9	0.884	0.9991
	Traditional - Listen Only	9.308	8.16	59.5	1.141	0.9913
	Listen and Speak - Listen Only	1.967	8.12	57.7	0.242	1.0000

Appendix G: Mean VOT by plosive

Listen and Speak

Post-test

```
   plosive Language ‘mean(fitted_vot)‘
 1 b       L1                      -12.0
 2 b       L2                        5.02
 3 b       TL                       15.5
 4 d       L1                      -15.8
 5 d       L2                       -6.75
 6 d       TL                        8.86
 7 p       L2                       90.5
 8 p       TL                      109.
 9 t       L1                       54.2
10 t       L2                       88.1
11 t       TL                      119.
```

Listen and Speak

Delayed post-test

```
   plosive Language ‘mean(fitted_vot)‘

 1 b       L1                    -18.4
 2 b       L2                    -17.9
 3 b       TL                     10.4
 4 d       L1                    -16.6
 5 d       L2                    -21.7
 6 d       TL                     13.9
 7 p       L2                     80.9
 8 p       TL                    115.
```

```
 9 t       L1                    53.2
10 t       L2                    70.2
11 t       TL                   114.
```

Listen Only

Post-test

```
  plosive Language ‘mean(fitted_vot)‘

 1 b       L1                   -25.9
 2 b       L2                    -4.75
 3 b       TL                     8.09
 4 d       L1                   -27.9
 5 d       L2                    -4.83
 6 d       TL                    16.2
 7 p       L2                    73.8
 8 p       TL                   116.
 9 t       L1                    50.1
10 t       L2                    65.5
11 t       TL                   113.
```

Listen Only

Delayed post-test

```
  plosive Language ‘mean(fitted_vot)‘

 1 b       L1                    -32.3
 2 b       L2                    -17.5
 3 b       TL                     12.4
 4 d       L1                    -49.3
 5 d       L2                    -23.3
 6 d       TL                     11.9
 7 p       L2                     67.2
 8 p       TL                    114.
 9 t       L1                     40.0
10 t       L2                     65.5
11 t       TL                    115.
```

Traditional

Post-test

```
   plosive Language ‘mean(fitted_vot)‘

 1 b       L1                   -36.7
 2 b       L2                    -7.51
 3 b       TL                   -59.2
 4 d       L1                   -39.9
 5 d       L2                   -14.3
 6 d       TL                   -68.7
 7 p       L2                    48.4
 8 p       TL                    26.8
 9 t       L1                    37.4
10 t       L2                    58.2
11 t       TL                    35.6
```

Traditional

Delayed post-test

```
   plosive Language ‘mean(fitted_vot)‘

 1 b       L1                    -53.0
 2 b       L2                    -27.2
 3 b       TL                    -64.1
 4 d       L1                    -51.8
 5 d       L2                    -22.6
 6 d       TL                    -63.7
 7 p       L2                     18.4
 8 p       TL                     20.2
 9 t       L1                     41.9
10 t       L2                     48.2
11 t       TL                     40.7
```

Appendix H: Post Hoc Tests for Vowel-onset F0

Table H.1 Results of post hoc within-Gender pairwise comparisons for vowel-onset f0 values by Language and Voice for Listen and Speak

			Post-test				
Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Male	voiceless	L2 - L1	2.191	6.50	35.2	0.337	0.9994
		L2 - TL	128.536	24.51	27.1	5.245	**0.0002**
	voiced	L2 - L1	1.703	5.55	35.6	0.307	0.9996
		L2 - TL	141.496	27.93	30.4	5.066	**0.0003**
	Language	contrast	estimate	SE	df	t.ratio	p.value
	L2	voiceless - voiced	11.208	6.79	39.2	1.651	0.5712
Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Female	voiceless	L2 - L1	-12.053	6.11	26.9	-1.971	0.3839
		L2 - TL	22.562	24.23	26.9	0.931	0.9348
	voiced	L2 - L1	-0.176	5.27	29.2	-0.033	1.0000
		L2 - TL	26.419	27.75	30.7	0.952	0.9293
	Language	contrast	estimate	SE	df	t.ratio	p.value
	L2	voiceless - voiced	2.825	6.28	36.0	0.450	0.9975
			Delayed Post-test				
Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Male	voiceless	L2 - L1	-8.30	5.87	93.0	-1.413	0.7193
		L2 - TL	128.23	26.57	15.4	4.826	**0.0023**
	voiced	L2 - L1	10.79	5.14	97.0	2.097	0.2976
		L2 - TL	147.55	26.54	20.0	5.559	**0.0002**
	Language	contrast	estimate	SE	df	t.ratio	p.value
	L2	voiceless - voiced	2.37	5.77	41.2	0.411	0.9984
Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Female	voiceless	L2 - L1	5.79	5.38	66.2	1.076	0.8890
		L2 - TL	16.77	26.33	15.5	0.637	0.9862
	voiced	L2 - L1	0.45	4.87	81.0	0.092	1.0000
		L2 - TL	16.03	26.33	20.2	0.609	0.9891
	Language	contrast	estimate	SE	df	t.ratio	p.value
	L2	voiceless - voiced	10.97	6.11	38.8	1.795	0.4808

Table H.2 Results of post hoc within-Gender pairwise comparisons for vowel-onset f0 values by Language and Voice for Listen Only

Post-test							
Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Male	voiceless	L2 - L1	3.95	5.07	26.7	0.779	0.9688
		L2 - TL	130.77	31.35	18.5	4.172	**0.0061**
	voiced	L2 - L1	9.03	4.43	29.6	2.039	0.3455
		L2 - TL	153.29	30.45	20.8	5.034	**0.0007**
	Language	contrast	estimate	SE	df	t.ratio	p.value
	L2	voiceless - voiced	2.37	5.77	41.2	0.411	0.9984
Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Female	voiceless	L2 - L1	6.20	5.32	32.2	1.166	0.8494
		L2 - TL	13.84	31.66	18.4	0.437	0.9976
	voiced	L2 - L1	3.94	4.57	33.6	0.863	0.9526
		L2 - TL	8.82	30.76	20.7	0.287	0.9997
	Language	contrast	estimate	SE	df	t.ratio	p.value
	L2	voiceless - voiced	10.97	6.11	38.8	1.795	0.4808
Delayed Post-test							
Male	voiceless	L2 - L1	0.0111	6.61	28.9	0.002	1.0000
		L2 - TL	131.3805	34.86	18.2	3.769	**0.0147**
	voiced	L2 - L1	-4.8356	5.85	33.3	-0.827	0.9604
		L2 - TL	142.2822	32.69	21.6	4.352	**0.0031**
	Language	contrast	estimate	SE	df	t.ratio	p.value
		voiceless - voiced	14.6599	6.34	42.0	2.313	0.2119
		voiceless - voiced	9.8132	6.75	23.3	1.454	0.6952
		voiceless - voiced	25.5616	20.35	152.8	1.256	0.8082
Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Female	voiceless	L2 - L1	-2.5833	6.70	30.5	-0.386	0.9988
		L2 - TL	14.3512	35.05	18.2	0.409	0.9983
	voiced	L2 - L1	2.4253	5.96	35.9	0.407	0.9984
		L2 - TL	11.4383	32.87	21.6	0.348	0.9992
	Language	contrast	estimate	SE	df	t.ratio	p.value
	L2	voiceless - voiced	8.2006	6.44	41.7	1.273	0.7979

Table H.3 Results of post hoc within-Gender pairwise comparisons for vowel-onset f0 values by Language and Voice for Traditional condition during post-test

Post-test							
Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Male	voiceless	L2 - L1	15.11	8.86	322.0	1.704	0.5301
	voiced	L2 - L1	5.82	7.09	321.0	0.820	0.9637
	Language	contrast	estimate	SE	df	t.ratio	p.value
	L2	voiceless - voiced	5.63	6.26	131.1	0.899	0.9460
Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Female	voiceless	L2 - L1	23.40	12.69	322.5	1.844	0.4390
		L2 - TL	49.50	36.01	13.9	1.374	0.7408
	voiced	L2 - L1	9.66	8.60	322.4	1.124	0.8713
		L2 - TL	63.31	31.85	15.1	1.988	0.3925
	Language	contrast	estimate	SE	df	t.ratio	p.value
	L2	voiceless - voiced	-3.22	7.60	120.5	-0.423	0.9982
Delayed Post-test							
Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Male	voiceless	L2 - L1	-17.599	6.63	68.5	-2.654	0.0985
	voiced	L2 - L1	3.460	5.84	72.3	0.592	0.9912
	Language	contrast	estimate	SE	df	t.ratio	p.value
	L2	voiceless - voiced	-8.953	6.67	134.3	-1.343	0.7605
Gender	Voice	contrast	estimate	SE	df	t.ratio	p.value
Female	voiceless	L2 - L1	3.870	6.30	56.9	0.614	0.9896
		L2 - TL	47.779	37.77	15.1	1.265	0.7987
	voiced	L2 - L1	4.834	5.46	57.9	0.885	0.9487
		L2 - TL	53.645	37.01	16.1	1.450	0.6984
	Language	contrast	estimate	SE	df	t.ratio	p.value
	L2	voiceless - voiced	4.735	6.31	116.6	0.751	0.9750

Table H.4 Results of post hoc pairwise comparisons for vowel-onset f0 by Instruction and Voice for L2

Voice	contrast	estimate	SE	df	t.ratio	p.value
voiceless	Listen and Speak - Listen Only	3.8788	8.01	39.0	0.484	1.0000
	Listen and Speak - Traditional	-11.7500	7.88	41.2	-1.492	0.9345
	Listen Only - Traditional	-15.6288	8.02	42.5	-1.949	0.7228
voiced	Listen and Speak - Listen Only	5.0625	7.55	44.3	0.670	0.9999
	Listen and Speak - Traditional	-14.9522	7.41	47.3	-2.018	0.6791
	Listen Only - Traditional	-20.0147	7.50	45.4	-2.668	0.2747

Appendix I: Post Hoc Tests for Spectral Tilt

Listen and Speak

Table I.1 Results of post hoc pairwise comparisons for spectral tilt by Language and Place for Listen and Speak in both tests

Vowel	Place	contrast	estimate	SE	df	t.ratio	p.value
			Post-test				
Back-High	bilabial	L2 - L1	2.6004	2.89	10.30	0.899	0.9381
		L2 - TL	1.5617	4.68	141.62	0.334	0.9994
	coronal	L2 - L1	2.5034	2.38	10.15	1.050	0.8901
		L2 - TL	-6.8029	5.34	216.83	-1.273	0.7992
Back-Low	bilabial	L2 - TL	14.2559	7.18	403.82	1.985	0.3528
	coronal	L2 - TL	-7.7669	7.13	392.91	-1.089	0.8858
Front-High	bilabial	L2 - L1	-0.0176	2.90	8.45	-0.006	1.0000
		L2 - TL	3.7170	5.33	215.55	0.697	0.9821
	coronal	L2 - L1	-1.5026	2.21	10.20	-0.681	0.9802
		L2 - TL	-2.4154	5.34	216.61	-0.452	0.9976
Front-Low	bilabial	L2 - L1	-4.2633	3.16	12.20	-1.349	0.7543
		L2 - TL	-7.3284	7.22	405.61	-1.015	0.9129
	coronal	L2 - L1	-14.7881	5.49	234.21	-2.691	0.0808
		L2 - TL	-11.5991	5.63	166.00	-2.061	0.3130
			Delayed Post-test				
Back-High	bilabial	L2 - L1	1.2264	2.75	9.45	0.445	0.9970
		L2 - TL	1.5320	4.21	92.91	0.364	0.9991
	coronal	L2 - L1	1.3253	2.36	11.80	0.561	0.9919
		L2 - TL	-5.9797	4.88	148.44	-1.226	0.8237
Back-Low	bilabial	L2 - TL	17.0784	6.55	269.27	2.606	0.0993
	coronal	L2 - TL	-0.3427	6.50	260.34	-0.053	1.0000
Front-High	bilabial	L2 - L1	-1.0537	2.87	8.79	-0.367	0.9988
		L2 - TL	4.4020	4.86	146.91	0.905	0.9446
	coronal	L2 - L1	0.0517	2.23	10.90	0.023	1.0000
		L2 - TL	-1.3230	4.85	145.81	-0.273	0.9998
Front-Low	bilabial	L2 - L1	3.4449	4.01	34.51	0.860	0.9535
		L2 - TL	-5.3816	7.12	299.38	-0.756	0.9745
	coronal	L2 - L1	3.7010	2.62	18.04	1.412	0.7195
		L2 - TL	-9.9710	4.99	157.82	-1.997	0.3486

Listen Only

Table I.2 Results of post hoc pairwise comparisons for spectral tilt by Language and Voice for Listen Only in both tests

Vowel	Place	contrast	estimate	SE	df	t.ratio	p.value
			Post-test				
Back-High	bilabial	L2 - L1	1.4903	1.93	8.99	0.774	0.9654
		L2 - TL	-0.5515	4.43	147.18	-0.125	1.0000
	coronal	L2 - L1	0.9610	1.60	9.90	0.601	0.9885
		L2 - TL	-7.4637	5.12	226.60	-1.456	0.6922
Back-Low	bilabial	L2 - TL	15.0098	6.94	391.66	2.163	0.2576
	coronal	L2 - TL	-5.7978	6.91	391.90	-0.839	0.9599
Front-High	bilabial	L2 - L1	0.8628	1.97	8.23	0.438	0.9971
		L2 - TL	3.6155	5.14	228.12	0.703	0.9814
	coronal	L2 - L1	-2.0522	1.61	11.33	-1.273	0.7931
		L2 - TL	-4.7417	5.14	228.09	-0.923	0.9403
Front-Low	bilabial	L2 - L1	0.2207	2.73	26.30	0.081	1.0000
		L2 - TL	-5.8765	7.14	403.10	-0.823	0.9631
	coronal	L2 - L1	-1.1103	1.91	15.21	-0.580	0.9909
		L2 - TL	-15.0617	5.22	237.22	-2.888	**0.0481**
			Delayed Post-test				
Back-High	bilabial	L2 - L1	2.545	1.40	8.03	1.822	0.5027
		L2 - TL	3.432	4.25	119.15	0.808	0.9656
	coronal	L2 - L1	1.730	1.30	12.38	1.328	0.7654
		L2 - TL	-4.210	4.95	193.09	-0.850	0.9576
Back-Low	bilabial	L2 - TL	17.669	6.65	351.77	2.655	0.0871
	coronal	L2 - TL	-4.894	6.66	352.72	-0.735	0.9774
Front-High	bilabial	L2 - L1	-0.750	1.61	11.20	-0.466	0.9965
		L2 - TL	5.537	4.99	196.75	1.110	0.8769
	coronal	L2 - L1	-2.440	1.24	11.81	-1.974	0.4085
		L2 - TL	0.281	4.94	191.79	0.057	1.0000
Front-Low	bilabial	L2 - L1	1.061	2.08	33.80	0.510	0.9954
		L2 - TL	-4.315	6.77	360.50	-0.637	0.9881
	coronal	L2 - L1	1.760	1.50	16.05	1.170	0.8444
		L2 - TL	-10.293	5.01	199.45	-2.054	0.3161

Traditional

Table I.3 Results of post hoc pairwise comparisons for spectral tilt by Language and Place for Traditional in both tests

Vowel	Place	contrast	estimate	SE	df	t.ratio	p.value
			Post-test				
Back-High	bilabial	L2 - L1	-0.51	2.74	14.4	-0.186	1.0000
		L2 - TL	4.19	4.23	27.4	0.989	0.9175
	coronal	L2 - L1	2.07	2.31	15.5	0.899	0.9409
		L2 - TL	4.54	4.11	24.2	1.104	0.8748
Front-High	bilabial	L2 - L1	1.76	3.25	30.4	0.540	0.9940
		L2 - TL	9.76	4.59	37.3	2.128	0.2959
	coronal	L2 - L1	4.51	2.49	15.3	1.811	0.4881
		L2 - TL	9.10	4.39	31.4	2.072	0.3276
Front-Low	bilabial	L2 - L1	-1.41	3.04	15.1	-0.464	0.9968
		L2 - TL	6.04	5.02	52.1	1.203	0.8332
	coronal	L2 - L1	3.40	2.48	12.6	1.367	0.7445
		L2 - TL	13.21	5.48	71.4	2.413	0.1657
			Delayed Post-test				
Back-High	bilabial	L2 - L1	2.073	1.65	10.49	1.259	0.7997
		L2 - TL	5.704	3.16	93.16	1.806	0.4669
	coronal	L2 - L1	0.808	1.39	12.79	0.581	0.9906
		L2 - TL	4.811	2.90	68.32	1.658	0.5635
Back-Low	coronal	L2 - L1	1.323	6.45	452.19	0.205	0.9999
Front-High	bilabial	L2 - L1	-0.944	1.60	7.76	-0.590	0.9889
		L2 - TL	8.036	3.60	145.84	2.232	0.2297
	coronal	L2 - L1	1.786	1.41	13.73	1.271	0.7955
		L2 - TL	9.603	3.36	115.01	2.860	0.0553
Front-Low	bilabial	L2 - L1	2.054	1.92	15.74	1.067	0.8872
		L2 - TL	9.101	4.16	222.59	2.189	0.2468
	coronal	L2 - L1	4.660	2.04	40.57	2.289	0.2219
		L2 - TL	15.606	4.97	335.18	3.142	**0.0223**

By-condition

Table I.4 Results of post hoc within-Place pairwise comparisons for spectral tilt by Language and Instruction

Place	Language	contrast	estimate	SE	df	t.ratio	p.value
bilabial	L1	Listen Only - Listen and Speak	0.0820	0.898	41.5	0.091	1.0000
		Listen Only - Traditional	1.5679	0.924	45.7	1.697	0.5405
		Listen and Speak - Traditional	1.4859	0.942	51.4	1.577	0.6172
	L2	Listen Only - Listen and Speak	0.7799	0.849	51.5	0.918	0.9401
		Listen Only - Traditional	1.3161	0.829	48.5	1.587	0.6110
		Listen and Speak - Traditional	0.5362	0.840	53.3	0.638	0.9875
coronal	L1	Listen Only - Listen and Speak	0.7699	0.917	51.2	0.839	0.9586
		Listen Only - Traditional	3.2307	0.923	52.7	3.501	0.0116
		Listen and Speak - Traditional	2.4608	0.932	58.5	2.641	0.1035
	L2	Listen Only - Listen and Speak	-0.3287	1.144	47.4	-0.287	0.9997
		Listen Only - Traditional	0.3167	1.144	47.2	0.277	0.9998
		Listen and Speak - Traditional	0.6454	1.136	50.6	0.568	0.9927

References

Ababneh, I (2018), ‘English pronunciation errors made by Saudi students’, *European Scientific Journal, ESJ*, vol. 14, no. 2.

Abdu, H (1988), ‘Italian loanwords in colloquial Libyan Arabic as spoken in the Tripoli region’. PhD thesis. Arizona: The University of Arizona.

Abramson, AS and Lisker, L (1970), ‘Discriminability along the voicing continuum: Cross-language tests’. In: *Proceedings of the sixth international congress of phonetic sciences*. Vol. 196. No. 7. Academia Prague, pp. 569–573.

Abramson, AS and Whalen, DH (2017), ‘Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions’, *Journal of Phonetics*, vol. 63, pp. 75–86. DOI: 10.1016/j.wocn.2017.05.002.

Abu-Mansour, M (1992), ‘Closed syllable shortening and morphological levels’. In: *Perspectives on Arabic Linguistics IV: Papers from the Fourth Annual Symposium on Arabic Linguistics*. Ed. by E Broselow, M Eid, and JJ McCarthy. Amsterdam: John Benjamins, pp. 201–213.

Abumdas, AA (1985), ‘Libyan Arabic phonology’. PhD thesis.

Acha, Joana (2009), ‘The effectiveness of multimedia programmes in children’s vocabulary learning’, *British Journal of Educational Technology*, vol. 40, no. 1, pp. 23–31. DOI: 10.1111/j.1467-8535.2007.00800.x.

Ahmed, A. (2008), ‘Production and perception of Libyan Arabic vowels’. PhD thesis. Newcastle University.

Ahmed, M (1984), ‘An experimental investigation of emphasis in Sudanese colloquial Arabic’. PhD thesis. Reading University.

Akahane-Yamada, R, Tohkura, Y, Bradlow, AR, and Pisoni, DB (1996), *Does training in speech perception modify speech production?* DOI: 10.1109/ICSLP.1996.607434.

Al Yaqoobi, ZM, Ali, F, and Sulan, N (2016), 'An analysis of errors caused by Omani EFL learners in pronouncing certain consonant sounds', *International Journal of Language Education and Applied Linguistics*.

Alghamdi, M (1990), 'Analysis, synthesis and perception of voicing in Arabic'. PhD thesis. University of Reading.

Allen, GD (1985), 'How the young French child avoids the pre-voicing problem for word-initial voiced stops', *Journal of Child Language*, vol. 12, no. 1, pp. 37–46.

Aloreibi, A and Carey, MD (2017), 'English language teaching in Libya after Gaddafi'. In: *English Language Education Policy in the Middle East and North Africa*. Ed. by R Kirkpatrik. Kuwait: Springer, pp. 93–114.

Alqattan, S (Feb. 2015), 'Early phonological acquisition by Kuwaiti Arabic children'. PhD thesis. Newcastle Upon Tyne: School of Education Communication and Language Sciences.

Altaha, FM (1995), 'Pronunciation Errors made by Saudi university students learning English: analysis and remedy', *International Journal of Applied Linguistics*, vol. 109, no. 1, pp. 110–123. DOI: 10.1075/itl.109-110.05alt.

Alwan, A, Narayanan, S, and Haker, K (1997), 'Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics', *The Journal of the Acoustical Society of America*, vol. 101, no. 2, pp. 1078–1089. DOI: 10.1121/1.417972.

Amayreh, M (2003), 'Completion of the Consonant Inventory of Arabic', *Journal of Speech, Language, and Hearing Research*, vol. 46, no. 3, pp. 517–529. DOI: 10.1044/1092-4388(2003/042).

Amayreh, M and Dyson, A (1998), 'The Acquisition of Arabic Consonants', *Journal of Speech, Language, and Hearing Research*, vol. 41, no. 3, pp. 642–653. DOI: 10.1044/jslhr.4103.642.

Amayreh, M and Dyson, A (2000), ‘Phonetic inventories of young Arabic-speaking children’, *Clinical Linguistics & Phonetics*, vol. 14, no. 3, pp. 193–215.

Al-Amer, A (2018), ‘Phonological acquisition and development in Arabic-English bilingual children’. PhD thesis. Birkbeck College, University of London.

Ammar, W and Morsi, R (2006), ‘Phonological development and disorders: colloquial Egyptian Arabic’, *Phonological development and disorders in children: A multilingual perspective*, pp. 204–232.

Anderson, J (1987), ‘The Markedness Differential Hypothesis and syllable structure difficulty’. In: *Interlanguage phonology: The acquisition of a second language sound system*. Ed. by G. Ioup and S. Weinberger. Cambridge, MA: Newbury House, pp. 279–291.

Al-Ani, SH (1970), *Arabic phonology; an accoustical and physiological investigation*. The Hague: The Hague, Mouton.

Aquil, R (2013), ‘Cairne Arabic Syllable Structure though Different Phonological Theories’, *Open Journal of Modern Linguistics*, vol. 3, no. 03, pp. 259–267. DOI: 10.4236/ojml.2013.33034.

Arlt, PB and Goodban, MT (1976), ‘A comparative study of articulation acquisition as based on a study of 240 normals, aged three to six’, *Language, Speech, and Hearing Services in Schools*, vol. 7, no. 3, pp. 173–180.

Aslin, RN, Pisoni, DB, Hennessy, BL, and Perey, AJ (1981), ‘Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience’, *Child Development*, vol. 52, no. 4, pp. 1135–1114.

Aurayieth, A (1982), ‘The phonology of the verb in Libyan Arabic’. PhD thesis. University of Washington.

Avery, P and Ehrlich, S (1992), *Teaching American English Pronunciation*. Oxford: Oxford University Press.

Ayyad, HS (2011), ‘Phonological development of typically developing Kuwaiti Arabic–speaking preschoolers’. PhD thesis. University of British Columbia.

Ayyad, HS, Bernhardt, BM, and Stemberger, JP (Sept. 2016), ‘Kuwaiti Arabic: acquisition of singleton consonants’, *International Journal of Language and*

Communication Disorders, vol. 51, no. 5, pp. 531–545. DOI: 10.1111/1460-6984.12229.

Baker, W, Trofimovich, P, Flege, JE, Mack, M, and Halter, R (2008), 'Child—Adult Differences in Second-Language Phonological Learning: The Role of Cross-Language Similarity', *Language and Speech*, vol. 51, no. 4, pp. 317–342. DOI: 10.1177/0023830908099068.

Bankson, N and Bernthal, JE (1998), 'Articulation and phonological disorders', *Analysis and interpretation of assessment data*, pp. 270–298.

Barros, AM (2003), 'Pronunciation difficulties in the consonant system experienced by Arabic speakers when learning English after the age of puberty'. Unpublished. MA thesis. Morgantown: USA: West Virginia University.

Bassetti, B (2006), 'Pinyin orthographic input and CFL learners' pronunciation'. In: *British Chinese Language Teachers' Seminar, Cambridge*, pp. 9–10.

— (2007), 'Effects of hanyu pinyin on pronunciation in learners of Chinese as a foreign language'. In: *The cognition, learning and teaching of Chinese characters*. Ed. by A Guder-Manitius, X Jiang, and Y Wan. Beijing, China: Beijing Language and Culture University Press.

— (2008), 'Orthographic input and second language phonology'. In: *Input Matters in SLA*. Ed. by T Piske and M Young-Scholten. May. Clevedon, UK: Multilingual Matters, pp. 191–206.

— (2017), 'Orthography affects second language speech: Double letters and geminate production in English', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 43, pp. 1835–1842. DOI: 10.1037/xlm0000417.

Bassetti, B and Atkinson, N (2015), 'Effects of orthographic forms on pronunciation in experienced instructed second language learners', *Applied Psycholinguistics*, vol. 36, no. 1, pp. 67–91. DOI: 10.1017/S0142716414000435.

Bassetti, B, Soković-Perović, M, Mairano, P, and Cerni, T (2018), 'Orthography-induced length contrasts in the second language phonological systems of L2 speakers of English: Evidence from minimal pairs', *Language and Speech*, vol. 61, no. 4, pp. 577–597. DOI: 10.1177/0023830918780141.

Bates, D, Mächler, M, Bolker, B, and Walker, S (2015), 'Fitting Linear Mixed-Effects Models Using lme4', *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48. DOI: 10.18637/jss.v067.i01.

Beckman, ME and Edwards, J (2000), 'Lexical frequency effects on young children's imitative productions'. In: *Papers in laboratory phonology V: Acquisition and the Lexicon*. Ed. by M Broe and Janet Pierrehumbert. Cambridge: Cambridge University Press, pp. 208–218.

Bell, A and Gibson, A (2008), 'Stopping and Fronting in New Zealand Pasifika English', *University of Pennsylvania Working Papers in Linguistics*, vol. 14, no. 2, p. 7.

Bernhardt, BM and Stemberger, JP (2017), 'Investigating typical and protracted phonological development across languages'. In: *Cross linguistic encounters in language acquisition: Typical and atypical development*. Ed. by E Babatsouli, D Ingram, and N Müller. Multilingual Matters, pp. 71–108.

Best, CT (1994), 'The emergence of native-language phonological influences in infants: A perceptual assimilation model'. In: *The development of speech perception: The transition from speech sounds to spoken words*. Ed. by JC Goodman and HC Nusbaum. Cambridge, MA: MIT Press, pp. 167–224. DOI: 10.1007/978-94-015-8234-6_24.

— (1995), 'A direct realist view of cross-language speech perception: New directions in research and theory'. In: *Speech perception and linguistic experience: Issues in cross-language speech research*. Ed. by W Strange. Baltimore: York Press, pp. 171–204.

Best, CT, McRoberts, GW, and Goodell, E (2001), 'Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system', *The Journal of the Acoustical Society of America*, vol. 109, no. 2, pp. 775–794. DOI: 10.1121/1.1332378.

Best, CT and Tyler, MD (2007), 'Nonnative and second-language speech perception: Commonalities and complementarities'. In: *Language experience in second language speech learning: In honor of James Emil Flege*. Ed. by Ocke-Schwen Bohn and Murray Munro J. Vol. 17. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Birdsong, D (2007), 'Nativelike pronunciation among late learners of French as a second language1', *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*, vol. 17, pp. 99–116. DOI: https://doiorg/10.1075/lllt.17.12bir.

Blevins, J (2004), *Evolutionary phonology: The emergence of sound patterns.* Cambridge: Cambridge University Press.

Boersma, P and Weeninck, D (2016), *Praat: doing phonetics by computer (version 6.0.092).* URL: http://www.praat.org.

Bohn, O and Flege, JE (1992), 'The Production of New and Similar Vowels by Adult German Learners of English', *Studies in Second Language Acquisition*, vol. 14, no. 02, p. 131. DOI: 10.1017/S0272263100010792.

Bond, ZS and Fokes, J (1991), 'Perception of English voicing by native and nonnative adults', *Studies in Second Language Acquisition*, vol. 13, no. 4, pp. 471–492. DOI: 10.1017/S0272263100010299.

Botagga, M (1991), 'A generative phonetic feature analysis of the vowel development of native Arabic speakers learning English as a second language', *Unpublished PhD dissertation, University of Wales.*

Boyce, SE, Hamilton, SM, and Rivera-Campos, A (2016), 'Acquiring rhoticity across languages: An ultrasound study of differentiating tongue movements', *Clinical Linguistics & Phonetics*, vol. 30, no. 3-5, pp. 174–201. DOI: 10.3109/02699206.2015.1127999.

Bradlow, AR, Akahane-Yamada, R, Pisoni, DB, and Tohkura, Y (1999), 'Training Japanese listeners to identify English/r/and /l/: Long-term retention of learning in perception and production', *Attention, Perception, & Psychophysics*, vol. 61, no. 5, pp. 977–985. DOI: 10.3758/BF03206911.

Bradlow, AR, Pisoni, DB, Akahane-Yamada, R, and Tohkura, Y (1997), 'Training Japanese listeners to identify English /r/and/l: IV. Some effects of perceptual learning on speech production', *The Journal of the Acoustical Society of America*, vol. 101, no. 4, pp. 2299–2310. DOI: 10.1121/1.418276.

Broselow, E and Finer, D (1991), 'Parameter setting in second language phonology and syntax', *Interlanguage Studies Bulletin (Utrecht)*, vol. 7, no. 1, pp. 35–59. DOI: 10.1177/026765839100700102.

Browning, SR (2004), *Analysis of Italian children's English pronunciation*. Unpublished report contributed to the EU FP5 PF STAR Project. URL: http://www.eee.bham.ac.uk/russellm/ItalianEnglishReport%5C%5CItalianEnglish_report_v2.htm.

Al-Buainain, H, Shain, K, Al-Timimy, F, and Khattab, G (2013), 'Baseline Data for Arabic Acquisition with Clinical Applications: Some Phonological Processes in Qatari Children's Speech', *International Journal of Business and Social Research*, vol. 2, no. 6. DOI: 10.18533/ijbsr.v2i6.131.

Bukshaisha, F (1985), 'An experimental phonetic study of some aspects of Qatari Arabic'. PhD thesis.

Bundgaard-Nielsen, RL, Best, CT, and Tyler, MD (2011), 'Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories', *Applied Psycholinguistics*, vol. 32, no. 1, pp. 51–67. DOI: 10.1017/S0142716410000287.

Busby, PA and Plant, GL (1995), 'Development of speech sounds in children', *The Journal of the Acoustical Society of America*, vol. 97, no. 4, pp. 2603–2606.

Canfield, RL and Jusko, TA (2009), 'Lead poisoning'. In: *Diseases and Disorders in Infancy and Early Childhood*. Ed. by JB Jensen and MM Haith. Oxford: Academic Press, pp. 237–250.

Chaker, S and Mettouchi, A (2006), 'Berber'. In: *Encyclopedia of Language & Linguistics*. Ed. by K Brown. 2nd ed. Oxford: Elsevier, pp. 738–744. DOI: 10.1016/B0-08-044854-2/02064-2.

Chirlian, NS and Sharpley, CF (1982), 'Children's articulation development: Some regional differences', *Australian Journal of Human Communication Disorders*, vol. 10, no. 2, pp. 23–30.

Cho, T and Ladefoged, P (1999), 'Variation and universals in VOT: evidence from 18 languages', *Journal of phonetics*, vol. 27, no. 2, pp. 207–229.

Cho, T, Whalen, DH, and Docherty, G (2019), 'Voice onset time and beyond: Exploring laryngeal contrast in 19 languages', *Journal of Phonetics*, vol. 72, pp. 52–65. DOI: 10.1016/j.wocn.2018.11.002.

Chun, DM, Jiang, Y, and Ávila, N (2013), 'Visualization of tone for learning Mandarin Chinese'. In: *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, pp. 77–89.

Colantoni, L and Steele, J (2006), 'Native-like attainment in the L2 acquisition of Spanish stop-liquid clusters'. In: *Selected proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*. Ed. by CA Klee and TL Face. Cascadilla. Somerville, MA, pp. 59–73.

— (2008), 'Integrating articulatory constraints into models of second language phonological acquisition', *Applied Psycholinguistics*, vol. 29, no. 3, pp. 489–534. DOI: 10.1017/S0142716408080223.

Colantoni, L, Steele, J, and Escudero, P (2015), *Second language speech: Theory and practice*. Cambridge: Cambridge University Press.

Costamagna, L (2000), *Insegnare e imparare la fonetica*. Torino, Italy: Paravia.

Couper, G (2011), 'What makes pronunciation teaching work? Testing for the effect of two variables: socially constructed metalanguage and critical listening', *Language Awareness*, vol. 20, no. 3, pp. 159–182. DOI: 10.1080/09658416.2011.570347.

Cruttenden, A (2008), 'Gimson's Pronunciation of English. London: Hodder Education'.

Cruttenden, A (2014), *Gimson's pronunciation of English.* 8th ed. Routlege.

Cucchiarini, C, Dawidowicz, M, Filimban, E, Tammelin-Laine, T, Craats, I van de, and Strik, H (2015), 'The Digital Literacy Instructor: developing Automatic Speech Recognition and selecting learning material for opaque and transparent orthographies'. In: *Adult Literacy, Second Language and Cognition.* Nijmegen: Centre for Language Studies, pp. 251–278.

Cucchiarini, C, Neri, A, and Strik, H (2009), 'Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback', *Speech Communication*, vol. 51, no. 10, pp. 853–863. DOI: 10.1016/j.specom.2009.03.003.

Cucchiarini, C, Strik, H, and Boves, L (2002), 'Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech', *The Journal of the Acoustical Society of America*, vol. 111, no. 6, pp. 2862–2873. DOI: 10.1121/1.1471894.

Cutler, A (2015), 'Representation of second language phonology', *Applied Psycholinguistics*, vol. 36, no. 1, pp. 115–128. DOI: 10.1017/S0142716414000459.

Davis, K (1995), *Phonetic and phonological contrasts in the acquisition of voicing: Voice onset time production in Hindi and English.* Vol. 22. 2, pp. 275–305. DOI: 10.1017/S030500090000979X.

Davis, K and Kuhl, PK (1992), 'Best exemplars of English velar stops: A first report'. In: *Second International Conference on Spoken Language Processing.*

— (1993), 'Acoustic correlates of phonetic prototypes: Velar stops', *The Journal of the Acoustical Society of America*, vol. 93, no. 4, p. 2392.

— (1994), 'Tests of the perceptual magnet effect for American English/k/and/g', *The Journal of the Acoustical Society of America*, vol. 95, no. 5, p. 2976.

DeKeyser, RM (2000), 'The robustness of critical period effects in second language acquisition', *Studies in second language acquisition*, vol. 22, no. 4, pp. 499–533.

Delattre, P and Freeman, DC (1968), 'A dialect study of American r's by x-ray motion picture', *Linguistics*, vol. 6, no. 44, pp. 29–68. DOI: 10.1515/ling.1968.6.44.29.

Demuth, K and Song, JY (2011), ‘Insights From Acquisition and Learning’. In: *The Oxford Handbook of Laboratory Phonology*. Ed. by AC Cohn, C Fougeron, and MK Huffman. New York: Oxford University Press.

Diehl, RL and Lindblom, B (2004), ‘Explaining the structure of feature and phoneme inventories: the role of auditory distinctiveness’. In: *Speech processing in the auditory system*. Ed. by A Greenberg and F Popper. Springer, pp. 101–162.

Dmitrieva, O, Llanos, F, Shultz, AA, and Francis, AL (2015), ‘Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English’, *Journal of Phonetics*, vol. 49, pp. 77–95. DOI: 10.1016/j.wocn.2014.12.005.

Docherty, G (1992), *The timing of voicing in British English obstruents*. 9. Walter de Gruyter.

Docherty, G, Foulkes, P, Tillotson, J, and Watt, D (2006), ‘On the scope of phonological learning: issues arising from socially structured variation’. In: *Laboratory Phonology 8: Varieties of Phonological Competence*. Ed. by L Goldstein, CT Best, and DH Whalen. Berlin: Mouton de Gruyter, pp. 393–422.

Dodd, B (1995), *The differential diagnosis and treatment of children with speech disorder*. Studies in disorders of communication. London: Whurr.

Dodd, B, Holm, A, Crosbie, SC, and Zhu, H (2013), ‘Children’s acquisition of phonology’. In: *Differential diagnosis and treatment of children with speech disorder*. Ed. by B Dodd. 2nd ed. London; Philadelphia, Pa: John Wiley & Sons, Incorporated.

Dodd, B, Holm, A, Hua, Z, and Crosbie, SC (2003), ‘Phonological development: a normative study of British English-speaking children’, *Clinical Linguistics & Phonetics*, vol. 17, no. 8, pp. 617–643.

Dyson, A (1988), ‘Phonetic inventories of 2-and 3-year-old children’, *Journal of Speech and Hearing Disorders*, vol. 53, no. 1, pp. 89–93.

Dyson, A and Amayreh, M (2000), ‘Phonological errors and sound changes in Arabic-speaking children’, *Clinical Linguistics & Phonetics*, vol. 14, no. 2, pp. 79–109. DOI: 10.1080/026992000298850.

— (2002), ‘Phonological errors and sound changes in Arabic-speaking children’, *Clinical Linguistics & Phonetics*, vol. 14, no. 2, pp. 79–109. DOI: 10.1080/026992000298850.

Dyson, A and Paden, EP (1983), ‘Some phonological acquisition strategies used by two-year-olds’, *Journal of Childhool Communication Disorders*, vol. 7, no. 1, pp. 6–18.

Eckman, F (1977), ‘Markedness and the contrastive analysis hypothesis’, *Language Learning*, vol. 27, no. 2, pp. 315–330. DOI: 10.1111/j.1467-1770.1977.tb00124.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-1770.1977.tb00124.x.

— (1991), ‘The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners’, *Studies in second language acquisition*, vol. 13, no. 1, pp. 23–41.

— (Dec. 2004), ‘From Phonemic Differences to Constraint Rankings: Research on Second Language Phonology’, *Studies in Second Language Acquisition*, vol. 26, pp. 513–549. DOI: 10.1017/S027226310404001X.

— (2008), ‘Typological markedness and second language phonology’. In: *Phonology and Second Language Acquisition*. Ed. by JGH Edwards and ML Zampini. Amsterdam: John Benjamins, pp. 95–115.

Edwards, J, Beckman, ME, and Munson, B (2004), ‘The interaction between vocabulary size and phonotactic probability effects on children’s production accuracy and fluency in nonword repetition’, *Journal of speech, language, and hearing research*, vol. 47, pp. 421–436.

Eguchi, S and Hirsh, I (1969), ‘Formant frequency values of vowels produced by preadolescent boys and girls’, *Acta Oto-Laryngol. Suppl*, vol. 257, pp. 2–51.

Ehbara, H (2015), ‘Vowel epenthesis in Libyan Arabic (C)CVCC syllables’. Newcastle University.

Elfitoury, A (May 1976), ‘A descriptive grammar of Libyan Arabic’. PhD thesis. Washington: Georgetown University.

Elgadi, A (1986), ‘Tripolitanean Arabic phonology and morphology: A generative approach’. PhD thesis. Georgetown: Georgetown University.

Elliott, AR (1995), ‘Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language’, *The Modern Language Journal*, vol. 79, no. 3, pp. 356–371. DOI: 10.1111/j.1540-4781.1995.tb01112.x.

Elramli, Y (2012), ‘Assimilation in the phonology of a Libyan Arabic dialect: a constraint-based approach’. PhD thesis. Newcastler upon Tyne.

Embarki, M (2013), ‘Phonetics’. In: *The Oxford Handbook of Arabic Linguistics*. Ed. by Jonathon Owens. Oxford: Oxford University Press, pp. 23–44.

Emhamed, EDH and Krishnan, KSD (2011), ‘Investigating Libyan teachers’ attitude towards integrating technology in teaching English in Sebha secondary schools’, *Academic Research International*, vol. 1, no. 3, p. 182.

Escudero, P and Boersma, P (2004), ‘Bridging the Gap Between L2 Speech Perception Research and Phonological Theory’, *Studies in Second Language Acquisition*, vol. 26, no. 04, pp. 551–585. DOI: 10.1017/S0272263104040021.

Escudero, P, Hayes-Harb, R, and Mitterer, H (2008), ‘Novel second-language words and asymmetric lexical access’, *Journal of Phonetics*, vol. 36, no. 2, pp. 345–360. DOI: 10.1016/j.wocn.2007.11.002.

Escudero, P, Simon, E, and Mulak, KE (2013), ‘Learning words in a new language: Orthography doesn’t always help’, *Bilingualism: Language and Cognition*, vol. 17, no. 2. DOI: 10.1017/S1366728913000436.

Escudero, P, Sisinni, B, and Grimaldi, M (2014), ‘The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels’, *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1577–1584. DOI: 10.1121/1.4864477.

Escudero, P and Wanrooij, K (2010), ‘The effect of L1 orthography on non-native vowel perception’, *Language and Speech*, vol. 53, no. 3, pp. 343–365. DOI: 10.1177/0023830910371447.

Evans, BG and Alshangiti, W (2018), ‘The perception and production of British English vowels and consonants by Arabic learners of English’, *Journal of Phonetics*, vol. 68, pp. 15–31. DOI: 10.1016/j.wocn.2018.01.002.

Fantazi, G (2003), ‘Perception and production of syllable structure and stress by adult Libyan Arabic speaker acquiring English in the UK’. PhD thesis. Durham University.

Ferguson, CA (1959), ‘Diglossia’, *Word*, vol. 15, no. 2, pp. 325–340. DOI: 10.1080/00437956.1959.11659702.

Flege, JE (1987a), ‘A critical period for learning to pronounce foreign languages?’, *Applied linguistics*, vol. 8, no. 2, pp. 162–177.

— (1987b), ‘The instrumental study of L2 speech production: Some methodological considerations’, *Language Learning*, vol. 37, no. 2, pp. 285–296. DOI: 10.1111/j.1467-1770.1987.tb00569.x.

— (1987c), ‘The production and perception of foreign language speech sounds’. In: *Human communication and its disorders: A review*. Ed. by H Winitz. Vol. 1. Norwood, pp. 224–401.

— (1987d), ‘The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification’, *Journal of Phonetics*, vol. 15, no. 1, pp. 47–65.

— (1988), ‘The development of skill in producing word-final English stops: Kinematic parameters’, *The Journal of the Acoustical Society of America*, vol. 84, no. 5, pp. 1639–1652.

— (1995), ‘Second language speech learning: Theory, findings, and problems’. In: *Speech perception and linguistic experience: Issues in cross-language research*. Ed. by W Strange. Baltimore: York Press, pp. 233–277.

Flege, JE, Birdsong, D, Bialystok, E, Mack, M, Sung, H, and Tsukada, K (2006), ‘Degree of foreign accent in English sentences produced by Korean children and adults', *Journal of Phonetics*, vol. 34, no. 2, pp. 153–175. DOI: 10.1016/j.wocn.2005.05.001.

Flege, JE and Fletcher, KL (1992), 'Talker and listener effects on degree of perceived foreign accent', *The Journal of the Acoustical Society of America*, vol. 91, no. 1, pp. 370–389. DOI: 10.1121/1.402780. eprint: 10.1121/1.402780.

Flege, JE and Hillenbrand, J (1984), 'Limits on phonetic accuracy in foreign language speech production', *The Journal of the Acoustical Society of America*, vol. 76, no. 3, pp. 708–721. DOI: 10.1121/1.391257.

Flege, JE, MacKay, I, and Meador, D (1999), 'Native Italian speakers' perception and production of English vowels', *The Journal of the Acoustical Society of America*, vol. 106, no. 5, pp. 2973–2987. DOI: 10.1121/1.428116.

Flege, JE, Munro, MJ, and MacKay, I (1995a), 'Factors affecting strength of perceived foreign accent in a second language', *The Journal of the Acoustical Society of America*, vol. 97, no. 5, pp. 3125–3134. DOI: 10.1121/1.413041.

— (1995b), 'The effects of age of second language learning on the production of English consonants', *Speech Commun*, vol. 16, pp. 1–26.

Flege, JE, Navarra, J, Sebastián-Gallés, N, and Soto-Faraco, S (2003), 'Assessing constraints on second-language segmental production and perception', *Language Comprehension and Production*, vol. 93, no. 3, pp. 1589–1608. DOI: 10.1037/0096-1523.31.5.912.

Flege, JE and Port, R (1981), 'Cross-Language Phonetic Interference: Arabic to English', *Language and Speech*, vol. 24, no. 2, pp. 125–146. DOI: 10.1177/002383098102400202.

Flege, JE and Schmidt, AM (1995), 'Native speakers of Spanish show rate-dependent processing of English stop consonants', *Phonetica*, vol. 52, no. 2, pp. 90–111.

Flege, JE, Yeni-Komshian, GH, and Liu, S (1999), 'Age constraints on second-language acquisition', *Journal of Memory and Language*, vol. 41, no. 1, pp. 78–104.

Foulkes, P, Docherty, G, and Watt, D (1999), ‘Tracking the emergence of sociophonetic variation’. In: *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 1625–1628.

— (2005), ‘Phonological Variation in Child-Directed Speech’, *Language*, vol. 81, pp. 177–206.

Fraser, Helen (2000), *Coordinating Improvements in Pronunciation Teaching for Adult Learners of English as a Second Language*. Armidale, NSW: University of New England.

Garib, A (Dec. 2014), ‘A VOT measurement of the pronunciation of word-initial /p/ by Libyan speakers of English’. MA thesis. Arizona State University.

Gentner, D (1982), ‘Why nouns are learned verbs: Linguistic relativity versus natural partitioning’, *Syntax and semantics*.

Ghazali, S (1977), ‘Back consonants and backing coarticulation in Arabic’. PhD thesis. Austin.

Ghitza, O and Goldstein, JL (1982), ‘Discrimination of formant frequency, intensity, and bandwidth in natural speech’, *The Journal of the Acoustical Society of America*, vol. 71, no. S1, S37–S37.

Giannakopoulou, A, Brown, H, Clayards, M, and Wonnacott, E (2017), ‘High or low? Comparing high and low-variability phonetic training in adult and child second language learners’, *PeerJ Life and Environment*. DOI: 10.7717/peerj.3209.

Giannini, A and Pettorino, M (1982), ‘The emphatic consonants in Arabic’, *Speech Laboratory Report*, vol. 4.

Gierut, JA (1999), ‘Syllable onsets: Clusters and adjuncts in acquisition’, *Journal of Speech, Language, and Hearing Research*, vol. 42, no. 3, pp. 708–726. DOI: 10.1044/jslhr.4203.708.

Gimson, AC (1970), *An introduction to the pronunciation of English*. London: Edward Arnold.

Gnanadesikan, A (2004), ‘Markedness and faithfulness constraints in child phonology’. In: *Fixing priorities. Constraints in phonological acquisition.*

Ed. by R Kager, J Pater, and W Zonneveld. Cambridge: Cambridge University Press, pp. 73–108.

Goldstein, MH, Schwade, JA, and Bornstein, MH (2009), 'The value of vocalizing: Five-month-old infants associate their own noncry vocalizations with responses from caregivers', *Child Development*, vol. 80, no. 3, pp. 636–644. DOI: 10.1111/j.1467-8624.2009.01287.x.

Green, JR, Moore, CA, Higashikawa, M, and Steeve, RW (2000), 'The physiologic development of speech motor control: Lip and jaw coordination', *Journal of Speech, Language, and Hearing Research*, vol. 43, no. 1, pp. 239–255. DOI: 10.1044/jslhr.4301.239.

Greenlee, M (1974), 'Interacting Processes in the Child's Acquisition of Stop-Liquid Clusters Papers and Reports on Child Language Development', no. 7.

Griffini, E (1913), *L'arabo parlato della Libia*. Ulrico Hoepli.

Grunwell, P (1985), *Phonological assessment of child speech*. College Hill Press.

— (1987), *Clinical phonology*. 2nd ed. London: Croom Helm.

Guilloteau, NC (1997), 'Modification of phonetic categories in French as a second language: Experimental studies with conventional and computer-based intervention methods'. PhD thesis. The University of Texas at Austin.

Gusbi, M (1966), *English for Libya*. London: University of London Press LTD.

Gusbi, M and John, R (1974), *Further English for Libya*. Longman.

Habara, HF (Jan. 2017), 'Phonological style-shift in Misrata Libyan Arabic: A socio-historical approach'. Misrata University.

Hadley, W (2016), *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. URL: https://ggplot2.tidyverse.org.

Haelsig, PC and Madison, CL (1986), 'A study of phonological processes exhibited by 3-, 4-, and 5-year-old children', *Language, Speech, and Hearing Services in Schools*, vol. 17, no. 2, pp. 107–114.

Hago, OE and Khan, WA (2015), ‘The pronunciation problems faced by Saudi EFL learners at secondary schools’, *Education and Linguistics Research*, vol. 1, no. 2, pp. 85–99.

Hakuta, K, Bialystok, E, and Wiley, E (2003), ‘Critical Evidence: A Test of the Critical-Period Hypothesis for Second-Language Acquisition’, *Psychological Science*, vol. 14, no. 1, pp. 31–38. DOI: 10.1111/1467-9280.01415.

Halle, M and Clements, GN (1983), *Problem Book in Phonology*. Cambridge: MIT Press.

Hamdi, R, Ghazali, S, and Barkat-Defradas, M (2005), ‘Syllable Structure in Spoken Arabic: a comparative investigation’. In: *Ninth European Conference on Speech Communication and Technology*.

Hancin-Bhatt, B and Bhatt, RM (1997), ‘Optimal L2 syllables: Interactions of transfer and developmental effects’, *Studies in second language acquisition*, vol. 19, no. 3, pp. 331–378.

Hansen, Pernille (2017), ‘What makes a word easy to acquire? The effects of word class, frequency, imageability and phonological neighbourhood density on lexical development’, *First language*, vol. 37, no. 2, pp. 205–225.

Hardcastle, W and Barry, W (1985), ‘Articulatory and perceptual factors in /I/ vocalisations in English’, *Journal of the International Phonetic Association*, vol. 15, no. 2, pp. 3–17.

Hardison, DM (2004), ‘Generalization of computer assisted prosody training: Quantitative and qualitative findings’, *Language Learning & Technology*, vol. 8, no. 1, pp. 34–52.

— (2005), ‘Contextualized computer-based L2 prosody training: Evaluating the effects of discourse context and video input’, *Calico Journal*, pp. 175–190.

Harris, KS (1958), ‘Cues for the discrimination of American English fricatives in spoken syllables’, *Language and speech*, vol. 1, no. 1, pp. 1–7. DOI: 10.1177/002383095800100101.

Hattori, K and Iverson, P (2009), ‘English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy', *The Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 469–479.

Hazan, V and Barrett, S (2000), 'The development of phonemic categorization in children aged 6-12', *Journal of Phonetics*, vol. 28, no. 4, pp. 377–396. DOI: 10.1006/jpho.2000.0121.

Hazan, V, Sennema, A, Iba, M, and Faulkner, A (2005), 'Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English', *Speech Communication*, vol. 47, no. 3, pp. 360–378. DOI: 10.1016/j.specom.2005.04.007.

Heimbach, RL (1994), 'The effect of interaction on the comprehension and acquisition of new lexical items by Kindergarten ESL learners'. PhD thesis. Temple University.

Herd, W, Jongman, A, and Sereno, J (2013), 'Perceptual and production training of intervocalic /d, ɾ, r/ in American English learners of Spanish', *The Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. 4247–4255. DOI: 10.1121/1.4802902.

Heselwood, B (1996), 'Glottal states and emphasis in Baghdadi and Cairene Arabic: synchronic and diachronic aspects', *University of Durham Center for Middle Eastern and Islamic Studies, Occasional Paper*, vol. 53.

Hodson, BW and Paden, EP (1981), 'Phonological processes which characterize unintelligible and intelligible speech in early childhood', *Journal of Speech and Hearing Disorders*, vol. 46, no. 4, pp. 369–373.

Holes, C (2007), 'Kuwaiti Arabic'. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by K Versteegh, M Eid, A Elgibali, M Woidich, and A Zaborski. Netherlands: Brill Academic Publishers, pp. 608–620.

Holm, A (1998), 'Speech development and disorder in bilingual children'. PhD thesis. Newcastle University.

Hombert, JM, Ohala, JJ, and Ewan, WG (1979), 'Phonetic explanations for the development of tones', *Language*, vol. 55, no. 1, pp. 37–58.

House, AS and Fairbanks, G (1953), ‘The influence of consonant environment upon the secondary acoustic characteristics of vowels’, *Journal of Acoustic Society of America*, vol. 25, no. 1, pp. 105–113.

Hurford, JR (1991), ‘The evolution of the critical period for language acquisition’, *Cognition*, vol. 40, no. 3, pp. 159–201.

Imssalem, N (2002), *Teaching and learning in Libya*. Benghazi: University of Garyounis Press.

Ingels, SA (2011), ‘The effects of self-monitoring strategy use on the pronunciation of learners of English’. PhD thesis. University of Illinois at Urbana-Champaign.

Ingram, D (1976), *Phonological disability in children*. London: Edward Arnold.

— (1999), ‘Phonological acquisition’, *The development of language*, pp. 73–97.

Ingram, D, Christensen, L, Veach, S, and Webster, B (1980), ‘The acquisition of word-initial fricatives and affricates in English by children between 2 and 6 years’. In: *Child Phonology*. Ed. by GH Yeni-Komshian, JF Kavanagh, and CA Ferguson. Elsevier, pp. 169–192. DOI: 10.1016/B978-0-12-770601-6.50014-0.

Ioup, G (2008), ‘Exploring the role of age in the acquisition of a second language phonology’. In: *Phonology and second language acquisition*. Ed. by Hansen EJG and ML Zampini. Amsterdam: John Benjamins, pp. 41–62.

Irwin, JV and Wong, SP (1983), *Phonological Development in Children 18 to 72 Months*. Carbondale: Southern Illinois University Press.

Issa, AGE (2016), ‘Phonetic and Phonological Aspects of Gemination in Libyan Arabic’. PhD thesis. University of Leeds.

Iverson, P, Hazan, V, and Bannister, K (2005), ‘Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults’, *The Journal of the Acoustical Society of America*, vol. 118, no. 5, pp. 3267–3278. DOI: 10.1121/1.2062307.

Iverson, P and Kuhl, PK (1995), ‘Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling’, *The Journal of* the *Acoustical Society of America*, vol. 97, no. 1, pp. 553–562. DOI: 10.1121/1.412280.

Iverson, P, Kuhl, PK, Akahane-Yamada, R, Diesch, E, Tohkura, Y, Kettermann, A, and Siebert, C (2003), 'A perceptual interference account of acquisition difficulties for non-native phonemes', *Cognition*, vol. 87, no. 1, B47–B57.

Jabali, OM and Abuzaid, YJ (2017), 'Pronunciation errors committed by Palestinian students at An-Najah National University: An analytical approach'.

Jakobson, R (1968), *Child language. Aphasia and phonological universals*. Janua linguarum. Series minor, nr. 72. The Hague, Paris: Mouton.

Jesry, M (1996), 'Some cognitively controlled coarticulatory effects in Arabic and English, with particular reference to VOT'. PhD thesis. University of Essex.

Jones, D (1956), *An outline of English phonetics*. 8th ed. Cambridge: Cambridge University Press.

Jongman, A, Wayland, R, and Wong, S (2000), 'Acoustic characteristics of English fricatives', *The Journal of the Acoustical Society of America*, vol. 108, no. 3, pp. 1252–1263.

Kang, K and Guion, SG (2006), 'Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals', *The Journal of the Acoustical Society of America*, vol. 119, no. 3, pp. 1672–1683. DOI: 10.1121/1.2166607.

Karimi, S (1987), 'Farsi speakers and the initial consonant cluster in English', *Interlanguage phonology: The acquisition of a second language sound system*, pp. 305–318.

Kartushina, N, Hervais-Adelman, A, Frauenfelder, UH, and Golestani, N (2015), 'The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds', *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 817–832. DOI: 10.1121/1.4926561.

Kartushina, N, Hervais-Adelman, A, Frauenfelder, UH, and Golestani, N (2016), ‘Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training’, *Journal of Phonetics*, vol. 57, pp. 21–39.

Kartushina, N and Martin, CD (2019), ‘Talker and acoustic variability in learning to produce nonnative sounds: evidence from articulatory training’, *Language Learning*, vol. 69, no. 1, pp. 71–105.

Kenstowicz, MJ (1994), *Phonology in generative grammar*. Vol. 7. Blackwell Cambridge, MA.

Kenworthy, JM (1987), *Teaching English Pronunciation*. New York: Longman.

Kewley-Port, D and Preston, MS (1974), ‘Early apical stop production: a voice onset time analysis’, *Journal of Phonetics*, vol. 2, pp. 125–150. DOI: 10.1016/S0095-4470(19)31270-7.

Kharma, N and Hajjaj, A (1989), *Errors in English among Arabic speakers: Analysis and remedy*. London: Longman.

Khattab, G (2002a), ‘Sociololinguistic competence and the bilingual’s adoption of phonetic variants: auditory and instrumental data from English-Arabic bilinguals’. PhD thesis. University of Leeds.

— (2002b), ‘VOT production in English and Arabic bilingual and monolingual children’. In: *Perspectives on Arabic Linguistics XIII-XIV: Papers from the Thirteenth and Fourteenth Annual Symposia on Arabic Linguistics*. Ed. by E Benmamoun and DB Parkinson. Amsterdam: John Benjamins, pp. 1–38.

Khattab, G, Al-Tamimi, F, and Heselwood, B (2006), ‘Acoustic and auditory differences in the /t/-/T/opposition in male and female speakers of Jordanian Arabic’. In: *Perspectives on Arabic Linguistics XIV*. Ed. by S Boudela. Amsterdam: John Benjamins, pp. 131–160.

Khattab, G and Al-Tamimi, J (2014), ‘Geminate timing in Lebanese Arabic: the relationship between phonetic timing and phonological structure’, *Laboratory Phonology*, vol. 5, no. 2, pp. 231–269. DOI: 10.1515/lp-2014-0009.

Kilminster, MGE and Laird, EM (1978), ‘Articulation development in children aged three to nine years’, *Australian Journal of Human Communication Disorders*, vol. 6, no. 1, pp. 23–30.

Kingston, J and Diehl, RL (1994), ‘Phonetic knowledge’, *Language*, vol. 70, no. 3, pp. 419–454. DOI: 10.2307/416481.

Kirby, JP and Ladd, DR (2015), ‘Stop voicing and F0 perturbations: Evidence from French and Italian’. In: *ICPhS*.

Kirk, C (2008), ‘Substitution errors in the production of word-initial and word-final consonant clusters’, *Journal of Speech, Language, and Hearing Research*. DOI: 10.1044/1092-4388(2008/003).

Kirk, C and Demuth, K (2003), ‘Onset/coda asymmetries in the acquisition of clusters’. In: *Proceedings of the 27th annual Boston University conference on language development*, pp. 437–448.

— (2005), ‘Asymmetries in the acquisition of word-initial and word-final consonant clusters’, *Journal of Child Language*, vol. 32, no. 4, pp. 709–734. DOI: 10.1017/S0305000905007130.

Kissling, EM (2013), ‘Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners?’, *The Modern Language Journal*, vol. 97, no. 3, pp. 720–744. DOI: 10.1111/j.1540-4781.2013.12029.x.

Klatt, DH (1975), ‘Voice onset time, frication, and aspiration in word-initial consonant clusters’, *Journal of Speech and Hearing Research*, vol. 18, no. 4, pp. 686–706.

Kløve, MH and Young-Scholten, M (2008), ‘Repair of L2 syllables through metathesis’, *International Review of Applied Linguistics in Language Teaching*, vol. 39, no. 2, pp. 103–133. DOI: 10.1515/iral.39.2.103.

Kortmann, B and Upton, C (2008), *Varieties of English. 1, The British Isles*. Berlin: Mouton de Gruyter.

Krashen, SD (1982), *Principles and practice in second language acquisition*. Oxford Pergamon.

— (1985), *The input hypothesis: issues and implications*. London: Longman.

Krashen, SD (1994), 'The input hypothesis and its rivals'. In: *Implicit and explicit learning of languages*. Ed. by NC Ellis. Academic Press, pp. 45–77.

— (2018), 'Down with forced speech'. In: *Reconceptualizing English language teaching in the 21st century*. Ed. by YN Leung, J Katchen, SY Hwang, and Y Chen. ETA-ROC: Crane Publishing Company, pp. 99–106.

Kriba, HA (2009), 'Acoustic parameters of emphasis in Libyan Arabic'. PhD thesis. Newcastle upon Tyne.

Kuhl, PK (1991), 'Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not', *Attention, Perception, & Psychophysics*, vol. 50, no. 2, pp. 93–107.

— (1994), 'Learning and representation in speech and language', *Current Opinion in Neurobiology*, vol. 4, no. 6, pp. 812–822. DOI: 10.1016/0959-4388(94)90128-7.

Kuhl, PK and Iverson, P (1995), 'Linguistic experience and the perceptual magnet effect'. In: *Speech perception and linguistic experience: Issues in cross-language research*. Ed. by W Strange. Baltimore: York Press, pp. 121–154.

Kuhl, PK, Tsao, F-M, and Liu, H-M (2003), 'Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning', *Proceedings of the National Academy of Sciences*, vol. 100, no. 15, pp. 9096–9101. DOI: 10.1073/pnas.1532872100.

Kuhl, PK, Williams, KA, Lacerda, F, Stevens, KN, and Lindblom, B (1992), 'Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age', *Science*, vol. 255, pp. 606–608.

Kulikov, V (2016), 'Voicing in Qatari Arabic: Evidence for prevoicing and aspiration'. In: *Qatar Foundation Annual Research Conference Proceedings Volume 2016 Issue 1*. Vol. 2016. No. 1. Hamad bin Khalifa University Press (HBKU Press). DOI: 10.5339/qfarc.2016.SSHAPP2330.

Kwon, B-Y (2006), *Korean speakers' production of English consonant clusters: Articulatory and perceptual accounts*. Vol. 1. ProQuest.

Ladefoged, P and Maddieson, I (1996), *The sounds of the world's languages*. Vol. 1012. Blackwell Oxford.

Laradi, W (1983), 'Pharyngealization in Libyan (Tripoli) Arabic: an instrumental study'. PhD thesis. University of Edinburgh.

Larsen-Freeman, D (2000), *Techniques and principles in language*. Oxford: Oxford University Press.

Laver, J (1994), *Principles of phonetics*. Cambridge University Press.

Lee, J (2004), 'L1 transfer in L2 variation: The case of "disappearing /w/" in the English of Korean learners', *University of Wisconsin Linguistics Students Organization Working Papers in Linguistics*, vol. 4, no. 1, pp. 53–64.

Lee, JY (2009), 'The effects of pronunciation instruction using duration manipulation on the acquisition of English vowel sounds by pre-service Korean EFL teachers'. PhD thesis. University of Kansas.

Lee, S, Potamianos, A, and Narayanan, S (1999), 'Acoustics of children's speech: Developmental changes of temporal and spectral parameters', *The Journal of the Acoustical Society of America*, vol. 105, no. 3, pp. 1455–1468. DOI: 10.1121/1.426686.

Lee, SA and Iverson, G (2012), 'Stop consonant productions of Korean - English bilingual children', *Bilingualism*, vol. 15, no. 2, pp. 275–287. DOI: 10.1017/S1366728911000083.

Lehiste, I and Peterson, GE (1961), 'Some basic considerations in the analysis of intonation', *Journal of Acoustic Society of America*, vol. 33, no. 4, pp. 419–425.

Lenneberg, EH (1967), *Biological foundations of language*. New York: New York, Wiley.

Lenth, R (2019), *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.3.4. URL: https://CRAN.R-project.org/package=emmeans.

Levelt, CC, Schiller, NO, and Levelt, WJ (2000), 'The acquisition of syllable types', *Language acquisition*, vol. 8, no. 3, pp. 237–264.

Levis, J (2008), 'Computer Technology in Teaching and Researching Pronunciation', *Annual Review of Applied Linguistics*, vol. 27, pp. 184–202. DOI: 10.1017/s0267190508070098.

Lindblom, B (1963), 'Spectrographic Study of Vowel Reduction', *The Journal of the Acoustical Society of America*, vol. 35, no. 11, pp. 1773–1781. DOI: 10.1121/1.1918816.

— (1983), 'Economy of Speech Gestures'. In: *The Production of Speech*. Ed. by PF MacNeilage. New York, NY: Springer New York, pp. 217–245. DOI: 10.1007/978-1-4613-8202-7_10.

— (1990), 'Explaining phonetic variation: A sketch of the H&H theory'. In: *Speech Production and Speech Modelling*. Ed. by W Hardcastle and A Marchal. Dordrecht: Kluwer Academic Publishers, pp. 403–439.

— (1992), 'Phonological units as adaptive emergents of lexical development', *Phonological Development: Models, Research, Implications*, vol. 131, p. 163.

— (1996), 'Role of articulation in speech perception: clues from production', *The Journal of the Acoustical Society of America*, vol. 99, no. 3, pp. 1683–1692. DOI: 10.1121/1.414691.

Lindblom, B, Brownlee, S, Davis, B, and Moon, S-J (1992), 'Speech transforms', *Speech communication*, vol. 11, no. 4-5, pp. 357–368. DOI: 10.1016/0167-6393(92)90041-5.

Lisker, L and Abramson, AS (1964), 'A cross-language study of voicing in initial stops: Acoustical measurements', *Word*, vol. 20, no. 3, pp. 384–422. DOI: 10.1080/00437956.1964.11659830.

Liu, Q-X and Shi, J-F (2007), 'An analysis of language teaching approaches and methods–effectiveness and weakness', *US-China Education Review*, vol. 4, no. 1, pp. 69–71.

Lively, SE, Pisoni, DB, Yamada, RA, Tohkura, Y, and Yamada, T (1994), 'Training Japanese listeners to identify English/r/and/l/. III. Long-term retention of new phonetic categories', *The Journal of the acoustical society of America*, vol. 96, no. 4, pp. 2076–2087.

Lleó, C and Prinz, M (1996), ‘Consonant clusters in child phonology and the directionality of syllable structure assignment’, *Journal of Child Language*, vol. 23, no. 1, pp. 31–56.

Logan, JS, Lively, SE, and Pisoni, DB (1991), ‘Training Japanese listeners to identify English /r/ and /l/: A first report’, *The Journal of the Acoustical Society of America*, vol. 89, no. 2, pp. 874–886. DOI: 10.1121/1.1894649.

Long, MH (1990), ‘Maturational constraints on language development’, *Studies in second language acquisition*, vol. 12, no. 3, pp. 251–285.

Lord, Gillian (2005), ‘(How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course’, *Hispania*, pp. 557–567. DOI: 10.2307/20063159.

Lüdecke, D (2018), ‘ggeffects: Tidy Data Frames of Marginal Effects from Regression Models’, *Journal of Open Source Software*, vol. 3, no. 26, p. 772. DOI: 10.21105/joss.00772.

Mack, M (2003), ‘The phonetic systems of bilinguals’, *Mind, brain, and language: Multidisciplinary perspectives*, pp. 309–349.

MacKain, KS, Best, CT, and Strange, W (1981), ‘Categorical perception of English/r/and/l/by Japanese bilinguals’, *Applied Psycholinguistics*, vol. 2, no. 4, pp. 369–390.

MacKay, I, Flege, JE, Piske, T, and Schirru, C (2001), ‘Category restructuring during second-language speech acquisition’, *The Journal of the Acoustical Society of America*, vol. 110, no. 1, pp. 516–528. DOI: 10.1121/1.1377287.

Macken, MA and Barton, D (1977), ‘A longitudinal study of the acquisition of the voicing contrast in American-English word-initial stops, as measured by voice onset time’, *Papers and Reports on Child Language Development*, no. 14, pp. 73–120.

— (1980a), ‘The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants’, *Journal of Child Language*, vol. 7, pp. 41–74. DOI: 10.1017/S0305000900007029.

Macken, MA and Barton, D (1980b), ‘The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants’, *Journal of Child Language*, vol. 7, no. 3, pp. 433–458.

Mackey, A (2007), *Conversational interaction in second language acquisition: a collection of empirical studies.* Oxford: Oxford University Press.

MacLeod, AAN (2016), ‘Phonetic and phonological perspectives on the acquisition of voice onset time by French-speaking children’, *Clinical Linguistics & Phonetics*, vol. 30, no. 8, pp. 584–598. DOI: 10.3109/02699206.2016.1152509.

MacNeilage, PF (1980), ‘The control of speech production’. In: *Child phonology.* Elsevier, pp. 9–21.

Maddieson, I (1984), *Patterns of sounds.* Cambridge studies in speech science and communication. Cambridge University Press.

Major, RC (1994), ‘Chronological and stylistic aspects of second language acquisition of consonant clusters’, *Language learning*, vol. 44, no. 4, pp. 655–680.

— (2001), *Foreign accent: The ontogeny and phylogeny of second language phonology.* Routledge.

— (2007), ‘Identifying a foreign accent in an unfamiliar language’, *Studies in Second Language Acquisition*, vol. 29, no. 4, pp. 539–556. DOI: 10.1017/S0272263107070428.

McCarthy, JJ (1989), ‘Linear order in phonological representation’, *Linguistic inquiry*, vol. 20, no. 1, pp. 71–99.

— (2007), *Hidden generalizations: phonological opacity in Optimality Theory.* Equinox Publishing.

McCormack, PF and Knighton, T (1996), ‘Gender differences in the speech patterns of two and a half year old children’. In: *Speech science and technology: Sixth Australian International Conference.* Australian Speech Science and Technology Association Adelaide, Australia, pp. 337–341.

McGurk, H and MacDonald, J (1976), 'Hearing lips and seeing voices', *Nature*, vol. 264, no. 5588, pp. 746–748. DOI: 10.1038/264746a0.

McIntosh, B and Dodd, B (2008), 'Two-year-olds' phonological acquisition: Normative data', *International Journal of Speech-Language Pathology*, vol. 10, no. 6, pp. 460–469.

Mcleod, S and Crowe, K (2018), 'Children's Consonant Acquisition in 27 Languages: A Cross-Linguistic Review', pp. 1546–1571. DOI: 10.23641/asha.

McLeod, S, van Doorn, J, and Reed, VA (1998), 'Homonyms in children's productions of consonant clusters'. In: *Clinical Phonetics and Linguistics*. Ed. by W Zeigler and K Deger. Whurr Publishers Ltd, pp. 106–114.

— (2001a), 'Consonant Cluster Development in Two-Year-Olds: General Trends and Individual Differences', *Journal of Speech, Language, and Hearing Research*, vol. 44, no. 5, pp. 1144–1171. DOI: 10.1044/1092-4388(2001/090).

— (May 2001b), 'Normal Acquisition of Consonant Clusters', *American Journal of Speech-Language Pathology*, vol. 10, pp. 99–110. DOI: 10.1044/1058-0360(2001/011).

Miller, JL and Eimas, PD (1996), 'Internal structure of voicing categories in early infancy', *Perception & Psychophysics*, vol. 58, no. 8, pp. 1157–1167. DOI: 10.3758/BF03207549.

Milroy, L (2007), 'Off the shelf or under the counter? On the social dynamics of sound changes'. In: *Managing chaos: Strategies for identifying change in English (Studies in the History of the English Language 3)*. Ed. by CM Cain and G Russom. Berlin and New York: Mouton de Gruyter, pp. 149–72.

Mitchell, TF (1993), *Pronouncing Arabic*. Vol. 2. Oxford: Oxford University Press.

Mitleb, F (2001), 'Voice onset time of Jordanian Arabic stops', *Journal of the Acoustical Society of America*, vol. 109, no. 5, p. 2474.

Monaghan, C (2014), 'Are norms for percent consonant correct and the age of acquisition of phonemes in the Irish DEAP valid for a sample of typically developing children in Limerick?' MA thesis. Limerick, Republic of Ireland: University of Limerick.

Mowrer, DE and Burger, S (1991), 'A comparative analysis of phonological acquisition of consonants in the speech of 21/2-6-year-old Xhosa-and English-speaking children', *Clinical Linguistics & Phonetics*, vol. 5, no. 2, pp. 139–164.

Muftah, A (2001), 'English and Arabic sound systems: similarities and differences and their effects on learning English by Arab students'. PhD thesis.

Muñoz, C (June 2014), 'Contrasting effects of starting age and input on the oral performance of foreign language learners', *Applied Linguistics*, vol. 35, no. 4, pp. 463–482. DOI: 10.1093/applin/amu024. eprint: http://oup.prod.sis.lan/applij/article-pdf/35/4/463/9353831/amu024.pdf.

Munson, B, Edwards, J, Beckman, ME, Cohn, AC, Fougeron, C, and Huffman, MK (2011), 'Phonological representations in language acquisition: Climbing the ladder of abstraction', *Handbook of laboratory phonology*, pp. 288–309.

Nagamine, T (2011), 'Effects of hyper-pronunciation training method on Japanese university students' pronunciation', *Asian EFL Journal Professional Teaching Articles*, vol. 53, pp. 35–50.

Neri, A, Mich, O, Gerosa, M, and Giuliani, D (2008), 'The effectiveness of computer assisted pronunciation training for foreign language learning by children', vol. 21, no. 5, pp. 393–408. DOI: 10.1080/09588220802447651.

Neufeld, GG (1978), 'On the acquisition of prosodic and articulatory features in adult language learning', *Canadian Modern Language Review*, vol. 34, no. 2, pp. 163–174. DOI: 10.3138/cmlr.34.2.163.

— (1980), 'On the Adult's Ability to Acquire Phonology', *TESOL Quarterly*, vol. 14, no. 3, pp. 285–298.

Al-Nuzaili, AM (1993), 'Experimental study of emphasis and voicing in the plosives of Yemeni Spoken Arabic with some implications for foreign language teaching and learning'. PhD thesis. University of Leeds.

O'Brien, MG (2014), 'L2 Learners' Assessments of Accentedness, Fluency, and Comprehensibility of Native and Nonnative German Speech', *Language Learning*, vol. 64, no. 4, pp. 715–748. DOI: 10.1111/lang.12082.

Odisho, E (1973), ‘The Role of the Rear Section of the Vocal Tract in Baghdadi Arabic’. PhD thesis. University of Leeds.

Ohala, DK (1999), ‘The influence of sonority on children’s cluster reductions’, *Journal of communication disorders*, vol. 32, no. 6, pp. 397–422.

— (2008), ‘Phonological acquisition in a first language1’. In: *Phonology in Second Language Acquisition*. Ed. by JGH Edwards and ML Zampini. Amsterdam: John Benjamins, pp. 19–39.

Ohala, JJ (1981), ‘Articulatory constraints on the cognitive representation of speech’. In: *Advances in Psychology*. Vol. 7. Elsevier, pp. 111–122.

— (1993a), ‘Sound change as nature’s speech perception experiment’, *Speech Communication*, vol. 13, pp. 155–61.

— (1993b), ‘The phonetics of sound change’. In: *Historical linguistics: Problems and perspectives*. Ed. by C Jones. London: Longman Academic, pp. 237–78.

— (1997), ‘Aerodynamics of phonology’. In: *Proceedings of the Seoul International Conference on Linguistics*. Vol. 92, p. 97.

Omar, MK (1973), ‘The acquisition of Egyptian Arabic as a native language’, *Janua Linguarum, Series Pratica*, vol. 160, pp. 199–205.

Omar, Y.Z. (2014), ‘Perceptions of selected Libyan teachers of English as a foreign language regarding teaching of English in Libya’. PhD thesis. University of Missouri-Columbia.

Orafi, SMS and Borg, S (2009), ‘Intentions and realities in implementing communicative curriculum reform’, *System*, vol. 37, no. 2, pp. 243–253.

Overal, JW (2014), *DigLin*. URL: http://diglin.eu/.

Owens, J (1984), *A short reference grammar of Eastern Libyan Arabic*. O Harrassowitz.

Owolabi, D (2012), ‘Production and perception problems of English dental fricatives by Yoruba speakers of English as a second language’, *Theory and Practice in Language Studies*, vol. 2, no. 6, p. 1108. DOI: 10.4304/tpls.2.6.1108-1113.

Panetta, E (1940), *Pratiche e credenze popolari libiche: testi in arabo bengasino tradotti e annotati.* Istituto per l'Oriente.

— (1943), *L'arabo parlato a Bengasi.* Vol. 1. Libreria dello stato.

Paradis, C and Prunet, J-F (1991), *Phonetics and phonology, Vol. 2: The special status of coronals: Internal and external evidence.*

Paradis, J (2008), 'Second language acquisition in childhood'. In: *Blackwell handbook of language development.* Ed. by E Hoff and M Shatz. Williston: John Wiley & Sons, Incorporated, pp. 387–405.

Parlak, O (2010), 'Does pronunciation instruction promote intelligibility and comprehensibility'. In: *Annual Meeting for TESOL.* Boston, MA, USA.

Pater, J and Barlow, JA (2003), 'Constraint conflict in cluster reduction', *Journal of child language*, vol. 30, no. 3, pp. 487–526.

Pearson, BZ, Velleman, SL, Bryant, TJ, and Charko, T (2009), 'Phonological milestones for African American English-speaking children learning mainstream American English as a second dialect', *Language, Speech, and Hearing Services in Schools*, pp. 229–244.

Pearson, P, Pickering, L, and Da Silva, R (2011), 'The impact of computer assisted pronunciation training on the improvement of Vietnamese learner production of English syllable margins'. In: *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference.* Iowa State University, pp. 169–180.

Pereira, C (2007), 'Urbanisation and dialect change the Arabic dialect of Tripoli (Libya)'. In: *Arabic in the city: Issues in dialect contact and language variation.* Ed. by C Miller, E Al-Wer, D Caubet, and JCE Watson. Routledge, pp. 77–96.

— (2008), 'Le parler arabe de Tripoli (Libye): phonologie, morphosyntaxe et catégories grammaticales', p. 628.

Perrachione, TK, Lee, J, Ha, LYY, and Wong, PCM (2011), 'Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design', *The Journal of the Acoustical Society of America*, vol. 130, no. 1, pp. 461–472. DOI: 10.1121/1.3593366.

Perry, TL, Ohde, RN, and Ashmead, DH (2001), 'The acoustic bases for gender identification from children's voices', *The Journal of the Acoustical Society of America*, vol. 109, no. 6, pp. 2988–2998.

Peterson, GE and Barney, HL (1952), 'Control methods used in a study of the vowels', *The Journal of the acoustical society of America*, vol. 24, no. 2, pp. 175–184.

Pfenninger, S (2016), 'All good things come in threes: Early English learning, CLIL and motivation in Switzerland', *Cahiers de IILSL*, vol. 48, pp. 119–147.

Phoon, HS (2010), 'The Phonological Development of Malaysian English Speaking Chinese Children: A Normative Study'. PhD thesis. Christchurch, NewZealand: University of Canterbury.

Piske, T (2012), 'Factors affecting the perception and production of L2 prosody: Research results and their implications for the teaching of foreign languages'. In: *Pragmatics, Prosody and English Language Teaching*. Ed. by J Romero-Trillo. Dordrecht: Springer, pp. 41–59.

— (2017), 'The earlier, the better? Some critical remarks on current EFL teaching to young learners and their implications for foreign language teacher education'. In: *The Professional Development of Primary EFL teachers. National and International Research*. Ed. by R Wilden and R Porsch. Münster, New York: Waxmann, pp. 45–57.

Piske, T, Flege, JE, MacKay, IRA, and Meador, D (2002), 'The Production of English Vowels by Fluent Early and Late Italian-English Bilinguals', *Phonetica*, vol. 59, pp. 49–71. DOI: 10.1159/000056205.

Piske, T, MacKay, I, and Flege, JE (2001), 'Factors affecting degree of foreign accent in an L2: A review', *Journal of Phonetics*, vol. 29, no. 2, pp. 191–215. DOI: 10.1006/jpho.2001.0134.

Poole, I (1934), 'The genetic development of consonant sounds in the speech of children', *The Elementary English Review*, vol. 11, pp. 159–161.

Powell, TW and Elbert, M (1984), ‘Generalization following the remediation of early-and later-developing consonant clusters’, *Journal of Speech and Hearing Disorders*, vol. 49, no. 2, pp. 211–218.

Prather, EM, Hedrick, DL, and Kern, CA (1975), ‘Articulation development in children aged two to four years’, *Journal of Speech and Hearing Disorders*, vol. 40, no. 2, pp. 179–191.

Preisser, DA, Hodson, BW, and Paden, EP (1988), ‘Developmental Phonology: 18–29 months’, *Journal of Speech and Hearing Disorders*, vol. 53, no. 2, pp. 125–130. DOI: 10.1044/jshd.5302.125.

Pulgram, E (1965), ‘Consonant cluster, consonant sequence, and the syllable’, *Phonetica*, vol. 13, no. 1-2, pp. 76–81. DOI: 10.1159/000258473.

R Core Team (2013), *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Rababah, Ghaleb (2003), ‘Communication problems facing Arab learners of English: A personal perspective’, *TEFL Web Journal*, vol. 2, no. 1, pp. 15–30.

Repp, BH (1982), ‘Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception’, *Psychology Bulletin*, vol. 92, no. 1, pp. 81–110.

Rifaat, K (2003), ‘Voice Onset Time in Egyptian Arabic: A case where phonological categories dominate’. In: *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 791–794.

Roach, P (1990), *English Phonetics and Phonology: A Practical Course (Tutor’s Book)*. Cambridge: Cambridge University Press.

— (2009), *English Phonetics and Phonology: Units 10-20*. Cambridge University Press.

Roark, B and Demuth, K (2000), ‘Prosodic constraints and the learner’s environment: A corpus study’. In: *Proceedings of the 24th Annual Boston*

University Conference on Language Development. Vol. 2. Citeseer, pp. 597–608.

Roberts, JE, Burchinal, M, and Footo, MM (1990), ‘Phonological process decline from 2 to 8 years’, *Journal of Communication Disorders*, vol. 23, no. 3, pp. 205–217. DOI: 10.1016/0021-9924(90)90023-R.

Rosenblum, LD (2008), ‘Speech Perception as a Multimodal Phenomenon’, *Current Directions in Psychological Science*, vol. 17, no. 6, pp. 405–409. DOI: 10.1111/j.1467-8721.2008.00615.x.

Saaid, N (2010), ‘The use of technology (audio-visual aids) in teaching English as a foreign language in Benghazi secondary schools’. PhD thesis. Leicester University.

Al-Saidat, EM (2010), ‘Phonological analysis of English phonotactics: A case study of Arab learners of English’, *The Buckingham Journal of Language and Linguistics*, vol. 3, pp. 121–134.

Sakai, M (2016), ‘(Dis) Connecting perception and production: Training native speakers of Spanish on the English /i/-/ɪ/ distinction’. PhD thesis. Georgetown University.

Saleh, M, Shoeib, R, Hegazi, M, and Ali, P (2007), ‘Early phonological development in Arabic Egyptian children: 12–30 months’, *Folia Phoniatrica et Logopaedica*, vol. 59, no. 5, pp. 234–240.

Saleh, N and Zakaria, N (2013), ‘Investigating the difficulties faced in understanding, and strategies used in processing, English idioms by the Libyan students’, *International Journal of English Language and Translation Studies*, vol. 1, no. 2, pp. 69–91.

Samuel, AG (2011), ‘Speech Perception’, *Annual Review of Psychology*, vol. 62, no. 1, pp. 49–72. DOI: 10.1146/annurev.psych.121208.131643.

Sardegna, VG (2011), ‘Pronunciation learning strategies that improve ESL learners’ linking’. In: *Pronunciation and intelligibility: Issues in research and practice. Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*. Iowa State University, pp. 105–121.

Schielzeth, H (2010), ‘Simple means to improve the interpretability of regression coefficients’, *Methods in Ecology and Evolution*, vol. 1, no. 2, pp. 103–113. DOI: 10.1111/j.2041-210X.2010.00012.x.

Schleef, E and Ramsammy, M (2013), ‘Labiodental fronting of /θ/ in London and Edinburgh: a cross-dialectal study’, *English Language and Linguistics*, vol. 17, no. 1, pp. 25–54. DOI: 10.1017/S1360674312000317.

Scobbie, JM (1998), ‘Interactions between the acquisition of phonetics and phonology’. In: *Papers from the 34th Annual Regional Meeting of the Chicago Linguistic Society*. Ed. by CM Gruber, D Higgins, K Olsen, and T Wysocki. Chicago: Chicago Linguistics Society, pp. 343–358.

Scobbie, JM, Gibbon, F, William, J, Hardcastle, WJ, and Fletcher, P (2000), ‘Covert contrast as a stage in the acquisition of phonetics and phonology’. In: *Papers in laboratory phonology V: Acquisition and the lexicon*. Ed. by Michael B Broe and Janet B Pierrehumbert. Cambridge: Cambridge University Press, pp. 194–207.

Scobbie, JM and Wrench, AA (2003), ‘An articulatory investigation of word final/l/and/l/-sandhi in three dialects of English.’ In: *Proceedings of the 15th International Congress of Phonetic Sciences*. Universitat Autnoma de Barcelona.

Segal, O, Hejli-Assi, S, and Kishon-Rabin, L (2016), ‘The effect of listening experience on the discrimination of /ba/ and /pa/ in Hebrew-learning and Arabic-learning infants’, *Infant Behavior & Development*, vol. 42, pp. 86–99. DOI: 10.1016/j.infbeh.2015.10.002.

Settinieri, J (2008), *Teaching pronunciation*. Ed. by S Gramley and V Gramley. Bielefeld: Aisthesis Verlag, pp. 1–12.

Shafiro, V, Levy, ES, Khamis-Dakwar, R, and Kharkhurin, A (2013), ‘Perceptual confusions of American-English vowels and consonants by native Arabic bilinguals’, *Language and Speech*, vol. 56, no. 2, pp. 145–161. DOI: 10.1177/0023830912442925.

Shaheen, K (1979), 'The acoustic analysis of Arabic speech'. PhD thesis. University of Wales, Bangor.

Shahin, KN (1996), 'Accessing pharyngeal place in Palestinian Arabic'. In: *Perspectives on Arabic linguistics IX*. Ed. by Mushira Eid and Dilworth B Parkinson. Vol. 4. Amsterdam studies in the theory and history of linguistic science series 5. Amsterdam: John Benjamins Publishing Company, pp. 131–150.

Sharwood Smith, M (1993), 'Input enhancement and instructed second language acquisition', *Studies in second language acquisition*, vol. 15, no. 2.

Shitaw, A (2014), 'An instrumental phonetic investigation of timing relations in two-stop consonant clusters in Tripolitan Libyan Arabic'. PhD thesis. The University of Leeds.

Showalter, CE and Hayes-Harb, R (2015), 'Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition', *Applied Psycholinguistics*, vol. 36, no. 1, pp. 23–42. DOI: 10.1017/S0142716414000411.

Shriberg, L (1993), 'Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders', *Journal of Speech, Language, and Hearing Research*, vol. 36, no. 1, pp. 105–140. DOI: 10.1044/jshr.3601.105.

Shriberg, L and Kwiatkowski, J (1980), *Procedures for natural process analysis of continuous speech samples*. New York: John Wiley & Sons.

— (1994), 'Developmental phonological disorders .1. a clinical profile', *Journal of Speech and Hearing Research*, vol. 37, no. 5, pp. 1100–1126. DOI: 10.1044/jshr.3705.1100.

Silverman, K (1986), 'F0 segmental cues depend on intonation: the case of the rise after voiced stops', *Phonetica*, vol. 43, pp. 76–91.

Singmann, H, Bolker, B, Westfall, J, Aust, F, and Ben-Shachar, MS (2019), *afex: Analysis of Factorial Experiments*. R package version 0.25-1. URL: https://CRAN.R-project.org/package=afex.

Smit, AB (1993), ‘Phonologic error distributions in the Iowa-Nebraska Articulation Norms Project: Word-initial consonant clusters’, *Journal of Speech, Language, and Hearing Research*, vol. 36, no. 5, pp. 931–947.

Smit, AB, Hand, L, Freilinger, JJ, Bernthal, JE, and Bird, A (1990), ‘The Iowa articulation norms project and its Nebraska replication’, *Journal of Speech and Hearing Disorders*, vol. 55, no. 4, pp. 779–798.

Smith, BL (1978), ‘Temporal aspects of English speech production: A developmental perspective’, *Journal of Phonetics*, vol. 6, no. 1, pp. 37–67.

Smith, NV (1973), *The acquisition of phonology: A case study.* Cambridge: Cambridge University Press.

Speck, BP (2002), ‘Markedness and naturalness in the acquisition of phonology’.

Steele, J (2005), ‘Assessing the role of orthographic versus uniquely auditory input in acquiring new L2 segments’, *7èmes Rencontres Internationales du Réseau Français de Phonologie*, pp. 2–4.

Stevens, KN (2000), *Acoustic phonetics.* MIT press.

Stevens, KN, Manuel, SY, and Matthies, M (1999), ‘Revisiting place of articulation measures for stop consonants: Implications for models of consonant production’. In: *Proceedings of the International Congress of Phonetic Sciences*, pp. 1117–1120.

Stockman, IJ (2006), ‘Alveolar bias in the final consonant deletion patterns of African American children’, *Language Speech And Hearing Services in Schools*, vol. 37, pp. 85–95.

Stoel-Gammon, C (1987), ‘Phonological skills of 2-year-olds’, *Language, Speech, and Hearing Services in Schools*, vol. 18, no. 4, pp. 323–329.

Stoel-Gammon, C and Buder, E (1999), ‘Vowel length, postvocalic voicing and VOT in the speech of two-year olds’. In: *Proceedings of the XIIIth International Conference of Phonetic Sciences.* Vol. 3, pp. 2485–2488.

Stoel-Gammon, C and Dunn, C (1985), *Normal and disordered phonology in children.* Baltimore: University Park Press.

Stokes, SF and Surendran, D (2005), 'Articulatory complexity, ambient frequency, and functional load as predictors of consonant development in children', *Journal of Speech, Language, and Hearing Research*, vol. 48, pp. 577–591. DOI: 10.1044/1092-4388(2005/040).

Storkel, HL (2004), 'The emerging lexicon of children with phonological delays', *Journal of Speech, Language, and Hearing Research*.

Suayeh, S (1994), *Report on teaching English in the pre-university education in Libya, Tripoli*. The General People's Committee for Education, Scientific Research in collaboration with The National Centre for Research in Education, and Training.

Suchato, A (2004a), 'Classification of stop consonant place of articulation'. PhD thesis. Massachusetts Institute of Technology.

— (2004b), 'Classification of stop consonant place of articulation: Combining acoustic attributes', *Speech Communication Group, Laboratory of Electronics*, pp. 197–202.

— (2005), 'Classification of Stop Consonant Place of Articulation', *Dissertation Abstracts International, B: Sciences and Engineering*, vol. 65, no. 9, pp. 197–202.

Sumdangdej, S (2007), 'Input and the acquisition of suprasegmental phonology in English by Thai school children'. PhD thesis. Durham University.

Swain, M (1985), 'Communicative competence: Some roles of comprehensible input and comprehensible output in its development'. In: *Input in second language acquisition*. Ed. by S Gass and C Madden. MA: Newbury House, pp. 235–53.

— (1988), 'Manipulating and complementing content teaching to maximize second language learning', *TESL Canada Journal*, pp. 68–83. DOI: 10.18806/tesl.v6i1.542.

— (1995), 'Thee functions of output in second language learning'. In: *Principles and practice in applied linguistics: Studies in honor of HG Widdowson*. Ed. by G Cook and B Seidlhoffer. Oxford: Oxford University Press, pp. 125–144.

Swain, M (2005), ‘The output hypothesis: Theory and research’. In: *Handbook of research in second language teaching and learning*. Ed. by E Hinkel. New York: Routledge, pp. 495–508. DOI: 10.4324/9781410612700.

Swe’iy, A (2013), *Allugha Allibiya alqadima wa kitabatiha*. Tripoli: Dar Alrowad.

Swei’y, A (Feb. 2009), ‘’Oroubat Allugha Allibiya Al-Qadima wa Kitabatuha: Muqaraba baina Allibiya wa Alamazighiya’. An optional note. PhD thesis. Damascus, Syria: St. Clements University.

Syed, NA (2013), ‘Acquisition of English Dental Fricatives by Pakistani Learners’, *Newcastle Working Papers in Linguistics*, p. 59.

Tabain, M (1998), ‘Non-sibilant fricatives in English: spectral information above 10 kHz’, *Phonetica*, vol. 55, pp. 107–130. DOI: 10.1159/000028427.

Al-Tamimi, F and Heselwood, B (2011), ‘Nasoendoscopic, videofluoroscopic and acoustic study of plain and emphatic coronals in Jordanian Arabic’. In: *Instrumental Studies in Arabic Phonetics*. John Benjamins, pp. 163–192.

Al-Tamimi, J (2007), ‘Static and dynamic cues in vowel production: A cross dialectal study in Jordanian and Moroccan Arabic’. In: *16th International Congress of Phonetic Sciences (ICPhS), Saarbrücken, Germany, August*.

— (2017), ‘Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations’, *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 8, no. 1, p. 28. DOI: 10.5334/labphon.19.

Al-Tamimi, J, Carré, R, and Marsico, E (2004), ‘The status of vowels in Jordanian and Moroccan Arabic: Insights from production and perception’.

Al-Tamimi, J and Khattab, G (2015), ‘Acoustic cue weighting in the singleton vs geminate contrast in Lebanese Arabic: The case of fricative consonants’, *The Journal of the Acoustical Society of America*, vol. 138, no. 1, pp. 344–360.

Tanner, M and Landon, M (2009), ‘The effects of computer-assisted pronunciation readings on ESL learners’ use of pausing, stress, intonation, and overall comprehensibility’, *Language Learning & Technology*, vol. 13, no. 3, pp. 51–65.

Temp1in, MC (1957), *Certain Language Skills in Children*. Minneapolis: University of Minnesota.

Templin, M (1957), 'Certain language skills in children; their development and interrelationships'.

Thompson, I (1991), 'Foreign accents revisited: The English pronunciation of Russian immigrants', *Language Learning*, vol. 41, no. 2, pp. 177–204. DOI: 10.1111/j.1467-1770.1991.tb00683.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-1770.1991.tb00683.x.

Thomson, RI (2011), 'Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation', *Calico Journal*, vol. 28, no. 3, pp. 744–765.

Thomson, RI and Derwing, TM (Dec. 2014), 'The effectiveness of L2 pronunciation instruction: A narrative review', *Applied Linguistics*, vol. 36, no. 3, pp. 326–344. DOI: 10.1093/applin/amu076. eprint: http://oup.prod.sis.lan/applij/article-pdf/36/3/326/7254651/amu076.pdf.

Tillmann, HG (1995), 'Early Modern Instrumental Phonetics'. In: *Concise History of the Language Sciences*. Ed. by EFK Koerner and RE Asher. Amsterdam: Pergamon, pp. 401–416. DOI: 10.1016/B978-0-08-042580-1.50063-5.

Tindall-Ford, S, Chandler, P, and Sweller, J (1997), 'When two sensory modes are better than one', *Journal of Experimental Psychology: Applied*, vol. 3, no. 4, p. 257. DOI: 10.1037/1076-898X.3.4.257.

Tollfree, L (1999), 'South East London English: discrete versus continuous modelling of consonantal reduction'. In: *Urban Voices, Accent Studies in the British Isles*. Arnold, pp. 163–184.

Trofimovich, P, Lightbown, PM, Halter, RH, and Song, H (2009), 'Comprehension-based Practice: The Development of L2 Pronunciation in a Listening and Reading Program', *Studies in Second Language Acquisition*, vol. 31, no. 4, pp. 609–639. DOI: 10.1017/S0272263109990040.

Trudgill, P (1984), *Language in the British isles*. CUP Archive.

Tyler, MD, Best, CT, Faber, A, and Levitt, AG (2014), 'Perceptual assimilation and discrimination of non-native vowel contrasts', *Phonetica*, vol. 71, no. 1, pp. 4–21.

UNESCO-IBE (2007), *Libyan Arab Jamahiriya*. http://www.ibe.unesco.org/. Accessed: 2016-05-01.

van der Linde, KJ (2001), 'Sonority substitutions'. Unpublished. PhD thesis. University of Groningen.

Versteegh, K (2014), *The Arabic language*. 2nd ed.

Vihman, M (1996), *Phonological development: the origins of language in the child*. Applied language studies. Cambridge, Mass: Blackwell.

Vokic, G (2011), 'When alphabets collide: Alphabetic first-language speakers' approach to speech production in an alphabetic second language', *Second Language Research*, vol. 27, no. 3, pp. 391–417. DOI: 10.1177/0267658310396627.

Vorperian, HK and Kent, RD (2007), 'Vowel acoustic space development in children: A synthesis of acoustic and anatomic data', *Journal of Speech, Language, and Hearing Research*, vol. 50, no. 6, pp. 1510–1545. DOI: 10.1044/1092-4388(2007/104).

Wang, X (2002), 'Training Mandarin and Cantonese speakers to identify English vowel contrasts: Long-term retention and effects on production'. PhD thesis. Simon Fraser University.

Wang, Y, Jongman, A, and Sereno, JA (2003), 'Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training', *The Journal of the Acoustical Society of America*, vol. 113, no. 2, pp. 1033–1043. DOI: 10.1121/1.1531176.

Watson, JCE (2002), *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

Watson, MM and Scukanec, GP (1997), 'Phonological Changes in the Speech of Two Year Olds: A Longitudinal Investigation', *Infant-Toddler Intervention: The Transdisciplinary Journal*, vol. 7, no. 1, pp. 67–77.

Weber, A and Cutler, A (2004), ‘Lexical competition in non-native spoken-word recognition’, *Journal of Memory and Language*, vol. 50, no. 1, pp. 1–25. DOI: 10.1016/S0749-596X(03)00105-0.

Weinberg, A and Knoerr, H (2003), ‘Learning French pronunciation: Audiocassettes or multimedia?’, *Calico Journal*, pp. 315–336.

Weinberger, S (1987), ‘The influence of linguistic context on syllable simplification’. In: *Interlanguage Phonology*. Ed. by G Ioup and . Weinberger. Rowley, MA: Newbury House.

Wellman, BL, Case, IM, Mengert, IG, and Bradbury, DE (1931), *Speech sounds by young children*. Vol. 5. 2. Iowa City: University of Iowa Child Welfare Research Station.

Wells, JC (1982), *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.

Wenk, BJ (1979), ‘Articulatory setting and de-fossilization’, *Interlanguage Studies Bulletin*, vol. 4, pp. 202–220.

— (1983), ‘Articulatory setting and the acquisition of second language phonology’, *Revue de phonetique appliquee*, vol. 65, pp. 51–65.

Werker, JF and Tees, RC (1984), ‘Phonemic and phonetic factors in adult cross-language speech perception’, *The Journal of the Acoustical Society of America*, vol. 75, no. 6, pp. 1866–1878.

Wester, F, Gilbers, D, and Lowie, W (2007), ‘Substitution of dental fricatives in English by Dutch L2 speakers’, *Language Sciences*, vol. 29, no. 2-3, pp. 477–491. DOI: 10.1016/j.langsci.2006.12.029.

Whiteside, SP, Dobbin, R, and Henry, L (2003), ‘Patterns of variability in voice onset time: a developmental study of motor speech skills in humans’, *Neuroscience Letters*, vol. 347, no. 1, pp. 29–32. DOI: 10.1016/S0304-3940(03)00598-6.

Whiteside, SP, Henry, L, and Dobbin, R (2004), ‘Sex differences in voice onset time: A developmental study of phonetic context effects in British English’, *The*

Journal of the Acoustical Society of America, vol. 116, no. 2, pp. 1179–1183. DOI: 10.1121/1.1768256.

Williams, L (1977), 'The perception of stop consonant voicing by Spanish-English bilinguals', *Attention, Perception, & Psychophysics*, vol. 21, no. 4, pp. 289–297.

Winter, B (2019), *Statistics for Linguists: An Introduction Using R*. Routledge.

Wode, H (1996), 'Speech perception and L2 phonological acquisition'. In: *Investigating Second Language Acquisition*. Ed. by P Jordens and J Lalleman. Berlin: Mouton de Gruyter, pp. 321–353.

World Popoulation Review (Jan. 2017), *World Population Review*. Online. URL: http://worldpopulationreview.com/countries/libya-population/.

Yeni-Komshian, GH, Caramazza, A, and Preston, MS (1977), 'A study of voicing in Lebanese Arabic', *Journal of Phonetics*, vol. 5, no. 1, pp. 35–48.

Young-Scholten, M (1995), 'The negative effects of 'positive' evidence on L2 phonology'. In: *The Current State of Interlanguage*. Ed. by L Eubank, L Selinker, and M Sharwood Smith. Amsterdam: John Benjamins Publishing Company Amsterdam, pp. 107–121.

— (1998), 'Second language syllable simplification; Deviant development or deviant input?' In: *New Sounds '97: Proceedings of the Third international symposium on the acquisition of second-language speech*. Ed. by J Allan and J Leather. Amsterdam: Amsterdam University, pp. 351–360.

— (2002), 'Orthographic input in L2 phonological development'. In: *An integrated view of language development: Papers in honour of Henning Wode*. Ed. by P Burmeister, T Piske, and A Rohde. Trier: Wissenschaftlicher Verlag Trier, pp. 263–279.

Young-Scholten, M, Akita, M, and Cross, N (1999), 'Focus on form in phonology: Orthographic exposure as a promoter of epenthesis'. In: *Pragmatics and Pedagogy: Proceedings of the Third PacSLRF*. Ed. by P Robinson and NO Jungheim. Vol. 2. Tokyo: Aoyama Gakuin University, pp. 227–233.

Zampini, ML (1994), ‘The role of native language transfer and task formality in the acquisition of Spanish spirantization’, *Hispania*, vol. 77, no. 3, pp. 470–481.

— (2008), ‘L2 speech production research: Findings, issues, and advances’. In: *Phonology and Second Language Acquisition*. Ed. by JGH Edwards and ML Zampini. Amsterdam: John Benjamins, pp. 219–250.

Zamuner, T (2003), *Input-based phonological acquisition*. Routledge.

Zamuner, T, Gerken, L, and Hammond, M (2004), ‘Phonotactic probabilities in young children’s speech production’, *Journal of child language*, vol. 31, no. 3, pp. 515–536.

Zughoul, MR (1980), ‘Diglossia in Arabic: investigating solutions’, *Anthropological Linguistics*, vol. 22, no. 5, pp. 201–217.