Non-Native Production of Arabic Pharyngeal
and Pharyngealized Consonants

by

Sawsan A. Alwabari

A thesis submitted in conformity with the requirements
for the degree of Master of Arts
in Applied Linguistics and Discourse Studies

Carleton University

©2013 Sawsan Alwabari
Abstract

Grounded on the theoretical constructs of cross-language speech learning, this research investigates English L1 speakers’ production of Arabic pharyngeal and pharyngealized consonants (/h, ʕ/ and /tˤ, sˤ, dˤ, ðˤ/, respectively). Central to this study is how English speakers’ proficiency in Arabic affects their ability to produce these sounds differently. In particular, it examines the effect of Arabic proficiency on the production of pharyngealized versus non-pharyngealized consonants; pharyngeals versus non-pharyngeals; pharyngeal versus pharyngealized sounds; intra-category differences among the target consonants; and pharyngeal and pharyngealized consonants which differ in the adjacent vowel. The study adopts the posttest-only control group design in which Arabic learners and non-learners constitute the research groups. This thesis exploits Best’s (1995) perceptual assimilation model (PAM); Flege’s (1995) speech learning model (SLM) and principles of gestural phonology in interpreting the split-plot ANOVA results. Theoretical and practical implications are discussed based on the study results.
Acknowledgement

First and foremost, I am immensely grateful to my co-supervisors: Erik Anonby and Lev Blumenfeld for their time dedication and ongoing, insightful feedback. Without their supervision and support, this thesis would never have existed. I would like to extend sincere thanks to the Defense Committee members: Bethany MacLeod, Carmen Leblanc and Randall Gess. I am grateful to them for accepting to take part in my thesis defense.

This research was completely funded by the Saudi Arabian Cultural Bureau. Special thanks to this sponsoring body for their generosity and supervision.

My parents (Abdulmohsen and Fatima Alwabari); my brothers and sisters; my friends (Zahra Sandal and all other friends) and my fiancée (Abbas Saed) deserve warm thanks for their sincere wishes and emotional support throughout my MA study. I am greatly indebted to them for their encouragement and willingness to help. “Thank you” is never enough.

Special thanks also to Peggy Hartwick and Mike Parker for their technical support in my thesis defense. I am honored to meet both of them. Thanks to the problem solver who has been always ready to help, Joan Grant, the Program Administrator at the School of Linguistics at Carleton University. Last but definitely not least, I am grateful to all my research participants for their cooperation and enthusiasm in being involved in this research.
Abbreviations and symbols

CA—Contrastive analysis hypothesis (Lado, 1957)

CI—Confidence interval

CV—Consonant-vowel sequence

ε —Epsilon

F1—First formant (acoustics)

F2—Second formant (acoustics)

$F_{\text{crit}}$—Critical value of $F$-ratio

$F_{\text{obs}}$—Observed value of $F$-ratio

L1—First language

L2—Second language

LB—Lower bound of confidence interval

LSD—Fisher’s least significance difference

$M$—Mean

$MD$—Mean difference

$Md$—Median

$MS$—Mean square

MSA—Modern Standard Arabic

NL—Native language

$\eta_p^2$—Partial eta squared (effect size statistic)

$p$—Probability coefficient
PAM—Perceptual assimilation model (Best, 1995)

REB—Carleton University’s Research Ethics Board

RM ANOVA—Repeated-measure analysis of variance

SD—Standard deviation

SE—Standard error

SLM—Speech learning model (Flege, 1995)

TL—Target language

UB—Upper bound of confidence interval

VC—Vowel-consonant sequence

χ²—Chi-squared distribution
Table of contents

ABSTRACT ................................................................................................................................. ii

ACKNOWLEDGEMENT ........................................................................................................ iii

ABBREVIATIONS AND SYMBOLS ....................................................................................... iv

TABLE OF CONTENTS ........................................................................................................ vi

LIST OF TABLES ................................................................................................................ xii

LIST OF FIGURES ............................................................................................................... xiii

LIST OF APPENDICES ......................................................................................................... xiv

CHAPTER I: INTRODUCING THE STUDY .............................................................................. 1

1.1 Background of the study ................................................................................................. 1

1.2 Definition of terms ......................................................................................................... 2

   1.2.1 Primary and secondary articulations ..................................................................... 2

   1.2.2 Pharyngeal and pharyngealized consonants ......................................................... 3

   1.2.3 Arabic examples of pharyngeal and pharyngealized sounds ................................... 4

1.3 Statement of the problem ............................................................................................... 4

1.4 Research questions ......................................................................................................... 5

1.5 Significance of the study ............................................................................................... 6

1.6 Organization of the study ............................................................................................. 7

CHAPTER II: LITERATURE REVIEW ................................................................................... 8

2.1 Speech learning ............................................................................................................. 8

   2.1.1 Approaches to speech learning research .............................................................. 10
2.1.1.1 The contrastive analysis (CA) hypothesis .......................................................................................... 10
2.1.1.2 Perceptual assimilation model (PAM) .......................................................................................... 12
2.1.1.3 Speech learning model (SLM) ....................................................................................................... 18
2.1.1.4 Summary of PAM and SLM .......................................................................................................... 23
2.2 Studies on Arabic pharyngeal and pharyngealized consonants ................................................................. 25
2.2.1 Non-native Arabic speakers’ perception and production of pharyngealized sounds ......................... 25
2.2.2 Articulatory and acoustic correlates of pharyngeal and pharyngealized sounds ................................. 28
2.2.3 Perception of coarticulation associated with pharyngealization ....................................................... 29
2.2.4 Gender-related differences in the production of pharyngealization ................................................... 29
2.2.5 Unaddressed issues in the study of Arabic sound learning ............................................................... 30
2.3 Methodological considerations in studying sound learning .................................................................... 31
2.3.1 Stimuli construction in sound perception and production research ..................................................... 31
2.3.2 Rapid-shadowing paradigm in speech production tests ....................................................................... 33
2.4 Summary .............................................................................................................................................. 34

CHAPTER III: METHODOLOGY .................................................................................................................. 36

3.1 Design of the study .................................................................................................................................. 36
3.1.1. First construct: Recruiting two equivalent groups ............................................................................. 37
3.1.2 Second construct: Introducing the independent variable .................................................................... 38
3.1.3 Third construct: Measuring the effects of the independent variable .................................................. 39
3.2 Selection of participants ......................................................................................................................... 40
3.2.1 Study population ................................................................................................................................ 40
3.2.2 Sampling method ............................................................................................................................... 40
4.2.3 Question I: English L1 speakers’ production of pharyngeal and pharyngealized consonants, and their counterparts................................................................................................................................. 62

4.2.3.1 English speakers’ production of pharyngealized consonants and their plain counterparts... 62

4.2.3.2 English speakers’ production of pharyngeal consonants and their counterparts.................. 67

4.2.4 Question II: Differences in Arabic learners’ and non-learners’ production of pharyngeal versus pharyngealized consonants............................................................................................................................................ 71

4.2.4.1 First: Homogeneity tests of variances in participants’ scores for pharyngeal and pharyngealized categories............................................................................................................................................. 71

4.2.4.2 Second: Omnibus tests of within-subjects and between-subjects main effects............... 72

4.2.4.3 Third: Interaction between Arabic proficiency and sound production.......................... 72

4.2.4.4 Fourth: Post-hoc test of pharyngeal and pharyngealized sounds’ differences................. 73

4.2.4.5 Summary of group differences in pharyngeal versus pharyngealized CVs......................... 74

4.2.5 Question III: Intra-category differences in learners’ and non-learners’ sound production...... 74

4.2.5.1 Group differences in the production of /h/ versus /ʕ/................................................................. 75

4.2.5.2 Group differences in the production of /tˤ/ vs. /sˤ/ vs. /dˤ/ vs. /ðˤ/................................. 80

4.2.6 Question IV: Differences in participants’ production of pharyngeal and pharyngealized consonants with different vowel combinations................................................................................................................................... 84

4.2.6.1 Group differences in the production of pharyngealized CVs with different vowels........... 84

4.2.6.2 Group differences in the production of pharyngeal CVs with different vowels.................. 89

4.3 Non-native Arabic speakers’ production patterns................................................................... 92

4.3.1 Arabic non-learners’ sound responses to pharyngealized stimuli /tˤ, sˤ, dˤ, ðˤ/..................... 92

4.3.1.1 Non-learners’ shadowing responses to /tˤ/............................................................... 93

4.3.1.2 Non-learners’ shadowing responses to /sˤ/............................................................ 93
4.3.1.3 Non-learners’ shadowing responses to /dˤ/ ................................................................. 94
4.3.1.4 Non-learners’ shadowing responses to /θˤ/ ............................................................... 94
4.3.2 Arabic non-learners’ sound responses to pharyngeal stimuli /h, ʕ/........................................ 95
4.3.2.1 Non-learners’ shadowing responses to /h/ ................................................................. 95
4.3.2.2 Non-learners’ shadowing responses to /ʕ/ ................................................................. 95
4.3.3 Arabic learners’ sound responses to pharyngeal stimuli /h, ʕ/ ........................................... 96
4.3.4 Arabic learners’ sound responses to pharyngealized stimuli /tˤ, sˤ, dˤ, ðˤ/ ......................... 96
4.3.4.1 Learners’ shadowing responses to /tˤ/ ................................................................. 96
4.3.4.2 Learners’ shadowing responses to /sˤ/ ................................................................. 97
4.3.4.3 Learners’ shadowing responses to /dˤ/ ................................................................. 97
4.3.4.4 Learners’ shadowing responses to /ðˤ/ ................................................................. 98
4.4 Chapter summary ............................................................................................................. 98
4.4.1 Summary of split-plot ANOVA findings for the four research questions .................... 98
4.4.2 Summary of learners’ and non-learners’ shadowing responses to pharyngeal and pharyngealized stimuli ............................................................................................................. 100

CHAPTER V: DISCUSSION AND IMPLICATIONS ...................................................................... 102
5.1 Overview of the study .................................................................................................... 102
5.2 Discussion of the results .............................................................................................. 103
5.2.1 Production of pharyngeal and pharyngealized consonants, and their counterparts .... 103
5.2.1.1 Production of pharyngealized consonants and their counterparts ....................... 104
5.2.1.2 Production of pharyngeal consonants and their counterparts ............................... 108
5.2.2 Production of pharyngeal versus pharyngealized consonants ................................ 112
5.2.2.1 Explanations of production differences among the learners ........................................ 112
5.2.2.2 Explanations of production differences among the learners and non-learners ............ 113
5.2.2.3 Sound learnability: Pharyngeal and pharyngealized consonants ................................ 113
5.2.3 Intra-category differences among pharyngealized as well as pharyngeal CVs ............ 114
   5.2.3.1 Intra-category differences between the two pharyngeals .................................... 114
   5.2.3.2 Intra-category differences among pharyngealized sounds .................................... 116
5.2.4 Vowel effect on the production of pharyngeal and pharyngealized CVs ....................... 119
   5.2.4.1 Vowel effects on pharyngealized consonants’ production .................................... 119
   5.2.4.2 Vowel effects on pharyngeal consonants’ production .......................................... 121
5.3 Implications for Arabic sound learning ......................................................................... 122
   5.3.1 Theoretical implications ......................................................................................... 122
   5.3.2 Practical implications ............................................................................................ 124
5.4 Conclusions of the study .............................................................................................. 126
   5.5.1 Consonant production differences ....................................................................... 127
   5.5.2 Group differences: Effects of speakers’ Arabic proficiency .................................. 128
5.5 Limitations and recommendations for future research ................................................ 129

REFERENCES .................................................................................................................. 132

APPENDICES .................................................................................................................. 140
List of tables

Table 3.1: Descriptive statistics of participants’ age and number of spoken languages ............................... 42
Table 3.2: Group differences in age and number of spoken languages ................................................................. 43
Table 3.3: Learners’ Arabic-speaking experience: Age at which they started speaking Arabic, until when, and for how long ............................................................................................................. 44
Table 3.4: Cronbach’s alpha results of internal consistency for each language participants learned .......... 46
Table 4.1: Descriptive statistics of the within-subjects factors (sound categories) .............................................. 58
Table 4.2: RM ANOVA summary of the within-subjects factor (pharyngealized and plain sounds) and its interaction with Arabic proficiency ........................................................................................................ 64
Table 4.3: RM ANOVA summary of the within-subjects factor (pharyngeals and their counterparts) and its interaction with Arabic proficiency ..................................................................................................... 68
Table 4.4: RM ANOVA summary of within-subjects factor (pharyngeal and pharyngealized sounds) and its interaction with Arabic proficiency ........................................................................................................ 72
Table 4.5: Levene’s test of equality of error variances of /ħ/ and /ʕ/ ........................................................................ 76
Table 4.6: RM ANOVA summary of the within-subjects factor (/ħ/ and /ʕ/) and its interaction with Arabic proficiency ........................................................................................................................................ 77
Table 4.7: RM ANOVA summary of the within-subjects factor (/tˤ, sˤ, dˤ, ðˤ/) and its interaction with Arabic proficiency ........................................................................................................................................ 81
Table 4.8: RM ANOVA summary of within-subjects factor (pharyngealized consonants with different vowels) and its interaction with Arabic proficiency ........................................................................................................ 87
Table 4.9: Summary of non-learners’ responses to pharyngealized stimuli /tˤ, sˤ, dˤ, ðˤ/ ........................................ 93
Table 4.10: Summary of non-learners’ responses to pharyngeal stimuli /ħ, ʕ/ ......................................................... 95
Table 4.11: Summary of learners’ responses to pharyngealized stimuli /tˤ, sˤ, dˤ, ðˤ/ ............................................. 97
List of figures

Figure 1.1: Scale of articulation strictures ................................................................. 3

Figure 4.1: Group differences in the production of pharyngealized and plain consonants .................. 65

Figure 4.2: Group differences in the production of pharyngeal consonants and their counterparts .... 69

Figure 4.3: Group differences in the production of pharyngealized consonants and pharyngeal consonants........................................................................................................ 74

Figure 4.4: Group differences in the production of pharyngeal consonants /h/ and /ʕ/ ....................... 79

Figure 4.5: Group differences in the production of pharyngealized consonants /tˤ, sˤ, dˤ, ðˤ/ ............. 83

Figure 4.6: Group differences in the production of pharyngealized consonants with different vowel combinations ....................................................................................................................... 88

Figure 4.7: Group differences in the production of pharyngeal consonants with different vowel combinations ....................................................................................................................... 91
List of appendices

APPENDIX A

Appendix A1: Arabic consonantal phonemic chart

Appendix A2: Hierarchy of the target sounds (Arabic pharyngeal and pharyngealized consonants) and their plain counterparts

APPENDIX B: Language questionnaire

APPENDIX C: Matrix of the stimuli illustrating the three cohorts of each consonant and the random order in which they were presented to participants

APPENDIX D: Inter-rater reliability test (Kappa) results for measuring the internal consistency between the two Arabic perceivers’ assessments of the shadowed sounds

APPENDIX E: Research ethics approval from Carleton University’s Research Ethics Board (REB)

APPENDIX F: Samples of research recruitment advertisement: recruitment cards and posters

APPENDIX G: The consent form that participants signed in compliance with research ethics protocols

APPENDIX H

Appendix H1: Descriptive statistics of groups’ scores for pharyngealized and plain consonants

Appendix H2: Descriptive statistics of both groups’ scores for pharyngeal consonants and their counterparts

Appendix H3: Descriptive statistics of groups’ scores for pharyngealized and pharyngeal consonant

Appendix H4: Descriptive statistics of both groups’ scores for pharyngeal phonemes /h, ṣ/

Appendix H5: Multivariate tests output for the within-subjects factor (/tˤ, sˤ, dˤ, ðˤ/) and its interaction with Arabic proficiency

Appendix H6: LSD test comparing participants’ production of /tˤ/ vs. /sˤ/ vs. /dˤ/ vs. /ðˤ/
Appendix H7: Descriptive statistics of both groups’ scores for pharyngealized consonants

Appendix H8: Post-hoc test (LSD) comparing participants’ production of pharyngealized consonants with varying vowels

Appendix H9: Descriptive statistics of both groups’ production scores for pharyngealized CVs with different vowels

Appendix H10: Descriptive statistics of both groups’ production scores for the pharyngeal CVs with different vowels: /ħaː, ʕaː/; /hiː, ʕiː/ and /łuː, ʕuː/

Appendix H11: Post-hoc test (LSD) comparing participants’ production of pharyngeal CVs varying in adjacent vowels

Appendix I1: Matrix of Arabic non-learners’ production of pharyngealized consonants

Appendix I2: Matrix of Arabic non-learners’ production of pharyngeal consonants /h, ʕ/

Appendix I3: Matrix of Arabic learners’ production of pharyngeal consonants /h, ʕ/

Appendix I4: Matrix of Arabic learners’ production of pharyngealized consonants
Chapter I
Introducing the study

1.1 Background of the study

Non-native speech learning is an interdisciplinary inquiry that emphasizes the integrative role of various disciplines including, but not exclusively, linguistics, applied linguistics and education. Research in this field has focused on the complex interplay of the relation between phonology of learners’ native language (NL) and target language (TL), and learner-related factors such as age and linguistic experience. The vast majority of research in these areas is grounded on the theoretical constructs of cross-language speech learning models and theories that have dominated the field for a few decades. These theoretical constructs include Lado’s (1957) contrastive analysis hypothesis (CA); Best’s (1995) perceptual assimilation model (PAM); and Flege’s (1995) speech learning model (SLM). Although they differ radically in their approach to cross-language sound learning, central to all of them is the relation between NL and TL phonology.

This thesis exploits both PAM and SLM in addition to the principles of gestural phonology in investigating the learnability of certain Arabic consonantal segments. In particular, it probes non-native Arabic speakers’ production of Arabic pharyngeal (i.e. /h, ʕ/) and pharyngealized consonants (i.e. /tˤ, sˤ, dˤ, ðˤ/) based on the relation between Arabic and NL phonology.
The importance of the study lies on the fact that Arabic language has gained a marginal
attention in the field of cross-language sound learning. In particular, researchers have not tested
the newly emerged models—such as PAM and SLM—with one of the least common segmental
features of the world languages, the pharyngeal feature. Also, in the field of Arabic sound
learning in particular, this thesis is the first attempt to examine whether gestural phonology can
account for the learnability of the pharyngeal features in Arabic among non-native Arabic
speakers. In a broader context, the study provides insights to reconcile the theories of linguistics
with the practices of applied linguistics.

1.2 Definition of terms

1.4.1 Primary and secondary articulations.

Essential to the differentiation between pharyngeal and pharyngealized consonants is the
difference between primary and secondary articulations. The vast majority of speech sounds are
produced at one place of articulation such as the lips in [b, p]; the velum in [k]; and uvula in [q,
ʁ]. Some consonants, however, are produced at two simultaneous places of articulation, one of
which has a wider approximation between the articulators than the other. To illustrate, one of
these places of articulations has a stricture that is higher in the scale of strictures shown in Figure
1.1. For example, a double articulation is harnessed in the production of [sʷ] in *sue*. One is a
narrow constriction between the tongue tip and the alveolar ridge which results in producing the
voiceless fricative. The other is a wide approximation between the lips which results in the
rounding feature. The stricture that is narrower (i.e. the alveolar stricture in this case) is the
primary articulation, and the wider (i.e. the labial in [sʷ]) is the secondary articulation. This
secondary articulation, called labialization, is one of four types of double articulation, namely,
palatalization, velarization and pharyngealization (Ashby, 2011; MacKay, 1993).
1.4.2 Pharyngeal and pharyngealized consonants.

The difference between pharyngeal consonants and pharyngealized ones is that the former category requires only one place of articulation while the latter requires a double articulation. The pharynx is a common place of articulation for both types; however, while it is the only (primary) place of articulation for pharyngeals, it is the secondary articulation for the pharyngealized consonants whose primary articulation is in the anterior vocal tract. The pharyngeals’ stricture is narrower and more extreme than that of the pharyngealized sounds. Despite this distinction, both are produced with qualitatively the same constriction in which the root of the tongue approaches the vertical back wall of the pharynx. The vocal tract is thus split into two cavities: the pharyngeal cavity formed between the tip of the glottis and the stricture, and the oral cavity above the constriction (see Delattre, 1971; Esling, 1999; and Laufer et. al., 1988 for a detailed description of articulation).

Figure 1.1: Scale of articulation strictures
1.4.3 Arabic examples of pharyngeal and pharyngealized sounds.

These sounds characterize the Semitic languages including Arabic which has two pharyngeal and four pharyngealized consonants. Arabic pharyngeals are both fricatives; voiceless /ħ/ whose non-pharyngeal counterpart is the glottal /h/; and voiced fricative pharyngeal /ʕ/ whose non-pharyngeal counterpart is the glottal stop /ʔ/ (Hoberman, 1985, pp. 227). All Arabic pharyngealized segments have, in addition to the secondary constriction at the pharynx, a coronal constriction either at the alveolar ridge (as in /tˤ, dˤ, sˤ/) or at the teeth (as in /ðˤ/). Each of these pharyngealized consonants (commonly called “emphatics”; Laufer et. al., 1988) have a plain (i.e. non-pharyngealized) counterpart which requires only the coronal approximation. Hence, the pharyngealized sounds and their counterparts constitute four contrastive pairs: /t, tˤ/; /s, sˤ/ /d, dˤ/ and /ð, ðˤ/. Appendix A1 presents the phonemes of Modern Standard Arabic (MSA) including the pharyngeal and pharyngealized phonemes (see Watson, 2002; and Ryding, 2005 for details on Arabic phonology). These pharyngeal and pharyngealized consonants are the target sounds in this thesis.

1.3 Statement of the problem

The vast majority of research in Arabic pharyngeal and pharyngealized segments has focused on the acoustic features of pharyngeal consonants (Alwan, 1989; Jongman, Herd, Al-Masri, Sereno & Combest, 2011); their articulatory mechanisms (Esling, 1999; Laufer & Baer, 1988; Delattre, 1971; Al-Tamimi, Alzoubi & Tarawnah, 2009); perceptual and acoustic evidence of the coarticulation of pharyngealization (Ali & Daniloff, 1974); and cross-language perception of pharyngealized fricatives (Alosh, 1987; Zahid, 1996). This thesis forms a departure from these recurrent themes in the sense that it explores the production of both pharyngeal and
pharyngealized segments in light of the theories of non-native speech learning. The study compares these two sound categories in terms of non-native speakers’ potential for learning them and, in particular, producing them. It also accounts for the effects of several factors—such as vowel change and intra-category differences—on non-native production of these sounds. Furthermore, it investigates the developmental production differences between (i.e. the effect of Arabic input quantity on) these two categories by examining the production differences across two distinct stages of Arabic learning—that is, the production ability among beginning and advanced learners of Arabic. To achieve this, the study used a posttest-only control group design in parallel with the rapid-shadowing paradigm. Sound production was tested among speakers of English as a first language (L1), both advanced learners and non-learners of Arabic.

1.4 Research questions

In order to address the aforementioned issues (in Section 1.3), the study examines four research questions, some of which are divided into two sub-questions:

I. Do English speakers who vary in their Arabic proficiency produce pharyngeal and pharyngealized consonants as accurately as the counterparts of these sound categories?
   A. Testing production differences between pharyngealized sounds and their plain counterparts.
   B. Testing production differences between pharyngeal and non-pharyngeal sounds.

II. Are there production differences between pharyngeal and pharyngealized consonants?
   III. Are there intra-category production differences among these consonants?
A. Testing production between the pharyngeals /h/ and /ʕ/.

B. Testing production differences among the pharyngealized sounds /tˤ, sˤ, dˤ, ðˤ/.

IV. Does vowel quality affect the production of these sounds?

A. Examining production differences among pharyngealized CVs with varying vowels: /tˤaː, sˤaː, dˤaː, ðˤaː/ versus /tˤiː, sˤiː, dˤiː, ðˤiː/ versus /tˤuː, sˤuː, dˤuː, ðˤuː/.

B. Examining production differences among pharyngeal CVs with varying vowels: /haː, ʕaː/ versus /hiː, ʕiː/ versus /huː, ʕuː/.

Importantly, all these questions and sub-questions address English L1 speakers’ production of the sounds under investigation, as evidenced by Arabic L1 listeners’ perception. Also, the study compares these production differences across two groups of English L1 speakers with different proficiency levels in Arabic. Appendix A2 illustrates the hierarchical distribution of the sounds that has been addressed in the aforementioned questions.

1.5 Significance of the study

This study contributes to the field of non-native sound learning, in general, and Arabic learning in particular. It seeks to fill a huge gap in non-native sound learning research which has long focused on languages such as English, German, Italian and Spanish, and marginalized non-European languages such as Arabic. This research in Arabic sound learning not only extends the array of languages examined in this field, but it also provides insights for the enrichment of the learning models (particularly SLM) based on evidence of one of the least common—and perhaps most cross-linguistically challenging—speech sounds. For instance, the study argues for what constitutes the primitive of sound learning; offers new dimensions for assessing the relation
between NL and TL (i.e. articulatory spatial proximity and gesture dynamics); and predicts the shape of the relationship between NL and TL (i.e. whether it is a linear or a U-shaped relationship).

The significance of the study is not restricted to the theories of this discipline; rather, it offers several tools for the practitioners of Arabic sound learning. In particular, it assists learners and teachers in predicting areas of learning difficulty and helps curriculum developers in sequencing the content of Arabic phonology curricula.

1.6 Organization of the study

I have organized this study in five chapters, including this introductory chapter. Chapter II presents the theoretical framework and reviews relevant literature in Arabic sounds. This is followed by Chapter III which describes the methodology of the study including its design, participants, instrumentation, and a description data collection and analysis procedures. Chapter IV is divided into two main sections. The first is a statistical analysis of the four research questions using Arabic perceivers’ numerical assessment of the sounds that English speakers produced. The second section discusses the most frequent sounds that Arabic perceivers’ transcribe in response to English speakers’ sounds. I have consolidated these two sections in the discussion of the results in Chapter V which also includes the implications and limitations of the study, several recommendations for future research, and final conclusion of the study.
Chapter II

Literature Review

This chapter aims to provide a solid ground of the present study by reviewing relevant research in different areas. These areas include research in second language (L2) speech learning, studies exploring Arabic pharyngeal and pharyngealized consonants, and studies related to the methodologies employed in speech perception and production experiments.

Thus, the chapter is organized in three main sections, and each tackles one of these domains. The discussion of speech learning research is presented according to the chronological order in which speech learning models and hypotheses emerged, with concrete examples from studies that exploited these models and hypotheses. The discussion of the second section, on the other hand, is organized in terms of the most common research threads in the perception and production of Arabic pharyngeal and pharyngealized consonants. The last section addresses two issues pertaining to stimuli construction as well as test design in speech perception and production experiments.

2.1 Speech learning

The term “speech learning”—coined by Flege and Port (1981)—refers to the processes in which language learners “articulate or perceive a speech sound differently after (as compared to
before) massive exposure to a foreign language [italics added]” (Flege, 1988, p.226). This terminology denotes two properties that are essential to my study.

First, as stated in Flege’s (1988) definition, speech learning encompasses two processes that function collaboratively in enhancing sound learning: speech perception and production. In support of this, James (1988, p.30) asserts that learning new sounds involves learning “new patterns of articulation and perception [italics added]”. These two processes (i.e. perception and production) are indispensably inseparable; however, there is a lack of consensus among scholars on which one precedes the other. In contrast to what Sheldon (1982) and Brière (1966) argue for, some researchers (such as Lenneberg, 1962) provide evidence that perception is a prerequisite to production which emerges later in language learning. In the same vein, the Motor Theory of speech perception (Liberman, Harris, Hoffman & Griffith, 1957, cited in Carroll, 2008) postulates that listeners use implicit articulatory knowledge to aid perception. In other words, they perceive sounds by detecting the gestural movements involved in sound production. The present study is concerned primarily with sound production; nonetheless, both perception and production integrally formed the core of the study, and from this fact stems the use of the term speech learning.

Another characteristic of the term speech learning that is fundamental to this research is that it differentiates between speech acquisition and learning. Applied linguistics researchers draw a clear distinction between these two concepts. They ascertain that while acquisition is an exposure to language models without formal teaching, learning is a conscious process that involves breaking down language content into its analytic parts so that the learner gains a meta-knowledge about the language. In practice, the former concept is often associated with acquiring L1 while latter is associated with learning non-native languages (Gee, 1996).
These properties of speech learning are essential to this study in the sense that, although its primary focus is sound production, it encompasses the two processes: perception and production. In addition, the terminology is suitable to the context of the study given the fact that it is concerned with speaker’s ability to learn non-native sounds, not to acquire them. The study, nevertheless, probes the perception and production of segmented speech sounds rather than speech in general.

2.1.1 Approaches to speech learning research.

2.1.1.1 The contrastive analysis (CA) hypothesis.

Among the early approaches to segmental sound learning is Lado’s (1957) Contrastive Analysis Hypothesis (CA) which is based on structural comparison between the native language (NL) sound system to that of the target language (TL). Three elements, according to Lado, are instrumental in this comparison: phonemic inventory, allophonic membership within these phonemes, and positional distribution of phonemes within the language.

The first criterion for determining phonological similarity is comparing the phonemic inventory of the NL and TL sound systems which entails identifying the target phoneme in the TL and a phoneme that it is perceptually analogous to it in the NL. If the TL phoneme fails to satisfy this condition, TL learner may encounter difficulty in learning this particular sound.

The phonemic inventory is not the only element that is predictive of potential learning difficulties; rather, the allocation of all its allophones in both languages is more persistent. Lado (1957) provides a concrete example of the phoneme /d/ that exists in both English and Spanish; however, the two languages differ in their allophonic distribution. That is, while Spanish has two allophones for the same phoneme, English splits these allophones into different phonemes /d, ð/.
This phenomenon has been latter called “phonemic split” (Eckman & Iverson, 1997, pp. 192-193) and researchers argue that it induces the maximal learning difficulty.

In addition, according to CA, phonological rules that govern the distribution of these allophones may also induce learning difficulty if proven to be different in the NL and TL. To illustrate, both English and French have the fricative /ʒ/ and all its allophones; nonetheless, this consonants can occur in initial position in French (for example, jamais) but not in English. This distributional constraint causes English learners of French to transfer this phonological rule and, thus, encounter a difficulty learning /ʒ/ in initial position (Lado, 1957, pp. 12-17). By employing Lado’s structural comparison of sound systems in terms of these three elements, potential sound learning difficulties are predictable and consistent with the type of difference between these sound systems.

Exploiting Lado’s hypothesis, two opposing streams of research dominate the field of L2 speech learning at that time. One stream of scholars (such as Brière, 1966; Eckman et. al., 1997; Flege & Port, 1981; Gass & Selinker, 2001) argues that language learners produce TL sounds that are similar to the NL more easily and authentically than dissimilar sounds. In contrast, some studies including Major (1994) provide evidence that learning a TL sound that does not have an equivalent in the NL incurs less difficulty than sounds that are similar to NL phoneme inventory.

However, in 1960s and 1970s, claims about CA inadequacy began to arise (Major, 1994, pp. 181-204). Namely, Brière (1966), which used a composite language combining Arabic, Vietnamese, and French utterances as a TL learned by monolingual speakers of American English, presented contradicting results. For instance, some TL sounds which are divergent from participants’ NL (for example, /x/) are learned quite rapidly, whereas other sounds are learned in
accordance to CA hypothesis. Hence, as evidenced by Brière (1966) and others, the structuralist comparison proposed by Lado (1957) cannot account for all speech learning patterns. Therefore, new approaches to the study of non-native speech learning has emerged; for example, the Motor Theory of speech perception (Liberman et. al., 1957); the psychoacoustic approach (Jusczyk, 1993, cited in Best, 1995); the perceptual assimilation model (Best, 1995); and the speech learning model (Flege, 1995). I will elaborate on the latter two models only because they directly address non-native speech learning (i.e. the Motor Theory and the psychoacoustic approach tackle speech perception in general rather than cross-language speech perception which is relevant to the present study).

2.1.1.2 Perceptual assimilation model (PAM).

The basic premise of the perceptual assimilation model, developed by Best (1995), is that perception of non-native segments is determined according to the segments’ similarity to (or difference from) the closest NL segment. This relation between TL segment and the native segment is defined articulatorily according to two basic principles: the universal phonetic domain and native phonological space. The latter is exemplified in a model of gestural phonology developed by Browman and Goldstein (1986, 1989, 1990a, 1990b). Hence, it is instrumental to explain the universal phonetic domain and the native phonological space in order to better judge the similarities and differences between languages in general, and Arabic and English in particular.

Universal phonetic domain and native phonological space.

The phonetic gesture, the basic unit of the universal phonetic domain, is the coordinated formation and release of a varying-degree constriction that is harnessed for sound production and
formed along the vocal tract (Browman et. al., 1990a). The phonetic gesture is also defined in terms of motion *dynamics*, the spatial and temporal representation of the articulatory motion. An example of a simple phonetic gesture is lowering the velum in the production of the nasal consonant /n/.

The multidimensional, universal phonetic domain defines the “biodynamic constraints” of the vocal tract (Best, 1995). To illustrate, this domain is the summation of all spatial and temporal properties of *all* possible phonetic gestures that are harnessed in the world’s languages. Hence, this domain can range from the lips—as most of the world’s languages have bilabial consonants (Ladefoged & Maddieson, 1990)—to the glottis. Although only a few languages make use of a posterior, glottal constriction (such as in Arabic /ʔ/ and in ejective consonants /p’, t’, k’, s’, Ashby, 2011), the universal phonetic domain encompasses the glottis.

A native phonological space, on the other hand, is a portion of the universal phonetic domain that a language uses for its own sound articulations. This space, which is language-specific, is determined according to four aspects which form the benchmark for gauging cross-language similarities and differences. I discuss these determinants in relation to Arabic and English phonological spaces.

The first is the spatial dimension along the tube geometry (i.e. vocal tract) that encompasses all locations of a given language’s gestures. For instance, since Arabic has a bilabial plosive /b/ and pharyngeal consonants /h, ḫ/, it can be said that its spatial domain ranges from the lips to the pharynx. These spatial boundaries encompass all possible gesture locations that fall between these two extremes (i.e. the lips and the pharynx) regardless of whether a gesture location is actually harnessed in Arabic or not. Importantly, since American English has
a pharyngeal constriction in the production of /t/ (Delattre, 1971), it can be assumed that the pharynx falls within its native phonological space. Thus, Arabic and English overlap in their circumscription of the tube geometry. In other words, their native phonological spaces are similar in their spatial dimension.

The second aspect is the spatiotemporal invariants of language gestures (i.e. the space and time parameters of a language’s gestures). For example, as the voice-onset time of English /t/ is different from that of French /t/ (Whalen, Levitt & Goldstein, 2007), the temporal parameter of the English gesture is assumed to differ from the French one. Another example or instance, the production of Arabic /sˤ/ necessitates special timing of the formation and release of the pharyngeal constriction relative to the coronal constriction (Laufer et. al., 1988). This temporal feature of Arabic gestures is quite different from the English gesture used in articulating /s/ whose coronal constriction is independent on (and unrelated to) any pharyngeal gesture.

The degrees of constriction of the articulatory gestures represent the third aspect of native phonological space, and refer to the degree of proximity between the active and passive articulators involved in gesture formation. There are varying degrees of constriction in segmental production—for example, closed (as in plosives); critical (as in voiced fricatives); narrow (as in approximants); mid (as in most vowels); and wide such as the opening of the glottis in the production of voiceless stops (Ashby, 2011; and Best, 1995). Although both Arabic and English have a pharyngeal constriction, the degrees of constriction for Arabic pharyngeal and pharyngealized consonants are different from the degree of pharyngeal constriction for English pharyngeal /a/. Namely, the pharyngeal constriction for Arabic /h, s/ is narrower than the mid vowel constriction for /a/ (see Laufer and Baer, 1988 for detailed elaboration on pharyngeal degree of constriction in Arabic). This narrow constriction harnessed in Arabic consonants
cannot be incorporated within English phonological space; therefore, it could be a source of sound learning difficulty.

The assembly of multiple simple gestures in the production of a single utterance—called “gestural constellation” (Browman et. al., 1990a)—is the fourth aspect that characterizes a native phonological space. An example of how languages differ in their gestural constellations is evident in the case of Arabic pharyngealized consonants. Although American English and Arabic have the pharynx as a common spatial invariant in the production of English /a/ and Arabic /tˤ, sˤ, dˤ, ḏˤ/, the two languages have discrepant combinations of the simple gestures required in producing these sounds. To illustrate, multiple gestures coordinate together in articulating Arabic /sˤ/. These gestures include forming a median depression (groove) along the upper surface of the tongue; raising the tongue blade to touch the alveolar ridge (anterior, close constriction); retracting the root of the tongue to approach the back wall of the pharynx; elevating the hyoid bone and raising the larynx. The combination of these simple gestures is not harnessed in the articulation of English /a/ which requires smooth tongue surface (not grooved) and which does not require a coronal stricture. Therefore, the gestural constellation of Arabic /sˤ/ does not fall within English phonological space and, consequently, it can be a source of difficulty for English learners of Arabic.

To summarize, a native phonological space is a portion of the universal phonetic domain to which the respective language circumscribes for its phonetic functions. As the name implies, the native phonological space of a given language differs from that of other languages although there are some commonalities between different phonological spaces. These differences are judged in terms of the spatial boundaries, the spatiotemporal properties of phonetic gestures, the degrees of constriction, and gesture constellations of the native phonological spaces in question.
These determinants of native phonological space are the basis of my interpretations of the present study findings. That is, based on the differences between Arabic and English phonological spaces, I explain English speakers’ production of Arabic sounds which do not belong to English phonological space.

In general, these parameters of judging similarities and differences between languages (i.e. the spatial and dynamic characteristics of gestures) are central to the tenets of PAM.

*Perceptual patterns of PAM.*

Best’s (1995) PAM rests on the fact that perceivers are attuned to the native segments of their NL and that they perceive non-native segments according to their gestural similarity to, and difference from, the native segments. To paraphrase, attunement to native segments is the increased detection of phonetic *gestures* produced in the native phonological space of the perceiver’s NL. For example, according to PAM, native Arabic speakers demonstrate, a higher ability in detecting the gestural cues of pharyngeal consonants—because these consonants are produced in Arabic phonological space—than speakers of other languages whose native phonological space does not include the pharyngeal gestures. Perception of non-native segments, on the other hand, is influenced by the perceiver’s NL. Best (1995, pp. 193-195) categorizes the perception of a non-native *contrast* (i.e. a pair of two contrastive segments) into several perceptual patterns in accordance to the principles of the native phonological space:

- **Non-assimilable:** Both segments of the non-native contrast are produced from outside the native phonological space of the perceiver’s NL.
- **Both uncategorizable:** Both non-native segments belong to the native phonological space, but they do not resemble any native category.
• *Uncategorized versus categorized:* One segment of the contrast is assimilated to a native category while the other, although it falls within the native phonological space, does not resemble any native category.

• *Two-category assimilation:* Both non-native segments fall within the phonological space; however, each segment is assimilated to a different native category.

• *Category-goodness difference:* Both non-native segments are assimilated to a single native category; however, one segment is more divergent from the native counterpart than the other segment.

• *Single-category assimilation:* Both non-native segments are equally divergent from (or similar to) the native contrastive segments.

Several empirical studies support these perceptual patterns including studies of Zulu contrastive clicks (Best, McRoberts & Sithole, 1988 cited in Best, 1995); Ethiopian ejective contrast (Best, 1990); English glide contrast (Best & Strange, 1992); and other studies in vocalic contrasts.

In brief, the native phonological space, on which PAM rests, is a portion of the universal phonetic domain that a language delineates for its sounds’ production. This phonological space is defined in accordance to several parameters of native gestures’ spatial and dynamic characteristics. These parameters are fundamental to PAM’s classifications of non-native segmental contrasts. That is, the perception of non-native contrast depends on its segments’ belonging to the native phonological space of the perceiver’s NL, and on the degree of similarity between native and non-native segmental categories. Above all, the perceptual primitive for PAM is the articulatory gesture by which phonetic similarity can be estimated. Several subsequent studies investigate this model on various languages; however, it has not been tested against Arabic, in general, and Arabic pharyngeal and pharyngealized consonants, in particular.
Although the primary concern of this study is the production of these sounds, it accounts for the theoretical basis of PAM. Precisely, I have exploited PAM’s classifications of TL sounds (that are defined based on articulatory information) and related them to the production patterns I observed; however, I have not made judgments on participants’ perception abilities.

2.1.1.3 Speech learning model (SLM).

The speech learning model (SLM), proposed by Flege (1995), is different from PAM in the sense that the former accounts for non-native speech production as being mediated by perception while the latter accounts for non-native speech perception only. Nonetheless, fundamental to both SLM and PAM is the definition of the native phonological space.

Central premise of SLM is the proposal that individuals’ NL phonetic systems “remain adaptive” over time and that they are susceptible to reconfigurations and additions as a result of learning TL sounds. The following discussion of SLM’s hypotheses elaborates on this premise.

First hypothesis (H1).

NL and TL sounds are perceptually related at an allophonic level rather than at a phonemic level, and ability to perceive and discriminate TL sound correlates with its position within the word. This means that the model draws the comparisons between the allophones of the NL and TL, rather than the phonemes. SLM also claims that the discriminability of TL allophones depends on their position within the word—i.e. whether they are in word-initial; -middle; or -final position.

For example, Sheldon and Strange (1982) provides evidence that Japanese learners of English learn the allophones of the English phonemes /l/ and /ɹ/ at different rates according to
their position in the word. That is, these sounds are better perceived and produced in word-final position than in initial position. When the allophones of English /l/, for example, occur in word-initial position, Japanese learners encounter difficulty in perceiving them because the acoustic cues are less robust. These findings are in agreement with H1 in the sense that Japanese learners of English are capable of perceiving the allophones of the English phoneme /l/, whereas other allophones of this phoneme pose perceptual difficulty.

Second hypothesis (H2).

A new phonetic category can be configured to represent TL sounds that differ from the NL sound perceptually. An earlier study by Lively and Logan (1990) tests native Japanese learners of English in their perception of English liquids /ɹ/ and /l/. The researchers have found evidence supporting the claim that learners’ “perceptual space” can be extended or squeezed in response to the TL sounds. Presumably, this perceptual space is analogous to the phonetic category that Flege (1995) predicts (i.e. both the perceptual space and the phonetic category are abstractions of how sounds are perceptually represented). Furthermore, in support to H2, Flege, Takagi and Mann (1995b, cited in Flege, 1995) argues that native Japanese learners of English may create a new phonetic category that represents English /ɹ/ rather than /l/. The reason is that, according to the study findings, the English liquid /ɹ/ is perceptually distinct from its Japanese counterpart /r/ while the other liquid /l/ is not.

Third hypothesis (H3).

The ability to discern the phonetic differences between a TL sound and the closest NL sound increases as the divergence (i.e. difference) between the sounds increases. This hypothesis is founded on the fact that Japanese /r/ is closer to the English liquid /l/ than to the other English
liquid (i.e. /s/); therefore, Japanese learners of English are better in discriminating Japanese /r/ and English /s/ (which are less similar to each other) than in discriminating Japanese /r/ and English /l/ (which are more similar to each other).

**Fourth hypothesis (H4).**

The ability to discern phonetic differences between TL and NL sounds, and between TL sounds that are not contrastive in the NL, correlates negatively with the age of learning. Evidence from Baker and Trofimovich (2005)—which compares vowel perception and production between early and late bilinguals of Korean and English—shows that learner’s age has an impact on the degree of interaction between the phonetic systems of the two languages. That is, early bilinguals demonstrate a higher degree of the discriminability between NL and TL vowels than late bilinguals. The study also shows that the direction of the interaction between NL and TL phonetic systems is influenced by the age of learning, with younger bilinguals particularly exhibiting a bidirectional interaction. A similar study (Flege, Mackay & Meador, 1999) concludes that the perception and production abilities deteriorate as the age of learner increases. The findings of Baker et. al. (2005) and Flege et. al. (1999) are consistent with the hypothesis of SLM.

**Fifth hypothesis (H5).**

The model predicts that the mechanism of *equivalence classification* may prevent new phonetic categories for TL sounds from being established. This means that the increased phonetic similarity (i.e. equivalence) between TL and NL sounds may prevent TL learners from creating new phonetic categories for the TL sound. If so, a single phonetic category will be harnessed in perceiving TL and NL similar sounds. Consequently, the learner (who is unable to
discern the perceptual differences between NL and TL sounds) will produce these sounds in a similar way.

*Sixth Hypothesis (H6).*

In certain circumstances, the new phonetic category that a learner establishes to perceptually process TL sounds may differ from the native speakers’ phonetic category of the same sound. This occurs, according to SLM, when the learner creates the category based on perceptual features different from the features perceived by native speakers. This hypothesis leads Flege to predict that although TL learners do create new phonetic categories for TL sounds, this does *not* necessarily permit them to produce the TL sounds as authentically as the native speakers do (Flege, 1995, pp. 243). In other circumstances, a learner’s phonetic category diverges from the native speaker’s category in order to maintain contrast between NL and TL categories that have the same phonological space. In other words, if learners’ NL phonetic category belongs to the same phonological space of the TL category, this may lead the TL phonetic category to deviate from the *typical* TL category which native speakers of TL establish.

Flege’s hypothesis is based on evidence found in Lindblom (1990b, cited in Flege, 1995) which shows that learners’ production of a TL vowel diverges from their production of the NL vowel although both NL and TL vowels occur in the same phonological space. Flege (1995) infers that learners do so in order to maintain contrast between NL and TL vowels.

*Seventh hypothesis (H7).*

Learners’ production of a TL sound is consistent with the phonetic category which they use to perceive the sound. In other words, their sound production mirrors the phonetic category representation. Consequently, if a learner’s phonetic category of a TL sound concords with
native speakers’ category, learner’s and native speakers’ sound productions will be similar. The model predicts that, if the NL has fewer vowels than the TL, TL learners will establish new phonetic categories to represent the new vowels (according to H2), and they will be able to produce these new vowels in a way consistent with their respective phonetic categories (H7). By inference, since Arabic has fewer vowels than English, Arabic learners of English may establish new phonetic categories to represent “new” English vowels. In the case that these categories match English native speakers’ categories, Arabic learners of English may be able to produce these vowels in the same way as a typical English monolingual speakers’ production.

Of particular importance to the present research are H2, H3, H5, H6 and H7 as I argue that learners of Arabic are able to create new phonetic categories from some (but not all) Arabic consonants. In fact, H7 does not account for other variables that may affect learners’ production of TL sounds. That is, implied in this hypothesis is that perceptual accuracy (as evidenced by new category formation) is the only source for authenticity in TL sound production. There are, however, other sources for TL misarticulation such as motoric difficulties and insufficient auditory input. Furthermore, Waldman, Singh and Hayden (1978) does not find a significant correlation between the ability to perceive and discriminate sounds, on the one hand, and the ability to produce the sounds, on the other hand. Participants’ perception abilities surpass their production ability, which suggests that misarticulation is not necessarily caused by perceptual deficit (and other variables might have caused their misarticulation). The fact that non-native speakers’ production ability is not restricted to perception deficiency is crucial to this thesis as I attribute several production errors to articulatory reasons.
**2.1.1.4 Summary of PAM and SLM.**

Although both PAM and SLM deal with non-native speech perception, there are some commonalities and differences between them in addressing TL sound perception. A common principle shared between PAM and SLM is that the key element in predicting TL perception is the phonetic similarity between TL and NL sounds. However, the models differ radically in predicting TL perception as influenced by the phonetic similarity between TL and NL. To illustrate, SLM predicts a linear relation between phonetic similarity and perceptual attainment of TL sounds (i.e. the ability to form new phonetic categories increases as the perceived similarity increases). Conversely, category formation according to PAM follows a U-shaped pattern. That is, the highest likelihood of establishing new TL categories is predicted to be when the TL and NL sounds are moderately similar, while category formation decreases for TL sounds that are highly similar/dissimilar to NL sounds. Another difference between the two models is that PAM deals with *naïve* TL speech perception (i.e. TL sound perception by non-native speakers who have not been exposed to the TL). SLM, on the other hand, exploits the ultimate perceptual attainment by non-native speakers who have been exposed to a TL perceptual input intensively. In addition, Flege (1995) extends SLM to account for speech production as well.

*Assessing TL and NL phonetic similarity.*

Assessing phonetic similarity—which is the key in predicting sound learnability—is independent on (and essential for) the tests of non-native speakers’ perception of the TL contrast (Strange, 2007). There are various ways for estimating the phonetic similarity between the NL and TL sounds; for instance, it can be assessed according to gestural differences between the sounds in question (PAM), or according to learners’ perception of the similarity (SLM).
Alternatively, evidence on phonetic similarity can be obtained either through quantitative analysis of the acoustic similarity between the TL and NL sounds, or through “perceptual assimilation paradigms” derived from tests of judging the perceptual similarity between TL and NL sounds (Strange, 2007). In fact, in the field of learning Arabic sounds, there is a lack of sufficient empirical evidence on the phonetic similarity between Arabic and English sounds to account for in predicting English speakers’ ability to learn Arabic sounds. I have interpreted English learners’ production of Arabic segments based on inferences from the articulatory description of these sounds. In some instances—such as the comparison between /h/ and /ʕ/ where the articulatory differences cannot justify the production patterns—I have grounded my explanation on acoustic differences between the sounds.

*Current research trends inspired by PAM and SLM.*

The two theoretical constructs, PAM and SLM, have been influential in the field of second language speech learning. An array of issues in L2 speech learning have been empirically investigated in light of PAM and SLM including, but not exclusively, non-segmental speech learning; age-related perception and production patterns; the relation between musical training and L2 speech learning; and segmental speech learning of L2 sounds. The latter, to which my study belongs, is perhaps the most dominant thread of research related to learning L2 vowels and consonants. The studies that exploit these models at the segmental level focus on languages such as Italian (Flege et. al., 1999) and Japanese (Lively et. al., 1990; and Sheldon et. al., 1982). However, Arabic is among the languages that have received less attention in the study of L2 speech learning, in general, and in applying the principles of PAM and SLM on its speech learning, in particular. Therefore, one of the goals of the present study is to extend the field of
inquiry on Arabic speech learning by exploiting PAM and SLM principles in interpreting the learning patterns I observed.

2.2 Studies on Arabic pharyngeal and pharyngealized consonants

There is a scarcity of studies conducted on learning the phonology of Arabic as an additional language, in general, and on learning its pharyngeal and pharyngealized consonants, in particular. The dominant themes in the field of Arabic phonology include phenomena such as the coarticulation effect of pharyngealization, the articulatory mechanisms harnessed in the production of pharyngeal and pharyngealized consonants and, less commonly, gender-related differences in pharyngealization production. A limited number of researchers have examined the perception and production of these consonants among non-native speakers of Arabic. I have divided the subsequent sub-sections according to the major themes of inquiry in the perception and production of Arabic pharyngealized and pharyngeal consonants.

2.2.1 Non-native Arabic speakers’ perception and production of pharyngealized sounds.

Perhaps the only study that (to my knowledge) tests the perception of pharyngealized consonants in terms of Best’s (1995) PAM is Zahid (1996). This study examines the discriminability and identification of a pharyngealized fricative /sˤ/ and its plain counterpart /s/ among native Arabic and French speakers. The author hypothesized that French perceivers’ discrimination would be moderate to very good according to the predictions of PAM’s “category-goodness difference” (Best, 1995). The reasoning for Zahid’s (1996) prediction is that French perceivers would assimilate both members of this pair (i.e. /s/ and /sˤ/) to a single French category /s/. However, French perceivers’ would perceive the two Arabic sounds differently
because /sˤ/ is more divergent from the French sound than the plain Arabic member. The author uses a continuum of synthesized stimuli in which the F2 onset frequency is gradually reduced, so that the stimuli range between /s/ (100 Hz) and /sˤ/ (200 Hz). As the sound deviates from /s/, according to the author’s hypothesis, the perception accuracy would decrease. The study finds evidence of categorical perception of /s, sˤ/ among the Arabic and French perceivers with a significant difference between the two groups in the category boundary. The results also indicate that French speakers’ perception is less accurate than the Arabic speakers’.

I assume that the stimulus in Zahid (1996), although it reveals the perceptual pattern for the /s, sˤ/ continuum, these perception findings might have been induced by the sounds that were between /s/ and /sˤ/ rather than by the authentic pharyngealized /sˤ/. Furthermore, although the difference between groups’ perceptual category boundary is statistically significant, this does not necessarily mean that the French group’s perception is actually poor. Hypothetically, it is likely that French listeners’ perception of /sˤ/ is accurate even if their perception is significantly different from Arabic listeners’ perception. Actually, the difference between both groups’ performance is marginal. In fact, there is not a pre-determined benchmark for judging whether the perception is accurate, moderate or poor.

A more extensive study in cross-language perception and production of pharyngealized consonants is a PhD dissertation by Alosh (1987). The study aims to exploit the difference in perceptual cues between native and non-native speakers of Arabic, and the developmental transition in non-native speakers’ perception of pharyngealized fricatives /sˤ, dˤ/. To illustrate, the perception of pharyngealization could be triggered by either the pharyngealized consonant itself or by the adjacent vowel. This is tested using synthetic sounds in which the coarticulated vowels adjacent to pharyngealized consonants are replaced by plain vowels, in addition to un-
The study finds a difference between native and non-native Arabic speakers’ perceptual cue of pharyngealization (i.e. detecting pharyngealization on the pharyngealized vowel); however, this difference was statistically non-significant. As for the second part of the study, in order to test whether these perceptual cues are proficiency-related, (i.e. the developmental transition in perception of pharyngealization), the researcher divides the American-English speakers into three groups according to their proficiency level. The results show a significant interaction between American-English speakers’ proficiency in Arabic and detecting pharyngealization on the coarticulated vowel. The study also extends to testing the dependency of production upon perception. Although it is not an essential question in Alosh (1987), when interpreting the statistical findings of my study, I found that although the vowel /aː/ results in more accurate perception of pharyngealization, the difference in perceiving the pharyngealization associated with the three vowels /aː, iː, uː/ is statistically non-significant. This indicates that vowel change may not induce any difference perceiving pharyngealized fricatives, a finding that is of a particular relevance to my study.

In general, pharyngealized fricatives have been the focus of these studies on non-native speakers’ perception and production. Other pharyngealized consonants in the Arabic phonemic inventory (such as the pharyngealized plosive consonants) have not been examined. Although it might be predicted that since the articulatory mechanisms for all Arabic pharyngealized consonants are similar, their perception and production would have been also similar; therefore, the findings of these two studies could be generalized to all pharyngealized consonants. In fact, this is not necessarily the case. There might be some intra-category differences between these consonants, and this is one of the goals of the present study. On the other hand, non-native speakers’ learnability of pharyngeal consonants—on which the current study focuses—has
received less attention than pharyngealized consonants. Inquiries into pharyngeal consonants are limited to their articulatory gestures and acoustic characteristics as produced by Arabic native speakers.

2.2.2 Articulatory and acoustic correlates of pharyngeal and pharyngealized sounds.

A commonly cited piece of research is Laufer and Baer (1988) which addresses the articulatory and acoustic properties of pharyngeal /ħ, ʕ/ and pharyngealized /tˤ, sˤ, dˤ, zˤ, q/ in Hebrew and Arabic. The study demonstrates that these consonants are characterized by raising F1 and lowering F2 which extend to the steady state of the adjacent vowel. The authors also conclude that the constriction at the pharynx is qualitatively the same whether it is primary or secondary articulation. However, the degree of constriction in the case of secondary articulation (i.e. pharyngealized consonants) is wider than the degree of constriction for the pharyngeal consonants. This is particularly interesting given the fact that two of the pharyngealized consonants (i.e. /tˤ, dˤ/) are plosives in which the constriction is expected to be narrower than that of the pharyngeal fricatives. The authors do not draw a comparison between fricative pharyngeal and pharyngealized consonants in terms of the degree of constriction; however, they compare between plosive pharyngealized consonants /tˤ, dˤ/ and the fricative pharyngealized ones /zˤ, sˤ/, showing that constriction is narrower for the plosives. It is worth noting here that there was a lack of consensus among scholars of Arabic phonology on labelling /q/ as either a uvular or pharyngeal consonant, yet Laufer et. al. (1988) includes it in the pharyngealized category. Importantly, this study is concerned with only the pharyngeal constriction for both sound categories, and this does not rule out the fact that pharyngealized consonants require also an anterior primary articulation somewhere in the oral cavity. Delattre’s (1971) X-ray results, which encompass Arabic /h, ʕ, x, q/ as pharyngeal phonemes, are also consistent with Laufer et. al.
(1988). Intensive elaborations in this domain of research, the articulatory mechanisms of pharyngeal production in different languages, can be tracked in studies conducted by John Esling and colleagues (for example, Esling, 1999; Esling, Fraser & Harris, 2005; Carlson & Esling, 2003; Zeroual, Esling & Crevier-Buchman, 2008).

2.2.3 Perception of coarticulation associated with pharyngealization.

Another stream of research on pharyngeal and pharyngealized Arabic consonants concerns the coarticulation effect and its influence on perception. In Jongman, Herd, Al-Masri, Sereno and Combest (2011), for example, the coarticulation occurs in both a progressive and regressive manner. This means that both the following and preceding vowels adjacent to a pharyngealized consonant were pharyngealized. This acoustic evidence is aligned with the findings from the perceptual test which reveals that coarticulated vowels help in detecting pharyngealization. This latter finding regarding the perception of pharyngealization as cued by the pharyngealized vowel is also consistent with the findings of Alosh (1987). Evidence for bidirectional coarticulation associated with pharyngealization is also found in Ali and Daniloff (1974). While Jongman et. al. (2011) provides both an acoustic and perceptual evidence of coarticulation, Ali et. al. (1974) examines this phenomenon from a perception perspective only.

2.2.4 Gender-related differences in the production of pharyngealization.

The acoustic enquiry in Khan (1975) reveals gender-related differences in articulating pharyngealized consonants. Female native speakers of Arabic demonstrate less acoustic differentiation between pharyngealized and non-pharyngealized consonants than native male speakers. These gender differences are apparent between Arabic-speaking groups rather than the English-speaking groups.
Evidence for gender-related differences in pharyngealization from an articulatory perspective is derived from Al-Tamimi, Alzoubi and Tarawnah (2009). This study primarily questions the articulatory differences between pharyngealized and plain consonants, rather than gender differences. Nonetheless, when analyzing its findings, I found evident differences between the two genders in the measurement values of three phonetic gestures: hypoid bone elevation, larynx raising and oropharyngeal narrowing. It should be noted that these differences, which might have been induced by the difference in anatomical measures between genders, are not tested for statistical significance.

2.2.5 Unaddressed issues in the study of Arabic sound learning.

To summarize, the most commonly researched phenomena in field of Arabic phonology include the coarticulation of pharyngealization as evidenced by acoustic and perceptual tests, the articulatory gestures harnessed in the production of pharyngealized and pharyngeal consonants, and gender differences in the production of pharyngealization.

Although some researchers explore non-native Arabic speakers’ perception and production of these consonants, their inquiry is limited to pharyngealized fricatives /sˤ, dˤ/ which differ articulatorily from pharyngealized plosives /tˤ, dˤ/. This articulatory difference may lead to differences in non-native speakers’ production of pharyngealized plosives. Similarly, non-native Arabic speakers’ perception and production of pharyngealized plosives and pharyngeals has not been addressed in literature.

In addition, the studies that have examined non-native speakers’ perception and production use synthesized stimuli which is appropriate for addressing either coarticulation (Alosh, 1987) or categorical speech perception (Zahid, 1996). I have also identified some issues
in Zahid’s (1996) experimental design that might account for its findings. The present study takes these issues into consideration and attempts to bridge the gap in the field of Arabic phonology learnability.

2.3 Methodological considerations in studying sound learning

There are, in fact, some commonalities in the methods employed in the research of sound perception and production. These commonalities pertain to, for example, stimuli construction and test design. In this section, I highlight two method-related aspects that are frequently implemented in this type of research: the use of meaningless CV sequences as opposed to meaningful words, and a speech production test design known as rapid-shadowing paradigm.

2.3.1 Stimuli construction in sound perception and production research.

Depending on the exact purpose of the research, some studies employ meaningful words to stimulate participants’ responses whereas others use fragmented speech in form of CV, VC, and so forth. An important psycholinguistic phenomenon may result from specific stimuli constructions and, consequently, interfere with the intended research purpose. This phenomenon, known as top-down processing, is mostly associated with full, meaningful words, sentences, and connected speech. The following is an illustration of the concept of top-down processing.

In order to comprehend an utterance—whether it is a word, sentence or a longer utterance—it must be processed in the brain at different levels. Listeners must, for example, identify the phonemes that constitute the utterance (i.e. the basic level process called phonological processing). The listeners, then, use this phonological information to retrieve the corresponding word from their lexical repertoire (i.e. lexical processing). At later stages, they process this utterance at other cognitive levels (i.e. syntactic and semantic processing). At the
syntactic level, listeners identify the phrase or sentence structure in which the already-processed word exists. Finally, at the highest level (the semantic level), listeners identify the meanings of the phrases and sentences so that they can comprehend the whole utterance. If language comprehension occurs in this upward sequential processing (i.e. starting from the basic phonological level to the higher-order levels), language comprehension occurs in a bottom-up processing. Conversely, if the utterance is processed first at the higher-order levels (semantic level) downwards to the basic phonological level, this processing is called top-down (Carroll, 2008, pp. 45-65).

Each of these two concepts (i.e. top-down and bottom-up processing) is associated with a specific type of stimulus used in sound perception and production study. Using meaningful utterances in an experiment, whether they are words or longer utterances, is likely to stimulate higher-order language processes. If so, these processes influence the identification of phonemes. For instance, an English native speaker is presented with the sentence *The kids were sleeping in their loom* to test his/her ability to identify English liquids /r, l/. It is likely, in this case, that the participant will hear the last word as *room* rather than *loom* because, at the semantic level, the word *room* fits coherently with the rest of the sentence. If so, the test scores do not really reflect the participant’s ability to identify the sound because the semantic processing interfered with the phonological process which is the purpose of the test. In contrast, if the perceiver is tested in CV sequences (such as /ru:/ versus lu:/), the likelihood of higher-order processes taking place is minimized. In methodological terms, the stimulus construction determines the likelihood of the interference of the confounding variables (i.e. unintended factors) and, therefore, I used CV sequences in the present study which investigates sound production not language comprehension.
2.3.2 Rapid-shadowing paradigm in speech production tests.

A commonly used test method in speech production studies is the rapid-shadowing paradigm, which originated in cognitive psychology to test episodic memory (Goldinger, 1998). In rapid shadowing, participants are presented with an audio stimulus and they immediately repeat the utterance they have heard. Goldinger (1998) and other researchers report that there is a strong alignment between the exemplar (stimulus) and the shadowed sound in terms of their acoustic properties, such as fundamental frequency and word duration. From an articulatory perspective, based on previous research findings, Goldinger (1998) predicted that shadowing results in immediate configuration of the speech articulators so as to imitate the gestural constellations of the stimulus. The author’s prediction is further reinforced by the claims of the Motor Theory of speech perception. This does not suggest that rapid shadowing is a shallow, mechanistic process because Goldinger (1998) has found evidence of cognitive processes (required to retrieve the utterance from the episodic memory) that mediate sound production. This exemplar-based paradigm is reminiscent of Flege’s (1987b) “phonetic norm” in speech learning research. The notion of phonetic norm is simply an exemplar, or model speech, provided by monolinguals of the TL and it serves two purposes in speech learning research: it stimulates non-native speakers’ speech production, and it is used as a benchmark to gauge the acoustic characteristics of non-native speakers’ speech. In brief, rapid shadowing results in a remarkable consistency between the exemplar and the shadowed sound in terms of their acoustic features and articulatory gestures. Based on these findings, I have used rapid-shadowing paradigm in the sound production test. In addition, this model is appropriate for testing naïve perception and production, in which non-native speakers’ responses cannot be stimulated otherwise. The exemplar that I used in eliciting participants’ sound production in this test was
also used as a phonetic norm against which native-Arabic speaker perceptually gauged the production accuracy of non-native speakers.

2.4 Summary

To recapitulate, the discussion so far has focused on three major threads: the development of speech learning research, common themes in the field of the perception and production of Arabic pharyngealized and pharyngeal sounds, and finally some methodological commonalities in studies of sound learning.

Several hypotheses, models and theories address issues related to L2 speech learning at the phonological level. Namely, Lado’s (1957) pioneer hypothesis, the contrastive analysis hypothesis, encourages a long debate on whether similar TL sounds are easier to learn than new sounds. It has turned out later that the structural comparison between TL and NL sounds proposed by Lado (1957) cannot account for phenomenal learning difficulties. A few decades later, two models based on detailed phonetic description of the NL and TL have emerged. These models, the perceptual assimilation model (Best, 1995) and the speech learning model (Flege, 1995) have been influential in contemporary speech learning studies. These models have been tested in several languages such as Italian, German, Spanish, Japanese and others; however, learning Arabic phonology as a TL has received little attention particularly in testing it with PAM and SLM.

The dominant themes in Arabic phonology research are the articulatory mechanisms; the perceptual and acoustic correlates of the spread of emphasis; and non-native Arabic speakers’ perception and production of pharyngealization. A limited number of studies have explored the perception and production of pharyngealized consonants, rather than pharyngeal ones. I have
identified major gaps in this field of research, proposed reasons why these research findings cannot be generalized to unstudied pharyngealized and pharyngeal consonants, and raised some issues in these studies that need to be addressed in subsequent research.

In the previous sub-section, I have highlighted two major methodological concepts that are common in speech learning research that are particularly relevant to my study design. One is the psycholinguistic concepts of bottom-up and top-down language processing that result from the use of meaningful words as opposed to non-words. I have argued that using non-words (such as CV sequences) in the stimuli minimizes the confounding variables that may interfere with the intended results. Another methodological concept is the rapid-shadowing paradigm employed in language production experiments. The purpose of this sub-section is to present a testified ground of my study design.
Chapter III

Methodology

To reiterate, the primary goal of this study is to explore the learnability of Arabic sounds as evidenced in sound learning potential among non-native Arabic speakers with different proficiency levels. The study also aims to provide empirical evidence for predicting areas of difficulty in learning certain Arabic consonants. Thus, central to the study are non-native Arabic speakers’ proficiency (i.e. the independent variable) and their ability for learning Arabic pharyngeal and pharyngealized consonants (i.e. the dependent variable).

This chapter presents the methodology I employed in examining the production of Arabic pharyngeal and pharyngealized consonants. Namely, I explain the experimental design of the study and discuss some methodological limitations. After introducing the research participants, I describe the instrumentations adopted in this study. I then describe the procedures followed in data collection. Finally, in the data analysis section, I outline the statistical tests used to address each the research question.

3.1 Design of the study

The structure of the study was based on the posttest-only control group design presented in Campbell and Stanley (1963). Principally, this design encompasses three experimental elements: recruiting two equivalent groups of participants, introducing the independent variable,
and measuring the effect of the independent variable on the dependent variable (Cozby, 1993). In this section, I explain these constructs in relation to my study.

My study consisted of two groups of participants: experimental and control group. The main difference between them was that the experimental group had learned Arabic as an additional language while the control group had not. This means that the two groups essentially differed in their exposure to Arabic language. It is worth noting that this exposure occurred in real-life settings, which means that participants learned the language before being recruited for this study. In order to measure the effect of the independent variable on participants’ ability to produce the target Arabic sounds (i.e. the dependent variable), all participants conducted the same sound perception-production test. For the same measurement purpose, I employed a subsequent perception test using the data obtained from the previous test. Given this, the study encompassed the three constructs of the posttest-only control group design.

3.1.1 First construct: Recruiting two equivalent groups.

The first element of this research design does not merely require obtaining two groups, it also emphasizes the equivalence of the two groups—i.e. the experimental and control groups. Research groups can be considered equivalent if participants’ individual characteristics are equally distributed between the two groups. The basic principle behind this element is that the equivalence of research groups assures the sample’s representativeness of the study population, as well as the lack of bias against them which would undermine the internal validity of the study. The comparability of the two groups can be achieved in different ways such as randomization; pretesting (which is not feasible in this design); and assigning the same subjects to both groups (Cozby, 1993).
Various operational measures were adopted in order to ensure the representativeness and equivalence of the groups. For instance, participants were selected randomly to eliminate bias and ensure the groups’ equivalence. Furthermore, I administered diagnostic tests for all participants in order to assess the groups’ comparability. This assessment consisted of two measurements including a written questionnaire which assessed individual characteristics that might influence participants’ performance. The second measurement was a listening-and-speaking diagnostic which revealed possible speaking or hearing impairments that might interfere with the test. Implementing these two measurements validated the assumption that participants in both groups had comparable individual characteristics and language background, and that they were equally capable of undertaking the sound perception-production test.

Moreover, statistical analysis of participants’ individual characteristics provided further evidence of the equivalence of the two groups. In brief, my study attempted to satisfy the first construct of the posttest-only control group design in the sense that it consisted of two research groups which were equivalent and representative of the study population.

### 3.1.2 Second construct: Introducing the independent variable.

The second element of the posttest-only control group design is the intervention of the independent variable into one of the research groups. This divides participants into two distinct groups: an experimental group (exemplified by Arabic learners in this study) and a control group (which the non-learners of Arabic represented).

In my study, English-speaking participants who learned Arabic as an additional language exemplified the experimental group whereas the non-learners (who never had any systematic exposure to Arabic or any other Semitic language) represented the control group. English was
the dominant language of all participants in the two groups, and they shared comparable language background except for Arabic.

Importantly, the experimental group had been exposed to Arabic for a minimum of two years in real-life settings, rather than in experimental settings (i.e. exposing them to the language in laboratory settings or in language courses designed especially to meet the experimental requirements by controlling the quantity and quality of Arabic input). This is crucial to the study for several reasons. First, it was not feasible to teach participants the language for such a long period of time. If participants were taught the language for two years in experimental settings, the study would be vulnerable to attrition—also called “mortality” (Dörnyei, 2007; Campbell et al. 1963)—which threatens its internal validity (Marczyk, DeMatteo & Festinger, 2005). Because of the long time commitment and effort in learning Arabic, it is likely that participants would have dropped out from the study. Secondly, learning Arabic in real-life settings reinforces—to some extent—the ecological validity of the study results. The conditions in which participants learned the language are representative of real life and, thus, the results can be generalized to real-life conditions. Of course, the generalizability of the results is not exclusively subject to ecological validity (Gliner & Morgan, 2000, pp. 393-395).

3.1.3 Third construct: Measuring the effects of the independent variable.

The final construct of the study design was measuring the effects of the independent variable on the dependent variable, which constituted the core of the study. I examined these effects using a sound perception-production test in which all participants took part. Then, using the data obtained from this test, two Arabic L1 speakers performed a perceptual judgment test to assess the divergence (or convergence) between participants’ sounds and the stimuli. These tests
were used to examine the effects of learning Arabic on the ability to produce pharyngeal and pharyngealized consonants.

3.2 Selection of Participants

3.2.1 Study population.

The population of this study consisted of all English L1 speakers who have never learned (or been exposed to) any Semitic language; all native speakers of Arabic; and all speakers whose dominant language is English and who had learned Arabic for two years and more. Importantly, the geographic boundaries of the study population are limited to the Province of Ontario, Canada. Using various recruitment methods, this population was targeted in densely-populated areas such as public libraries, social organizations, immigration centers, and different educational institutions.

3.2.2 Sampling method.

The sample of the study was randomly selected. For both participant groups, stratified random sampling was applied in order to attain a higher level of representativeness and equivalence (Babbie, 2007, pp.189-207 and Lunenburg & Irby, 2008, pp.169-172). Since speaker’s gender, in particular, was reported to influence the production of pharyngealization, gender was the criterion according to which I stratified the research groups. Khan (1975), for example, found that female Cairene Arabic speakers exhibited less pharyngealization effect than the male Cairene Arabic speakers. Both research groups—i.e. the learner and non-learner groups—had an equal number of male and female participants. This is how stratification was reinforced. Randomness, on the other hand, was achieved in two ways. In some cases, for example, I selected random serial numbers from several lists of high school students. Other
participants were also randomly selected from a list of respondents to research advertisements at different sites.

### 3.2.3 Participants’ demographic information.

According to Dörnyei (2007), a total number of 15 participants in each group is the minimum requirement for quantitative, comparative and experimental research. Given this fact, I recruited a total of 47 participants ($N = 47$). None of whom reported any speaking or hearing impairment. There were 20 participants ($n = 20$) in each group—the learners’ and non-learners’ groups. Four participants performed a pilot study, and a 25-year-old Arabic native speaker participated in constructing the audio stimuli. Given Khan’s (1975) findings that male Arabic speakers exhibit more salient pharyngealization than female speakers, a male speaker was selected for recording the stimuli. This Arabic speaker had spent about 3.5 years in an English speaking country where he was doing his undergraduate degree. Two other Arabic L1 female speakers participated in a perceptual judgment task: one was a 31-year-old speaker and the other was 27-year old. The former had lived in an English-speaking environment for 9 months while the latter had spent 2.5 years in the same environment. However, both of them had finished their undergraduate studies in an Arabic-speaking country where they spent most of their lives.

### 3.2.3.1 Learner and non-learner groups.

Participants in the two groups shared similar demographic characteristics to some extent. Since the sample was stratified based on gender, there were 10 female participants in each group, and an equal number of the other sex—constituting a total of 20 participants of each gender. Excluding the participant who did not declare her age, participants’ ages ranged between 16 and 33 years ($M = 21.26; Mdn = 20; SD = 4.68$).
The language backgrounds of participants in both groups were also comparable; for instance, all participants of the two groups were multilingual. As Table 3.1 demonstrates, the minimum number of languages participants spoke was three. In addition, English was the only dominant language for most participants except three of them who reported themselves as bilingual. English was the dominant language among these three bilinguals as well. The ‘dominant’ language, in this context, refers to the language that a participant spoke most frequently (and with greatest fluency). The study emphasized the dominance, rather than the ‘nativeness’, of the language for several reasons. In essence, the study was concerned about the difficulty of using the pharynx as a primary or secondary place of articulation for consonantal sound production. This can be estimated by determining each participant’s most frequently spoken language and its phonetic system. Moreover, the self-assessment of language proficiency revealed that some participants were more proficient in their dominant language than in other languages including their native language (NL). In addition, all participants’ English-language use was further reinforced by the fact that they were all employed in English-speaking institutions, either as students or as workers. These observations suggest that English was the dominant language for both groups, and justify the preference of language dominance over nativeness. These demographic characteristics apply to the learner and non-learner groups alike.

Table 3.1
Descriptive statistics of participants’ age and number of spoken languages

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>39</td>
<td>17</td>
<td>16</td>
<td>33</td>
<td>21.26</td>
<td>4.68</td>
<td>21.93</td>
</tr>
<tr>
<td>No. of languages</td>
<td>40</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td>3.90</td>
<td>.96</td>
<td>.91</td>
</tr>
<tr>
<td>Valid N</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing N</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. Only the participants in the two groups are included in this table. Arabic L1 speakers are not included.*
Table 3.2 demonstrates group differences and similarities in terms of participants’ age and the number of languages used including the dominant language. Although the two groups were not identical, they were comparable—to some extent—in terms of their demographic characteristics. Evidence of comparability included, for example, participants’ dominant language; their multilingualism; employment in English-speaking institutions; age group and gender.

Table 3.2
Group differences in age and number of spoken languages

<table>
<thead>
<tr>
<th></th>
<th>Valid N</th>
<th>Missing N</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>20</td>
<td>0</td>
<td>18.05</td>
<td>2.21</td>
<td>8</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>Languages*</td>
<td>20</td>
<td>0</td>
<td>3.85</td>
<td>.67</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Non-learners</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>19</td>
<td>1</td>
<td>24.63</td>
<td>4.21</td>
<td>14</td>
<td>19</td>
<td>33</td>
</tr>
<tr>
<td>Languages*</td>
<td>20</td>
<td>0</td>
<td>3.95</td>
<td>1.19</td>
<td>4</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

* The number of languages includes participants’ L1.

3.2.3.2 Arabic language background.

An essential difference between the two groups, however, pertains to participants’ experience in Arabic language, the independent variable. Whereas the control group had never received Arabic input on a regular basis, the other group had varying degrees of exposure to the language. Their Arabic language experience was examined according to the following criteria:

- Means of learning the language (for example, formal classes; self-learning; media; and social interaction).
- Self-assessment of proficiency in Arabic, including the assessment of the overall proficiency and the four language skills: speaking, listening, reading and writing.
- Arabic input reoccurrence (i.e. how frequently they were exposed to Arabic).
• Age at which they started pronouncing Arabic sounds, and until which age.
• Length of their Arabic speaking experience in particular.
• Perception of the most difficult and easy aspects of the language (for example, grammar, sounds, orthography, and vocabulary).

As shown in Table 3.3, participants varied in the length of time in which they had practiced speaking Arabic (or even pronouncing Arabic sounds) in spite of their speaking fluency level. To illustrate, some participants had been speaking Arabic for two years—the minimum amount of time—while others had been speaking it for up to 15 years ($M = 11.05$, $Mdn = 12$, $SD = 3.56$).

Some participants ($n = 2$) started pronouncing Arabic sounds as young as 2-year old. All participants in the learners group had formal Arabic classes as the main source of learning the language, and some of them had had additional means of learning it.

Table 3.3

Learners’ Arabic-speaking experience: Age at which they started speaking Arabic, until when and for how long

<table>
<thead>
<tr>
<th></th>
<th>Valid N</th>
<th>Missing N</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>From age</td>
<td>20</td>
<td>0</td>
<td>5.30</td>
<td>4.61</td>
<td>2</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>To age</td>
<td>20</td>
<td>0</td>
<td>16.35</td>
<td>2.21</td>
<td>11</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>Time length$^a$</td>
<td>20</td>
<td>0</td>
<td>11.05</td>
<td>3.56</td>
<td>13</td>
<td>2</td>
<td>15</td>
</tr>
</tbody>
</table>

$^a$Measured in the number of years.

3.3 Instrumentation

Four core instruments were used in the process of data collection. Namely, a language-background questionnaire and auditory stimuli were both used in two tasks: a perception-production task and perceptual judgment task. This section discusses these instruments in terms of their format, content, purpose, and reliability of data obtained from them.
3.3.1 Language background questionnaire.

Besides its methodological importance in measuring groups’ comparability, the language questionnaire was planned to study potential relationships between individuals’ language background and their performance in the sound production test. The language questionnaire consisted of six subsets constituting a total of 21 items (Appendix B). The subsets were multidimensional in the sense that the first subset, for example, measured the languages a participant learned, the age at which they learned them and the learning sources. Using a 3-point rating scale, the second subset assessed the receptive and productive skills and the overall proficiency in each language. The other subsets examined participants’ experience in learning Arabic. While the third subset measured Arabic aural input that a participant might have received, the fourth one examined Arabic speaking experience. Both the third and fourth subsets used roughly similar scales—forced-choice questions and temporal frequency scales. The pre-final subset dealt with the easiest and most difficult aspects of learning Arabic, including Arabic sounds, and the last subset was a ‘free space’ for further information or comments which participants might wish to share.

The internal consistency of data derived from the questionnaire was estimated using Cronbach’s alpha (Cronbach, 1951) reliability test of interrelated items. This psychometric test was applied on the items of the first and second subsets to test the data related to each language separately. To illustrate, all data pertaining to the second language (L2) were tested jointly—such as age of learning L2, L2 speaking proficiency, L2 listening proficiency, etc. The same process was repeated for all other languages that participants had learned. Table 3.4 presents the results of the internal consistency tests of subsets 1 and 2. As Table 3.4 shows, the lowest and highest reliability coefficients were .72 (L3) and .87 (L5), respectively, which suggest a high
internal consistency in this data. Since a very few participants reported learning up to six or seven languages, the internal consistency test could not be administered on data related to L6 and L7.

Table 3.4
Cronbach’s alpha results of internal consistency for each language participants learned

<table>
<thead>
<tr>
<th>L#</th>
<th>Cases</th>
<th>Reliability statistics</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Valid</td>
<td>Excluded(^a)</td>
<td>Cronbach’s alpha</td>
<td>C. alpha for standardized items</td>
</tr>
<tr>
<td>L1</td>
<td>36</td>
<td>4</td>
<td>.78</td>
<td>.88</td>
</tr>
<tr>
<td>L2</td>
<td>38</td>
<td>2</td>
<td>.85</td>
<td>.85</td>
</tr>
<tr>
<td>L3</td>
<td>35</td>
<td>5</td>
<td>.72</td>
<td>.75</td>
</tr>
<tr>
<td>L4</td>
<td>20</td>
<td>20</td>
<td>.76</td>
<td>.81</td>
</tr>
<tr>
<td>L5</td>
<td>7</td>
<td>33</td>
<td>.88</td>
<td>.85</td>
</tr>
</tbody>
</table>

Note. Reliability tests could not be applied on L6 and L7 because there were too few cases for the analysis.

\(^a\)Cases excluded due to missing values in a participant’s data.

In general, Cronbach’s alpha values showed that the data obtained from the questionnaire was highly consistent. This provided evidence for the reliability of participants’ responses in the questionnaire and, therefore, valid conclusions can be drawn based on these responses.

### 3.3.2 Auditory stimuli and perception-production task.

The auditory stimuli consisted of 54 sounds that were recorded and sliced using digital audio editing software called Audacity (version 2.0.0). Each stimulus was approximately 1.40-seconds long, and consisted of a consonant followed by a vowel (CV sequences). The consonants included:
• Arabic pharyngeal consonants \((n = 6)\): /ħ, ʕ/.

• Counterparts of pharyngeal consonants \((n = 6)\): /h, ʔ/.

• Pharyngealized consonants \((n = 12)\): /tˤ, ðˤ, sˤ, dˤ/.

• Plain counterparts of the pharyngealized consonants \((n = 12)\): /t, ð, s, d/, respectively.

• Fillers \((n = 18)\): /w, y, q, b, z, k/.

The first four sound classes (that is, the pharyngeal and pharyngealized consonants, and the counterparts of both) constituted the target stimuli. Among the six Arabic vowels, only the three long vowels were used in the whole set of stimuli—namely, /iː/; /uː/ and /aː/. The CV sequences were constructed in such a way as to repeat each consonant thrice, and each time it was accompanied by a different vowel. For instance, three CV cohorts were created for the consonant /t/: /tiː/; /tuː/; /taː/. The same process was applied on the target sounds and the fillers alike. In order to prevent participants from realizing any sequencing patterns and to minimize potential memory-related effects, the auditory stimuli were arranged in a random order. A complete list of the randomly ordered CVs is provided in Appendix C.

The previously recorded audio CVs were employed in the perception-production test in order to stimulate participant’s response. The test was designed using Microsoft PowerPoint 2011. This software is ideal for the functions it provides; for example, uploading the audio files that contained the stimuli, recording participants’ sounds in response to the stimuli, and extracting all audio files from the original PowerPoint file so that they can be processed by other software. Furthermore, when the audio files are extracted, they can be easily matched to the slide in which they had been recorded.
3.3.2.1 **Design of the perception-production task.**

Each slide in the PowerPoint file contained an audio stimulus and allowed the participant to record his/her sound in the same slide. This task used a laboratory-based *rapid-shadowing paradigm* (Goldinger, 1998; Honorof, Weiheing & Fowler, 2011). It was also a *self-paced* task as participants were allowed to pause and proceed at their own pace. This task was divided into two stages. In the first stage participants performed the rapid-shadowing task in which they listened to the disembodied sounds and imitated them immediately. The next stage was a self-comparison between the stimulus and a participant’s shadowed sound. A five-point *numerical rating scale* (Dörnyei, 2007, p. 106) was used to compare the two sounds. This one-to-five scale represents a continuum of ‘completely similar’ to ‘completely different’. The purpose of the rapid-shadowing task was to examine a participant’s ability to produce the target sounds, which denoted sound learnability relative to speakers’ Arabic experience. The second part of the task aimed to measure a participant’s perception of their sounds’ divergence from (or convergence to) the stimuli.

3.3.2.2 **Reliability of task responses.**

Psychometric tests of internal consistency of the self-assessment of shadowed sounds provided evidence of instrument validity. These reliability tests were administered on each sounds category separately. This means that the psychometric test of responses related to pharyngeal sounds was separate from the reliability test of filler sounds, for example.

The internal consistency of shadowed fillers was estimated by Cronbach’s alpha (Cronbach, 1951). The reliability coefficient of the self-comparison scores of shadowed fillers was .86 (*N* = 40), which revealed a high reliability measurement of the self-assessment of
shaded fillers. *Split-half* reliability coefficients (Jackson, 1979) of the same data were .83 ($N = 9$) and .70 ($N = 9$). Split-half coefficients of the responses on shadowed pharyngealized sounds were .59 ($N = 6$) and .74 ($N = 6$). Cronbach’s alpha reliability test of the self-assessment scores of shadowed non-pharyngealized consonants also yielded a fairly high coefficient ($\alpha = .76$, $N = 12$). Similarly, the scores obtained from the self-assessment of the pharyngeal sounds and their counterparts yielded a high measure of reliability of responses ($\alpha = .86$, $N = 12$).

### 3.3.3 Perceptual judgment task.

#### 3.3.3.1 Task design.

Only two Arabic L1 speakers participated in this dual-function task. It was a dual-function in the sense that participants performed two sub-tasks simultaneously. That is, in addition to the *sound identification* sub-task in which the Arabic perceivers identified the shadowed sound, they rated the degree of similarity between the stimulus and the corresponding shadowed sound. This later sub-task (i.e. perceptual judgment) employed the same five-point scale that was used in the perception-production task. Sound identification purposefully preceded the perceptual judgment so that the stimulus did not interfere with the perceiver’s ability to identify the shadowed sound. The perceptual judgment task aimed to measure the perceptual divergence between the stimuli and the shadowed sounds. A high degree of convergence between the stimulus and the shadowed sound indicated participant’s potential of producing (and more generally learning) the sound. Conversely, a great degree of divergence between the stimulus and participant’s response signaled sound learning difficulty.
3.3.3.2 Reliability of perceptual judgment responses.

In order to measure the reliability of data obtained from this task, I used an interrater reliability test called Kappa (Cohen, 1960). This test is appropriate for testing the reliability of data obtained from this task data because, in this task, two raters responded to the same variables. In contrast to an intuitive way of calculating agreement (such as the percentage of agreement), Kappa takes into consideration the cases in which both raters agree by mere chance. In fact, Kappa is applicable on categorical data; however, it can be used on the data obtained from the perceptual judgment task because numerical scale (which I used in this study) is convertible to a semantic, categorical scale (Dörnyei, 2007). Some statisticians (such as Landis & Koch, 1977; Munoz & Bangdiwala, 1997) offered different interpretations of Kappa coefficient which are similar to a great extent. I interpreted the coefficients I obtained according to Landis et. al. (1977). At a .05 alpha, Kappa coefficients were statistically significant for all target sounds assessed by Arabic perceivers, except for /su:/ and /sˤuː/: $\kappa = .16, p > .05; \kappa = .14, p > .05$, respectively. Appendix D provides Kappa coefficients for measuring the internal consistency between the two Arabic perceivers’ assessment of the shadowed sounds. In general, Kappa suggested that the ratings provided by the two Arabic perceivers were consistent; therefore, this evidence of the reliability of the perceptual judgment task responses ensured the validity of the conclusions derived from these responses.

3.4 Data collection procedures

The study was conducted over three distinct phases: prior to start collecting data in two phases, I complied with research logistic requirements and prepared the materials that I used later in collecting data.
3.4.1 First phase of data collection.

I first obtained a research ethics approval from Carleton University’s Research Ethics Board (Appendix E). I also started constructing the stimuli and randomizing them (Appendix C illustrates the target and filler CV sequences and the random order in which they were presented to participants). In structuring the CV sequences, I created three cohorts of each target consonant—pharyngeals (i.e. /ħ, ʕ/); pharyngealized consonants (i.e. /tˤ, sˤ, dˤ, ðˤ/) and the counterparts of both (/h, ʔ/ and /t, s, d, ð/, respectively). Each cohort had one of three vowels: /aː, iː/ and /uː/. I constructed the filler sounds (i.e. /w, j, b, z, k/ and /q/) in the same way before randomizing the whole sounds set. In addition to ethics clearance and stimuli construction, I prepared advertisements to recruit prospective participants. Research advertisements included posters, emails and cards, all of which had similar content (see Appendix F for samples of research recruitment advertisement). The pre-final procedure in this phase was recording the audio stimuli after being able to recruit a male Arabic L1 speaker. The stimuli were recorded in a quiet room to minimize the background noise. Using Audacity, the Arabic speaker produced all CVs continuously in the same randomized order with a 4-sec time interval between successive CV stimuli. Using the same software, I separated the stimuli from each other so that each stimulus was saved in a separate audio file. I inserted these sound files in a PowerPoint file, and each slide contained an audio stimulus and its random number in addition to a ‘hidden’ reminder of the 5-point numerical scale. I conducted several pilot studies which revealed that some CVs were not typically representative of the Arabic sound in the sense that Arabic L1 speakers provided inconsistent sound identification responses; therefore, I re-recoded them and repeated the pilot study.
3.4.2 Second phase of data collection.

The second phase, in which the perception-production task was conducted, was the longest phase in data collection. Only English L1 speakers (N = 40) were involved in this phase. The test took place in a quiet room at Carleton University or at the participant’s respective institution. Prior to starting the production test, participants were required to sign a consent form (Appendix G) and fill out a language questionnaire (Appendix B). In compliance with research ethics protocols, oral instructions about participants’ rights preceded the test. As a precaution, I informed participants that there might be some degree of divergence from (i.e. dissimilarity to) the stimuli and that all what they had to do was to “try their best”. Importantly, I avoided explaining the details of the second part in the initial instruction in order to prevent participants from thinking metalinguistically while producing the sounds. The instructions were also intended to explain the purpose of the test and to explain how to perform it. This was aided by a brief PowerPoint orientation that looked exactly the same as the actual test; however, it consisted of only two filler sounds, not the target sounds. The purpose of this orientation was not restricted to illustrating the test; rather, it was intended to diagnose any speaking or hearing impairment and potential technical issues. Participants were offered a monetary compensation (10 CAD) for their participation in the study.

The first part of the perception-production test employed a rapid-shadowing paradigm in which the participants pressed on the microphone icon to listen to the stimulus and imitate it rapidly. Then, they proceeded to the next sound, and so forth. Throughout the whole shadowing test, the software was automatically recording their sounds without participants’ intervention.
Then, before starting the second part of the test, I explained it in detail; emphasized that it was self-paced; and revealed the ‘hidden’ reminder of the numerical scale. I told participants to ‘compare’ between the stimulus and the corresponding sound they had produced using a scale that ranged from 1 (which means that the two sounds were completely similar) to 5 (which means that they were completely different). I intervened in this part of the test, which was multifunctional, by playing the audio sounds so that participants’ performance in the intended test was not affected by the technical complexity of the task. Nevertheless, the two parts of the task were self-paced in the sense that it was the participants themselves that proceeded from one slide to another at their own pace.

3.4.3 Third phase of data collection.

The last phase of data collection was a perceptual judgment task in which only two Arabic L1 speakers participated. Similar to the previous test, the perceptual judgment task was conducted on an individual basis and was self-paced. This task resembled the second part of the perception-production test except for the fact that while the English participants were the ones who conducted the self-assessment, it was the Arabic listeners who performed the assessment in the perceptual judgment task. However, before assessing the divergence between the stimulus and the shadowed sound, Arabic listeners identified the shadowed sound and transcribed using Arabic orthography. Then, they listened to the stimulus and estimated the degree of dissimilarity between the two sounds. They applied the same procedures for all sounds, the target and filler sounds alike. Importantly, each Arabic speaker did the assessment for all shadowed sounds ($N = 2,160$) obtained from the perception-production task.
3.5 Data analysis

In this quantitative study, I conducted several statistical tests using SPSS 21.0 program. This analysis encompassed the data obtained from the perceptual judgment task. In addition to calculating the mean of both Arabic perceivers’ scores to segregate data, I employed descriptive statistics for analyzing most common responses represented in the Arabic orthographic transcription they provided for the shadowed sounds.

Three types of tests were repeatedly used in addressing the research questions: homogeneity tests; tests of significant variance; and a post-hoc analysis. All statistical analyses were calculated at an alpha level of .05. In order to test significant variances, I used split-plot ANOVA which is different from the one-way ANOVA in the sense that the former is designated for testing two or more levels of the independent variable against two or more levels of the dependent variable. Fundamental to the validity of split-plot ANOVA is the equality of distribution of the dependent variable data; therefore, I used homogeneity tests—including Box’s, Mauchly’s and Levene’s test—to validate ANOVA results. The post-hoc analysis I employed was Fisher’s least significant difference (LSD) which yielded similar results to Bonferroni statistic. In one instance where split-plot ANOVA produced non-significant results (i.e. in testing vowel effect on pharyngeal consonants’ production), I conducted a paired-sample t-test in addition to LSD to find out where exactly the insignificance came from. The independent variable for all ANOVA analyses was Arabic proficiency, which divided participants into two groups, learners and non-learners of Arabic. Participants’ production scores for these sound categories represented the levels of the dependent variable which differed across the research questions.
3.6 Summary

This chapter introduced methodological description of the study including its design, participants, instrumentations, procedures of data collection, and an overview of statistical tests employed in analyzing data.

The sample \( (N = 47) \) was randomly selected from different sites. Stratified random sampling was the selection procedure for the participants who were assigned to groups, and gender was the criterion according to which I stratified them. Both groups were similar in most aspects including their L1 (English), except for the fact that the learners group learned Arabic as an additional language while the other group did not.

The framework of the study resembled the posttest-only control group design in the sense that two groups of participants were involved in the study and that all of them were tested in the effects of the independent variable (i.e. Arabic proficiency). The dependent variable, on the other hand, was participants’ ability to produce the Arabic target sounds.

In order to measure these effects, two core instrumentations were used: perception-production test and perceptual judgment task. Both groups \( (N = 40) \) conducted the perception-production task which employed rapid shadowing using audio stimuli of Arabic pharyngeal, pharyngealized consonants and their counterparts. The subsequent task, the perceptual judgment task, was designed to measure the perceptual degree of divergence between the exemplars (i.e. stimuli) and the shadowed sounds. Two Arabic L1 speakers performed this task, each of whom transcribed all shadowed sounds \( (N = 2,160) \) and then compared between them and the exemplars using a five-point numerical scale. The following chapter explains in detail the statistical findings of the perceptual assessment data obtained from these Arabic speakers.
Chapter IV

Data analysis

The study was designed in a way such that Arabic proficiency was the between-subjects variable and sound categories constituted the within-subjects repeated factor with several levels. The repeated factor encompassed the pharyngealized consonants /tˤ, sˤ, dˤ, ðˤ/; pharyngeal consonants /ħ, ʕ/ and the plain counterparts of both (i.e. /tˤ, sˤ, dˤ, ðˤ/ and /h, ʔ/, respectively). I constructed three cohorts (CVs) for each sound, and each cohort has a different adjacent vowel (i.e. /aː, iː, uː/). This constituted a total of 36 target CVs as illustrated in Appendix A.

The research design served the primary purpose of the study—that is, to investigate English L1 speakers’ production of these sound categories as they vary in their Arabic proficiency level. Importantly, in addressing all research questions, English speakers’ production of all sounds was assessed according to Arabic L1 listeners’ perception of these sounds. In order to examine the effect of Arabic proficiency on the ability to produce these sounds, I compared Arabic learners’ (experimental group) sound production to that of the non-learners (control group). The same principle applied to the four research questions: non-native Arabic speakers’ production of pharyngeal, pharyngealized consonants and the counterparts of both; their production of pharyngeal consonants as opposed to pharyngealized ones (i.e. /h, ʕ/ versus /tˤ, sˤ, dˤ, ðˤ/); intra-category production differences among pharyngealized sounds (i.e. /tˤ/ vs. /sˤ/ vs. /dˤ/ vs. /ðˤ/) as well as the pharyngeal ones (i.e. /h/ vs. /ʕ/); and the effect of vowel manipulation
on the production of pharyngeal CVs (i.e. /haː, ʕaː/ vs. /hiː, ʕiː/ vs. /huː, ʕuː/) as well as the pharyngealized ones (i.e. /tˤaː, sˤaː, dˤaː, ðˤaː/ vs. /tˤiː, sˤiː, dˤiː, ðˤiː/ vs. /tˤuː, sˤuː, dˤuː, ðˤuː/).

This chapter is devoted to the statistical analysis of the stated research questions. A brief overview of descriptive statistics of the within-subjects factors is followed by a presentation of inferential-statistics findings arranged by research questions. I conducted Split-plot ANOVA in addressing most research questions and sub-questions. In a few cases where this statistic was not applicable, I reported the results of multivariate tests instead.

4.1 Descriptive statistics of the dependent variable

Participants of both groups—the learners and non-learners of Arabic—were tested using the whole set of sounds. Table 4.1 presents the dispersions and central tendencies of the scores given to each sound category produced by all participants ($N = 40$). There are some cross-category differences in participants’ mean scores. A high score (maximum of 5) indicates more similarity between the stimulus and the sound produced in response to it. Given this fact, the plain counterparts of pharyngealized consonants /t, s, d, ð/ gained the highest scores whereas the lowest mean score was associated with the pharyngealized ones /tˤ, sˤ, dˤ, ðˤ/.

The learners and non-learners groups differed in their experience in Semitic languages in general, and Arabic in particular. Since this study was concerned with Arabic sound production, Table 3.3 (in the Methodology chapter) highlights self-declared information of Arabic learners’ experience in Arabic speaking. The mean length of participants’ Arabic speaking experience was approximately 11 years. The mean age at which they started speaking Arabic was 5.3. This constituted a considerable difference between the two groups of participants—Arabic learners and non-learners groups. Thus, the question posed is whether this difference in Arabic
experience (and proficiency) yields significant differences in speakers’ production of various Arabic sound classes. The remainder of this chapter addresses this question.

<table>
<thead>
<tr>
<th>Table 4.1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive statistics of the within-subjects factors (sound categories)</strong></td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SD</td>
</tr>
</tbody>
</table>

Note. Phz. = pharyngealized; Non-phz. = plain counterparts of pharyngealized; Ph = pharyngeal; Non-phz. = plain counterparts of pharyngeal sounds.

4.2 Statistical testing of research questions

Prior to conducting inferential statistics, I reversed the 5-point comparison scale so that the higher the score, the more similar the sound is to the stimulus. Furthermore, the score of each shadowed sound was the mean of the two scores provided by Perceiver 1 and Perceiver 2. I computed sound categories and sub-categories by calculating the mean of all sounds that belong the corresponding (sub)-category. For instance, the pharyngealized consonants category was the mean of participants’ scores in /tˤ, sˤ, dˤ, ðˤ/. Similarly, /tˤ/ sub-category was the mean of the scores that participants obtained for its three cohorts /tˤaː, tˤiː, tˤuː/, and so forth.

4.2.1 Assumptions of split-plot ANOVA.

Split-plot ANOVA statistic (also called “repeated-measures ANOVA”) was used to address the research questions—testing significant differences between the within-subjects factors (sound categories) at all levels of the between-subjects variable (Arabic proficiency).
This statistic is appropriate for designs in which one independent variable—which divides participants into two groups—is tested against more than one level (or type) of the dependent variable. In the present study, Arabic proficiency divided participants into two groups each of which were tested in several sound categories, which constituted the levels of the within-subjects factor.

Importantly, split-plot ANOVA is based on the assumption of equality of variance (called *homoscedasticity*) and covariance. To paraphrase, the variances among the scores that participants obtained for the sounds in question must be equal. Also, since each participant obtained more than one score (i.e. there are multiple sounds, and each participant received a score for each of these sounds), the covariance matrix of these scores must be equal.

Hence, I reported the results of three tests of homogeneity (i.e. equality): Box’s test of equality of covariance matrices; Mauchly’s test of sphericity (i.e. circularity of scores distribution); and Levene’s test of equality of error *variances*. Since Mauchly’s test is the most commonly reported test in studies using repeated-measures (RM) ANOVA, this study emphasized *this* test in particular. In these three tests, the variance and covariance of production scores are said to be homogenous if the test yields non-significant results, in which case the null hypothesis cannot be rejected. If the test output is significant ($p < .05$), the test fails to support the null hypothesis that the covariance matrices are equal; however, this does not suggest that there is evidence of the inequality of the covariance matrices. In statistical terms, if the $p < .05$, the null hypothesis (which states that the variances are equal) is rejected; however, this does *not* prove the alternative hypothesis (which states that the variances are not equal). In other words, the covariance matrices can be equal, yet the homogeneity test may not be able to capture the equality of covariance matrices (when $p < .05$). Levene’s and Box’s tests are interrelated in the
sense that the former tests the variances of scores for each type of the dependent variable separately while the latter tests the covariance of scores in these two types of dependent variable together. If Levene’s test shows that the scores of at least one type of the dependent variable (e.g. Type A) cannot be assumed to be homogenous, Box’s test shows that the covariance between the matrix of variable A and the matrix of another variable(s) cannot be assumed to be equal.

If Mauchly’s test fails to support the null hypothesis that the covariance matrices are equal, there are three possible alternatives:

- Using a multivariate test: for example, Pillai’s Trace, Wilks’ Lambda, Hotelling’s Trace, and Roy’s Largest Root.
- Using Friedman test.
- Using corrected $F$-ratio of the split-plot ANOVA.

With respect to the third alternative, corrections such as Greenhouse-Geisser (Greenhouse and Geisser, 1959); Huynh-Feldt; and Lower-bounded alter the degrees of freedom ($df1$ and $df2$) so that an adjusted $F$-ratio can be produced. Each of these corrections (i.e. Greenhouse Geisser and the others) produces its own $F$-ratio that may (and may not) be different from the $F$-ratios associated with the other corrections. In the case Mauchly’s test fails to capture the sphericity, the adjusted $F$-ratios that are produced by these corrections are different from the original ANOVA $F$-ratio. For instance, Huynh-Feldt’s $F$-ratio is different from the usual $F$-ratio and may also be different from Greenhouse-Geisser’s $F$.

These corrections first quantify the degree to which the covariance matrices deviate from sphericity (Trujillo-Ortiz, 2006). If the covariance matrices are spherical, Epsilons indicate a value of 1. The less spherical the covariance matrices are, the further the value from 1 becomes.
Then, they multiply $df1$ and $df1$ by their own epsilon value in order to produce an adjusted $F$-ratio. Given this, if sphericity cannot be assumed and $\epsilon \neq 1$, the adjusted $F$-ratios are different from the original $F$. In cases where the dependent variable (i.e. sound production) has only two levels, the covariance matrices are spherical but Mauchly’s test cannot be computed. This claim is supported by epsilon value which estimates sphericity and which always equals 1 when there are only two types of the dependent variable.

**4.2.2 Selection of an alternative when sphericity is violated.**

In this study, Epsilon value was the criterion according to which I selected the appropriate alternative. To illustrate, when Epsilons were low, I reported the results of the multivariate tests instead of RM ANOVA. Otherwise, I reported a correction-associated $F$-ratio. The decision as to which correction to use was based also on Epsilon value. That is, when Epsilons were less than .75, I selected Greenhouse-Geisser adjustment. Otherwise, Huynh-Feldt’s adjusted $F$ was preferred (Leech, Barrett & Morgan, 2008, pp. 155-161). I excluded Friedman test in this study since it is a nonparametric test.

In brief, homogeneity of covariance is a condition for validating RM ANOVA. However, in cases where this assumption could not be assumed, I used either a multivariate test or a correction-associated $F$ instead of the invalid ANOVA $F$-ratio. The appropriate alternative was determined based on the exact degree of sphericity violation, Epsilon value.

The subsequent sections in this chapter discuss in detail the statistical findings for the research questions. I organized the discussions of all research questions in four successive elements: first, explaining the results of the homogeneity tests; second, discussing the results of both within-subjects and between-subjects main effects; third, introducing the results of the
interaction between the two variables; and, finally, presenting the results of a post-hoc test called Fisher’s least significant difference (LSD). LSD which is a combination of various paired t-tests gives an insight of the exact difference between groups by comparing the means of participants’ scores.

4.2.3 Question 1: English L1 speakers’ production of pharyngeal and pharyngealized consonants, and their counterparts.

This question was divided into two sub-questions each of which targeted a sound category and its corresponding plain category. The first sub-question tackled English speakers’ production of shadowed pharyngealized consonants and their plain analogues, whereas the second tackled their production of pharyngeal consonants and their plain counterparts.

4.2.3.1 English speakers’ production of pharyngealized sounds and their plain counterparts.

The goal was testing whether there was a significant difference in English speakers’ production of pharyngealized sounds /tˤ, sˤ, dˤ, ðˤ/ and non-pharyngealized sounds /t, s, d, ð/ as they vary in their Arabic proficiency levels. In order to achieve this goal, I used split-plot ANOVA after testing the homogeneity of variances.

First: Homogeneity of scores’ covariance matrices.

Three tests of the homogeneity of variance and covariance of participants’ production scores were conducted. Box’s test had a $p$ value of .48 which is not significant at a probability level of .05 (Box’s $M = 2.60, F (3, 259920) = .82, p > .05$). This suggested that the null hypothesis of equal variances between participants’ scores cannot be rejected.
Similarly, Levene’s test also yielded non-significant results for both sound categories. For the pharyngealized consonants category, the obtained probability was .98 (Levene’s $F(1, 38) = .00, p > .05$). Likewise, the error variances of the plain counterparts were also equal at the same degrees of freedom (Levene’s $F(1, 38) = 1.37, p > .05, ns$). This means that, all mean scores for pharyngealized phonemes and their plain counterparts were homogenous.

Since there were only two types of the dependent variable (i.e. production scores for two sound categories: pharyngealized and plain consonants), Mauchly’s test could not be computed. However, Epsilon value ($\epsilon = 1$) for the three correction tests, Greenhouse-Geisser and the others, indicated that the covariance matrices of the two sound categories under investigation were circular.

Given the results of Box’s, and Levene’s tests as well as Epsilon value, it can be said that the assumption of the homogeneity of scores’ variance and covariance held true. Fulfilling the homogeneity condition ensured that the results of split-plot ANOVA were not distorted. Thus, there was no need to refer to the multivariate tests or the adjusted F-ratio (all of which basically produced the same results as RM ANOVA).

**Second: Omnibus tests of within-subjects and between-subjects main effects.**

In this research sub-question, the within-subjects effects were the two sound categories labelled as pharyngealized consonants /tˤ, sˤ, dˤ, ðˤ/ and their plain counterparts /t, s, d, ð/. At a probability coefficient of .05, there were significant differences in the mean production scores of the two sound categories ($F(1,38) = 194.85, p < .05, \eta_p^2 = .84$, power = 1). These findings (illustrated in the first row of Table 4.2) suggested that participants’ scores in the two sound categories— pharyngealized and plain consonants—differed significantly with a large effect size
(\eta_p^2 = .84). With respect to the between-subjects main effect, participants also differed significantly in their Arabic language proficiency: \(F(1, 38) = 58.56, p < .05, \eta_p^2 = .61,\) power = 1. This large effect means that 61% of the between-subjects variance was accounted for by their Arabic language experience.

Table 4.2
RM ANOVA summary of the within-subjects factor (pharyngealized sounds and their counterparts) and its interaction with Arabic proficiency

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>\eta_p^2</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sounds (pharyngealized and plain)</td>
<td>1</td>
<td>47.99</td>
<td>194.85</td>
<td>.00</td>
<td>.84</td>
<td>1.00</td>
</tr>
<tr>
<td>Sounds * Group(^a)</td>
<td>1</td>
<td>8.26</td>
<td>33.55</td>
<td>.00</td>
<td>.47</td>
<td>1.00</td>
</tr>
<tr>
<td>Error (sounds)</td>
<td>38</td>
<td>.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)This means the interaction between the two variables: Arabic proficiency (group) and sound production.

Third: Interaction between Arabic experience and the repeated measures.

As displayed in Table 4.2, RM ANOVA showed a significant difference in participants’ production of pharyngealized consonants versus their production of the plain counterparts as influenced by the producer’s Arabic proficiency: \(F(1, 38) = 33.55, p < .05, \eta_p^2 = .47,\) power = 1. The \(F\)-ratio signifies two important findings (Gravetter & Wallnau, 2005, pp. 370-382). First, given the fact that the \(F\) is the ratio of the difference among participants scores in these sound categories to the difference expected to occur by chance (error), its large value means that the difference between sound categories was greater than what would be expected by error alone. Second, given the degrees of freedom in this test, the observed \(F\) \((F_{obs} = 33.55)\) was much greater
than the critical $F (F_{\text{crit}} = 4.10)$. This indicated a significant difference between learners’ and non-leaners’ production of pharyngealized sounds and their plain equivalents as perceived by Arabic L1 speakers. The observed power of this interaction was extremely high, and the $\eta^2_p$ value indicated that nearly half of the differences between scores were due to differences between the sound categories under question.

The scores given to both groups in pharyngealized consonants category were lower than the scores for /t, s, d, ð/; however, non-learners of Arabic (control groups) showed more considerable difference between these sound categories than learners did. Exact mean scores for the both groups in pharyngealized and plain CVs are provided in Appendix H1. Figure 4.1 further illustrates these findings.

**Fourth: Post-hoc test of pharyngealized and non-pharyngealized consonants’ production.**

The post-hoc analysis, Fisher’s LSD, tests the significance of two comparisons of means. The first is comparing between learners’ and non-learners’ mean scores for the two sound
categories together (between-subjects factor). It also compared between the mean scores of pharyngealized and non-pharyngealized consonants (within-subjects factor).

LSD showed a significant mean difference between Arabic learners’ and non-learners’ scores \((MD = .83, SE = .11, p < .05, 95\% CI [.61, 1.05])\), with learners scoring higher than non-learners for the two sound categories taken together. The post-hoc test also showed that the scores for pharyngealized consonants were lower than the scores for non-pharyngealized consonants, for both participant groups: \(MD = 1.55, SE = .11, p < .05, 95\% CI [1.32, 1.77]\).

Precisely, Arabic learners’ mean score for pharyngealized sounds \((M = 3.60, SD = .58)\) was substantially greater than non-learners’ mean score for the same sound category \((M = 2.13, SD = .52)\). However, the magnitude of the difference between participants groups decreases in the case of non-pharyngealized sounds, with learners’ mean score slightly higher than non-learners’ mean score \((M = 4.51, SD = .49; and M = 4.32, SD = .35, respectively)\).

**Summary of group differences in pharyngealized and non-pharyngealized CVs.**

In brief, since the homogeneity assumption held true for the within-subjects factor (i.e. the variances in participants scores for the sounds under investigation were equal), RM ANOVA was conducted to examine the variance between learners’ and non-learners’ scores in two sound categories: pharyngealized phonemes and their plain counterparts. These scores represented Arabic L1 speakers’ assessment (or perception) of the degree of divergence (or convergence) between the stimulus and the shadowed sound.

Omnibus tests of the main effects—i.e. within-subjects as well as between-subjects main effects—revealed significant difference in scores due to the difference in the respective variable
at a probability level of .05. In other words, the variability in participants’ Arabic proficiency as well the variability in their scores for the two sound categories was statistically significant.

In addition, split-plot ANOVA showed that the difference in English speakers’ proficiency in Arabic corresponds to the difference in their ability to produce pharyngealized sounds versus non-pharyngealized sounds. The difference between learners’ and non-learners’ mean scores for the pharyngealized consonants was greater than the difference between their scores for the non-pharyngealized consonants. This means that, pharyngealized consonants were more susceptible to the variability in speakers’ Arabic proficiency than the non-pharyngealized sounds.

### 4.2.3.2 English speakers’ production of pharyngeal consonants and their counterparts.

I conducted split-plot ANOVA in order to test whether there was a significant difference in English speakers’ production of pharyngeal and non-pharyngeal consonants as a result of participants’ proficiency in Arabic. The sound categories in question are the Arabic pharyngeal consonants /ḥ, ʕ/ and their non-pharyngeal (glottal) counterparts /h, ʔ/. Again, the independent variable divided participants into two groups: learners and non-learners of Arabic. It is fundamental to ensure whether the variances of production scores are homogenous prior to conducting the ANOVA test.

First: Homogeneity of covariance of production scores for pharyngeal and consonants and their counterparts.

According to Box’s test, the covariance matrices of participants’ scores in pharyngeal and non-pharyngeal sounds were equal (Box’s $M = 4.01, F(3, 259920) = 1.26, p > .05$). Levene’s test showed that the error variances for both sound categories were also equal:

- Pharyngeal sounds’ production scores: Levene’s $F(1, 38) = 4.10, p = .05$. 

• Non-pharyngeal sounds’ production scores: Levene’s $F(1, 38) = .09, p > .05, ns.$

Mauchly’s test provided distorted results because there were only two types of the dependent variable. Nonetheless, Epsilon value ($\varepsilon = 1$) demonstrated that the variances of differences between groups’ scores did not violate the sphericity assumption. In brief, given the findings of the homogeneity tests, it can be said that the variance and covariance of participants’ production scores for pharyngeal and non-pharyngeal consonants were significantly homogenous.

Second: Omnibus tests of within-subjects and between-subjects main effects.

The mean differences between the levels of the repeated measure were statistically significant: $F(1, 38) = 48.37, p < .05, \eta_p^2 = .56, \text{power} = 1$. This means that participants’ mean scores for pharyngeal and non-pharyngeal sounds were significantly different, regardless of their Arabic proficiency (Table 4.3). Likewise, the mean differences in participants’ Arabic proficiency (i.e. between-subjects main effect) were statistically significant: $F(1, 38) = 83.27, p < .05, \eta_p^2 = .69, \text{power} = 1$. As denoted in the large value of $F$-ratio, the obtained mean differences for participants’ Arabic proficiency level were greater than what would be expected to result by chance.

Table 4.3
RM ANOVA summary of the within-subjects factor (pharyngeal consonants and their counterparts) and its interaction with Arabic proficiency

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta_p^2$</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sounds (pharyngeal and non-pharyngeal)</td>
<td>1</td>
<td>15.24</td>
<td>48.37</td>
<td>.00</td>
<td>.56</td>
<td>1.00</td>
</tr>
<tr>
<td>Sounds * Group</td>
<td>1</td>
<td>34.56</td>
<td>109.70</td>
<td>.00</td>
<td>.74</td>
<td>1.00</td>
</tr>
<tr>
<td>Error (sounds)</td>
<td>38</td>
<td>.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Third: Interaction between speakers’ Arabic experience and their production of pharyngeal and non-pharyngeal sounds.

As presented in Appendix H2, Arabic learners’ mean score for the pharyngeal consonants \( M = 4.70, SD = .39 \) was greater than non-learners’ mean score for the same sound category \( M = 2.35, SD = .61 \). Conversely, non-learners’ mean score for the glottal category \( M = 4.54, SD = .56 \) was slightly greater than learners’ mean score \( M = 4.25, SD = .55 \). This means that the difference between the two groups’ production of pharyngeal and non-pharyngeal sounds differed in direction and magnitude. Figure 4.2 illustrates the direction and magnitude of these differences between the groups’ mean scores for the two sound categories.

Split-plot ANOVA revealed that this difference in the mean scores was significant as a result of its interaction with participants’ Arabic proficiency: \( F(1, 38) = 109.70, p < .05, \eta_p^2 = .74 \), power = 1. The difference in participants’ mean scores as a result of their Arabic proficiency level was greater than the difference expected to occur by chance. About three quarters of the
variability in participants’ scores in pharyngeal and glottal sounds was associated with the variation in their Arabic proficiency level ($\eta_p^2 = .74$).

*Fourth: Post-hoc test of pharyngeal and non-pharyngeal consonants’ mean differences.*

The post-hoc test, Fisher’s LSD, showed that learners’ production scores were significantly higher than non-learners’ scores for these two sound categories: $MD = 1.03$, $SE = .11$, $p < .05$, 95% CI [0.80, 1.26]. This does not mean that learners scored higher than non-learners for both sound categories; rather, it is the mean difference between learners’ total score for both sound categories and non-learners’ total score. The test also revealed that the mean score for non-pharyngeal CVs was significantly higher than the mean score for pharyngeal CVs: $MD = .87$, $SE = .13$, $p < .05$, 95% CI [.62, 1.13].

*Summary of group differences in pharyngeal versus non-pharyngeal CVs.*

To summarize the results of this sub-question, RM ANOVA demonstrated that participants’ scores for pharyngeal and non-pharyngeal sounds were significantly different. Also, the two groups’ Arabic proficiency was significantly different. The interaction between the two variables (Arabic proficiency and sound production) showed that participants’ Arabic proficiency corresponds to discrepant differences in their production of pharyngeal versus non-pharyngeal sounds. That is, the pharyngeal consonants produced by the learners were more similar to the stimuli than the pharyngeal consonants produced by the non-learners. The opposite was true in the case of the non-pharyngeal sounds—that is, non-learners’ scores surpassed the learners’ scores for the non-pharyngeal sounds. This indicates that Arabic proficiency induced different abilities to produce pharyngeal versus non-pharyngeal sounds.
4.2.4 Question II: Differences in Arabic learners’ and non-learners’ production of pharyngeal versus pharyngealized consonants.

To reiterate, the goal of this comparison was testing significant differences in non-native Arabic speakers’ mean scores for two sound categories: pharyngealized consonants /tˤ, sˤ, dˤ, ðˤ/ as opposed to the pharyngeal category /ħ, ʕ/. The mean scores represented Arabic L1 speakers’ estimation of the degree of perceptual similarity between the stimulus and the shadowed sound. In order to conduct split-plot ANOVA, the homogeneity of variance and covariance of the within-groups factor was investigated.

4.2.4.1 First: Homogeneity tests of variances in participants’ scores for the pharyngeal and pharyngealized categories.

It was found that the covariance matrices were equal as measured by Box’s test (Box’s $M = 6.98, F(3, 259920) = 2.19, p > .05$, $ns$). The non-significance of Box’s test results proved that the covariance of the two scores that each participant obtained for both pharyngealized and pharyngeal sounds was equal, regardless of the participant’s Arabic proficiency. As estimated by Levene’s test, equality of error variances in scores for the pharyngealized category was not significant (Levene’s $F(1, 38) = .001, p > .05$). The same test approached non-significance for the pharyngeal sounds category (Levene’s $F(1, 38) = 4.10, p = .05$). Furthermore, Epsilon values ($\varepsilon = 1$) showed that the variances of the differences in participants’ scores were spherical. According to the findings of Box’s and Levene’s tests, the variances and covariance matrices of participants’ scores in pharyngealized and pharyngeal sounds were equal. Thus, the within-subjects variable satisfied the homogeneity assumption that underlies split-plot ANOVA.
4.2.4.2 Second: Omnibus tests of within-subjects and between-subjects main effects.

A significant mean difference between participants’ scores for pharyngealized and pharyngeal consonants was found: $F(1, 38) = 44.62, p < .05, \eta_p^2 = .54$, power = 1 (Table 4.4). The differences in these scores which resulted from the difference between the two sound categories were greater than the expected error difference. The observed $F$-ratio was greater than the critical value ($F_{crit} = 4.10$) which denoted the significance of these findings.

### Table 4.4
RM ANOVA summary of the within-subjects factor (pharyngealized and pharyngeal sounds) and its interaction with Arabic proficiency

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>df.</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta_p^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sounds (pharyngealized and pharyngeal)</td>
<td>1</td>
<td>8.59</td>
<td>44.62</td>
<td>.00</td>
<td>.54</td>
<td>1.00</td>
</tr>
<tr>
<td>Sound * Group</td>
<td>1</td>
<td>3.81</td>
<td>19.80</td>
<td>.00</td>
<td>.34</td>
<td>.99</td>
</tr>
<tr>
<td>Error (sound)</td>
<td>38</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There was also a significant mean difference in participants’ proficiency in Arabic with a large effect size: $F(1, 38) = 196.96, p < .05, \eta_p^2 = .84$, power = 1. The percentage of difference associated with this main effect (Arabic proficiency) excluding individual differences was 84 as estimated by the partial eta squared (Gravetter & Wallnau, 2005, pp. 370-382).

4.2.4.3 Third: Interaction between Arabic proficiency and sound production.

Descriptive statistics showed that Arabic learners’ mean score for the pharyngealized consonants ($M = 3.60, SD = .58$) was less than their mean score for the pharyngeal category ($M = 4.70, SD = .39$). Likewise, non-learners’ mean scores for both sound categories showed the same direction of difference but with different magnitude. That is, their mean score for the
pharyngealized sounds ($M = 2.13, SD = .52$) was less than their score for the pharyngeal CVs ($M = 2.35, SD = .61$). The differences in groups’ mean scores for both sound categories were different in magnitude in the sense that the mean difference of learners’ mean scores (1.10) was greater than the mean difference in non-learners’ mean scores (.22) for the two sound categories (Appendix H3). This concept is illustrated in Figure 4.3 where the line that represents the Arabic learners group is steeper than the line of the other group. Arabic learners’ mean scores for both pharyngeal and pharyngealized sounds were considerably greater than that of the non-learners. By employing RM ANOVA, it was found that these differences in mean scores were statistically significant ($F(1, 38) = 19.80, p < .05, \eta_p^2 = .34$, power = .99). As indicated in the $F$-ratio, the differences in mean scores that were associated with the variability in Arabic proficiency were greater than the expected error differences. The effect size was moderate and the observed statistical power was high.

4.2.4.4 Fourth: Post-hoc test of pharyngeal and pharyngealized sounds’ differences.

Two post-hoc tests, Fisher’s LSD and Bonferroni, showed that the mean difference between non-learners’ and learners’ production was statistically significant: $MD = -1.91, SE = .14, p < .05, 95\% CI [-2.19, -1.63]$. The minus sign indicates that non-learners’ mean score was lower than learners’ mean score, so if the comparison was reversed, the mean difference ($MD$); standard error ($SE$); and confidence intervals ($CI$) would show values greater than zero. In addition, both groups’ mean scores for the pharyngeal CV’s production were significantly higher than their mean scores for the pharyngealized CV’s production: $MD = .66, SE = .10, p < .05, 95\% CI [.46, .85]$. 
4.2.4.5 Summary of group differences in pharyngeal versus pharyngealized CVs.

To summarize, the homogeneity of variance and covariance for participants’ scores in pharyngealized and pharyngeal categories was not rejected. Thus, split-plot ANOVA was conducted to examine the difference in participants’ production of these sound categories as accounted for by the difference in producers’ Arabic proficiency. The findings indicated that the difference in Arabic proficiency corresponds to significant difference in the ability to produce pharyngealized consonants compared to pharyngeal ones. However, the difference between the two sound categories was more obvious among Arabic learners than the non-learners. The lack of Arabic proficiency corresponds to lower production performance of both sound categories.

4.2.5 Question III: Intra-category differences in learners’ and non-learners’ sound production.

This question entailed two tests each of which addressed a different sound category. The first test compared the differences in participants’ scores of the two pharyngeal phonemes whereas the second test addressed those differences among pharyngealized consonants.
4.2.5.1 Group differences in the production of /h/ versus /ʕ/.

This sub-question addressed the differences between /h/ and /ʕ/ production as associated with the speaker’s proficiency level in Arabic. The goal was to examine whether non-native Arabic speaker’s proficiency in the language influenced the variability in their production of /h/ as compared to /ʕ/, and if it did, how the sounds differ from each other in production.

First: Homogeneity tests of variances in participants’ scores for /h/ and /ʕ/.

As with the previous tests, several tests were conducted in order to ensure the homoscedasticity of the dependent variable. Box’s test results rejected the null hypothesis that the covariance between the /h/ matrix and /ʕ/ matrix were equal (Box’s $M = 15.61, F(3, 259920) = 4.91, p < .05$). The exact $p$ value was .002 which is significant at an alpha level of .05 and non-significant using an alpha of .001. For consistency purpose, since the probability level of the previous analyses was set at .05, Box’s results of the homogeneity of /h/ and /ʕ/ scores was considered significant using the same alpha level (.05). In fact, although Box’s test could not support the null hypothesis, this does not mean it proved the alternative hypothesis which states that the matrices are heterogeneous.

The reason behind Box’s test rejection of the null hypothesis is that the error variances in participants’ scores for /ʕ/ were significantly unequal, as indicated in Levene’s test (Table 4.5). However, error variances of their scores in /h/ were homogenous ($p > .05$). The variances in scores for /ʕ/ were not equal but the variances in /h/ scores were equal; therefore, the covariance of scores in both /h/ and /ʕ/ were not homogeneous as shown in Box’s test.

The other homogeneity test, Mauchly’s test, could not be computed since there were only two matrices of scores. However, Epsilons indicated that the covariance among scores of the two
matrices (i.e. /ʕ/ matrix and /h/ matrix) were spherical (ε = 1), and this was true for Greenhouse-Geisser, Huynh-Feldt and Lower-bounded alike.

Table 4.5
Levene’s test of equality of error variances of /h/ and /ʕ/

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ʕ/</td>
<td>5.22</td>
<td>1</td>
<td>38</td>
<td>.028</td>
</tr>
<tr>
<td>/h/</td>
<td>2.94</td>
<td>1</td>
<td>38</td>
<td>.094</td>
</tr>
</tbody>
</table>

To sum up, there was an inconsistency in the results of the homogeneity tests. Box’s test could not reveal that the covariance matrices were equal (because the variances in scores for /ʕ/ were heterogeneous, as shown in Levene’s test). Epsilons, on the other hand, showed that these covariance matrices were spherical (homogenous). In such a case of inconsistency, and given the fact that studies using split-plot ANOVA had based their analysis on sphericity tests (Leech, Barrett & Morgan, 2008, pp. 155-161; and Kinnear & Gray, 2009, pp. 322-335), this study took into account only sphericity test results.

Second: Omnibus tests of within-subjects and between-subjects main effects.

The within-subjects factors were the pharyngeal consonants /h/ and /ʕ/ in which all participants were tested. As illustrated in Table 4.6, the difference in participants mean scores for /h/ and /ʕ/ was statistically significant: $F(1, 38) = 12.75, p < .05, \eta^2 = .25$, power = .94. Although this difference was significant, the effect size measurement indicated that only 25% of the difference in the given scores was accounted for by the variability between the phonemes (i.e. moderate effect size). The mean differences in participants’ Arabic proficiency had significantly a larger effect size: $F(1, 38) = 210.02, p < .05, \eta^2 = .85$, power = 1.
Third: Interaction between the Arabic proficiency and the two pharyngeal sounds.

If the variance and covariance matrices are claimed to be homogenous, the usual RM ANOVA results are valid. Otherwise (if they are proved heterogeneous), either a multivariate test or correction-associated $F$ must be reported. Assuming that the sphericity assumption held true (based on Epsilon value), I accounted for only usual RM ANOVA results. However, I reported the results of the multivariate tests and the corrected $F$-ratio in order to further support my argument that the covariance matrices were actually homogenous.

Split-plot ANOVA showed a significant difference in /h/ vs. /ʕ/ scores as they were interacting with the differences in participants’ Arabic proficiency: $F(1, 38) = 13.16, p < .05, \eta_p^2 = .26$, power = .94. The correction-associated $F$-ratios were obtained by multiplying $df1$ and $df2$ by Epsilon value. In the present case, since $\epsilon = 1$ (because the covariance matrices were spherical), the adjusted $F$-ratio was the same as the one produced by the usual ANOVA results.
For example, Huynh-Feldt’s correction showed the same results of the usual RM ANOVA \((F(1, 38) = 13.16, p < .05, \eta_p^2 = .26, \text{power} = .94)\). Table 4.6 illustrates this similarity between the usual RM ANOVA and the correction-associated results.

If homogeneity violation was to be accounted for, multivariate tests can be employed. All multivariate tests showed exactly the same results \((F(1, 38) = 13.16, p < .05, \eta_p^2 = .26, \text{power} = .94)\). These results were similar to the usual as well as adjusted RM ANOVA output.

In brief, even if Box’s tests results were taken into consideration, all multivariate tests and RM ANOVA indicated a significant difference in participants’ scores for /h/ vs. /ʕ/ as interacting with the difference in Arabic proficiency. Although the results were statistically significant, the effect size estimate showed that only 25% of the difference in their scores was associated with this interaction between the variables in question. However, the obtained \(F\)-ratio (which was greater than \(F_{crit} = 4.10\)) denoted that the difference in scores as associated with this interaction was greater than the expected error difference.

**Fourth: Post-hoc test of mean differences in both groups’ production of /h/ vs. /ʕ/**

As indicated by Fisher’s LSD, there was a significant mean difference between learners’ and non-learners’ production of /h/ and /ʕ/, with learners group scoring significantly higher: \(MD = 1.51, SE = .15, p < .05, 95\% \text{ CI} [1.21, 1.81]\). Furthermore, both groups’ mean production scores for /ʕ/ was significantly higher than the mean production scores for /h/: \(MD = 1.43, SE = .12, p < .05, 95\% \text{ CI} [1.19, 1.68]\).

Descriptive statistics of the mean scores that both groups received for /h/ and /ʕ/ further supported LDS’s results. It was found that Arabic learners’ mean score for pharyngeal /ʕ/ (\(M = \)
4.69, $SD = .48$) was almost similar to their mean score for the other pharyngeal consonant /h/ ($M = 4.70, SD = .46$). Non-learners’ mean scores for the two sounds were more distinct, with their mean score for /ʕ/ ($M = 2.88, SD = .95$) greater than their mean score for /h/ ($M = 1.82, SD = .76$). The overall mean score for /ʕ/ for both groups ($M = 3.79, SD = 1.18$) was also greater than the total mean score for /h/ ($M = 3.26, SD = 1.59$). These differences are presented in Figure 4.4. Further details of group differences in /h, ʕ/ production are provided in Appendix H4.

**Summary of group differences in the production of /h/ vs. /ʕ/.

In brief, as evidenced in the findings of RM ANOVA, the difference in speaker’s Arabic proficiency was associated with different production accuracy of /h/ compared to /ʕ/. The lack of Arabic experience is associated with more variability between /h/ and /ʕ/ production. In contrast, Arabic learners’ production of the two pharyngeal consonants showed similar accuracy levels.
4.2.5.2 **Group differences in the production of /tˤ/ vs. /sˤ/ vs. /dˤ/ vs. /ðˤ/.

The purpose was examining the difference between participants’ scores of different pharyngealized consonants as associated with the interaction with the difference of their Arabic proficiency.

**First: Homogeneity of variances in scores given to /tˤ, sˤ, dˤ, ðˤ/.

The homogeneity of participants’ scores in pharyngealized consonants was similar, in some respect, to that of the pharyngeal sounds. According to Mauchly’s test, the covariance matrices of the scores of these phonemes were spherical (Mauchly’s $W = .87, \chi^2(5) = 5.90, p > .05$). With respect to Epsilon values, Huynh-Feldt showed a value of 1 indicating that sphericity was met, and Greenhouse-Geisser was around 1.

However, Box’s test could not support the assumption of the equality of covariance matrices of pharyngealized consonants (Box’s $M = 35.71, F(10, 6903) = 3.16, p < .05$). Box’s test results can be interpreted in light of Levene’s test which proved the equality of error variances for only two consonants. According to Levene’s test, the variances in participants’ mean scores for /tˤ/ and /sˤ/ were homogenous (Levene’s $F(1, 38) = 3.20, p > .05$; and $F(1, 38) = .08, p > .05$, respectively). In contrast, the variances of the mean scores given to /dˤ/ and /ðˤ/ could not be assumed to be equal: $F(1, 38) = 6.24, p < .05$; and $F(1, 38) = 14.63, p < .05$, respectively. Hence, the variances in mean scores for two sounds were homogeneous but they might be heterogeneous for the other two sounds (as shown in Levene’s test); therefore, the covariance of the scores given to the four sounds together could not pass Box’s test.

Generally, there is some inconsistency in the results of Mauchly’s and Box’s tests. In compliance with studies on RM ANOVA (Leech et. al., 2008; and Kinnear et. al., 2009), my study emphasized Mauchly’s test results rather than Box’s test. Given this (and according to
Mauchly’s results), the assumption that the covariance matrices were spherical held true and, thus, I reported usual RM ANOVA results.

Second: Omnibus tests of within-subjects and between-subjects main effects.

Participants’ mean scores for the pharyngealized consonants /tˤ, sˤ, dˤ, ðˤ/ were significantly different: \( F(3, 114) = 46.17, p < .05, \eta_p^2 = .55, \text{power} = 1 \) (Table 4.7). The difference in participants’ Arabic proficiency was also significant at the same alpha level: \( F(1, 38) = 72.17, p < .05, \eta_p^2 = .66, \text{power} = 1 \).

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>( \eta_p^2 )</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonants (/tˤ, sˤ, dˤ, ðˤ/)</td>
<td>3</td>
<td>23.16</td>
<td>46.17</td>
<td>.00</td>
<td>.55</td>
<td>1.00</td>
</tr>
<tr>
<td>Consonants * Groups</td>
<td>3</td>
<td>4.60</td>
<td>9.17</td>
<td>.00</td>
<td>.19</td>
<td>1.00</td>
</tr>
<tr>
<td>Error (phonemes)</td>
<td>114</td>
<td>.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Third: Interaction between Arabic proficiency and the pharyngealized consonants.

The results obtained from RM ANOVA denoted a significant difference in participants’ scores for the four pharyngealized consonants as interacting with the difference in their Arabic proficiency: \( F(3, 114) = 9.17, p < .05, \eta_p^2 = .19, \text{power} = 1 \). The statistical power in this interaction, computed an alpha level of .05, was large. The difference between the obtained \( F \)-ratio and the critical value \( (F_{crit} = 2.69) \) was relatively small compared to the previous RM ANOVA tests. This means that the differences in the scores given to pharyngealized consonants were more susceptible to error than the differences of the other tests; however, this does not suggest that the error difference was larger than the observed difference in scores which resulted
from the variability in Arabic proficiency. Importantly, even if the variances were to be considered heterogeneous according to Box’s test, multivariate tests of interaction still showed high \( p \) value and large \( \eta_p^2 \). The output of these multivariate tests is provided in Appendix H5.

**Fourth: Post-hoc test of differences in groups’ production of the pharyngealized CVs.**

Despite the differences between the four pharyngealized consonants’ production, the learners’ mean score was significantly higher than non-learners’ mean score for the same sounds: \( MD = 1.47, SE = .17, p < .05, 95\% \text{ CI } [1.12, 1.82] \).

There were also significant mean differences between these consonants in terms of participants’ production; however, although participants’ mean production score for /tˤ/ was higher than /ðˤ/, this difference between /ðˤ/ and /tˤ/ was statistically non-significant (\( MD = .26, SE = .18, p > .05, 95\% \text{ CI } [-.11, .63] \). Among all pharyngealized consonants, the minimal significant difference was between the mean production score of /sˤ/ and /tˤ/—the production of /sˤ/ was more accurate than the production of /tˤ/: \( MD = .78, SE = .15, p < .05, 95\% \text{ CI } [.47, 1.09] \). Excluding the difference between /tˤ/ and /ðˤ/ which was not significant, the mean differences (\( MD \)) between all other pharyngealized consonants were greater than the mean difference between /sˤ/ and /tˤ/ (i.e. greater than .78). Appendix H6 illustrates the results of Fisher’s LSD for differences in the production of all pharyngealized consonants.

Learners’ mean score for each pharyngealized sound was greater than non-learners’ score for the respective sound. However, these differences were inconsistent—that is, the difference between the two groups’ mean scores for /tˤ/ was substantially different from the difference in their mean scores for /ðˤ/. This latter consonant gained the lowest mean score for both groups (\( M = 1.94, SD = .92 \)). Descriptive statistics table of their mean scores for each pharyngealized sound
is provided in Appendix H7, and graphical representation of these differences is displayed in Figure 4.5.

![Figure 4.5. Groups differences in the production of pharyngealized consonants](image)

**Summary of group differences in the production of pharyngealized sounds.**

In brief, homogeneity tests showed contradicting results. While Box’s test rejected the equality of the covariance matrices, Mauchly’s test indicated that these matrices were circular and homogenous. I argued that Mauchly’s test results should be taken into consideration for conducting split-plot ANOVA and, thus, the homogeneity of covariance matrices was assumed. Therefore, split-plot ANOVA—which yielded significant differences in participants’ production of the four sounds as they interacted with their Arabic proficiency—was reported. Even if the homogeneity of variances was not assumed, multivariate tests produced similar results to RM ANOVA but with a larger effect size: $F(3, 36) = 8.38, p < .05, \eta^2_p = .41$, power = .99.

In general, the difference in participants’ Arabic proficiency induced different variances in their ability to produce the pharyngealized consonants. Hence, the four pharyngealized
consonants were not the same in terms of Arabic learners’ and non-learners’ production. These differences between the sounds under investigation were statistically significant; however, the difference in the production of the four sounds that is associated with differences in Arabic proficiency was relatively small (as shown in the effect size measure.

4.2.6 Question IV: Differences in participants’ production of pharyngeal and pharyngealized consonants with different vowel combinations.

The goal of this question was to investigate whether the previous differences in English speakers’ production of pharyngealized and pharyngeal consonants was influenced by vowel manipulation. This question was also intended to examine the difference in production scores given to pharyngealized (and pharyngeal) consonants with different vowels as associated with the difference in producers’ Arabic proficiency. Since there was a significant difference in participants’ scores for pharyngealized versus pharyngeal consonants (as shown in Question II), the present question was decomposed into two secondary questions each of which tackled a separate sound category. A sub-question addressed vowel manipulation among pharyngealized consonants and another sub-question for pharyngeal ones.

4.2.6.1 Group differences in the production of pharyngealized CVs with different vowels.

The between-subjects factor was the same as used in the other analyses (i.e. Arabic proficiency); however, within-subjects factors were constructed by calculating mean scores:

- Pharyngealized consonant with /aː/ = (/tˤaː/ + /sˤaː/ + /dˤaː/ + /ðˤaː/) / 4 = Mean
- Pharyngealized consonant with /iː/ = (/tˤiː/ + /sˤiː/ + /dˤiː/ + /ðˤiː/) / 4 = Mean
- Pharyngealized consonant with /uː/ = (/tˤuː/ + /sˤuː/ + /dˤuː/ + /ðˤuː/) / 4 = Mean
Importantly, these vowels were similar in length, but not in shape or height. The vowel /aː/ had the features [high] and [-rounded] while /uː/ was [rounded], and /iː/ was [-high]. The present question was concerned with whether these vowel differences, when interacting with the differences in speakers’ Arabic proficiency, yielded differences in the production of the pharyngealized sounds. In other words, the goal was exploiting the differences between these vowel combinations among groups of different proficiency levels in Arabic. I conducted split-plot ANOVA to achieve this purpose after exploring the homogeneity of the variance and covariance of participants’ production scores in the given sounds.

First: Homogeneity of variances in participants’ scores for pharyngealized consonants with different vowels.

Box’s test rejected the homogeneity assumption among the covariance matrices (Box’s $M = 14.39, F(6, 10462.2) = 2.19, p < .05$). This violation of the homogeneity of covariance resulted from the inequality of error variances in participants’ scores for pharyngealized consonants accompanied by /uː/: Levene’s $F(1, 38) = 7.37, p < .05$. However, this test revealed the equality of variances in participants’ production scores for pharyngealized CVs with /aː/ as well as pharyngealized CVs with /iː/: $F(1, 38) = .46, p > .05$; and $F(1, 38) = .87, p > .05$. According to Levene’s test, error variances were equal for participants’ scores in pharyngealized consonants with /aː/ and /iː/, but not for the same consonants with /uː/ vowel.

Mauchly’s test result was also not significant at .05 alpha (Mauchly’s $W = .99, \hat{\chi}^2(2) = .32, p > .05$). Hence, the covariance matrices of participants’ scores in the three sound sub-categories were spherical and homogenous. In general, the covariance matrices were assumed to be circular according to Mauchly’s test. The results of Mauchly’s test in particular, were the basis for the results obtained from RM ANOVA.
**Second:** Omnibus tests of within-subjects and between-subjects main effects.

There was a significant mean difference in participants’ production scores for the three sound combinations—i.e. /aː, iː, uː/ preceded by pharyngealized consonants—with small effect size: $F(2, 76) = 5.55, p < .05, \eta^2_p = .13$, power = .84. Differences their Arabic proficiency (i.e. between-subjects factor) were also statistically significant beyond .05 alpha: $F(1, 38) = 72.17, p < .05, \eta^2_p = .66$, power = 1.

**Third:** Interaction between Arabic proficiency and the production of pharyngealized CVs with different vowels.

As shown in the omnibus test, the significant mean differences among scores for the three CV sets were true for both groups of participants despite their proficiency level in Arabic. The question of whether these differences held true when they were associated with the variability in Arabic proficiency was the main concern of split-plot ANOVA. As displayed in Table 4.8, these differences approached significance when they interacted with differences in Arabic proficiency: $F(2, 76) = 3.15, p = .05, \eta^2_p = .08$, power = .59. The effect size measure indicated that only 8% of the difference in participants’ scores for the three CV types was associated with the differences in Arabic proficiency. This test, in fact, produced the least effect size as compared to the previously reported RM ANOVAs. As Leech et. al. (2008) asserted, in the case that sphericity is met and Epsilons are high, multivariate tests “may be less powerful [and] less likely to indicate statistical significance”. This was true in the present case—that is, multivariate tests produced less significance than the RM ANOVA: $F(2, 37) = 2.89, p > .05, \eta^2_p = .14$. Although multivariate tests yielded a larger effect size than that of RM ANOVA, only the later test was taken into account in the subsequent discussion of these findings. The reason is that, RM ANOVA is preferred over multivariate tests when sphericity is met (Leech et. al., 2008).
Fourth: Post-hoc test of the mean differences in the production of pharyngealized consonants with different vowels.

As shown in LSD test, there was a significant mean difference between learners’ and non-learners’ production of pharyngealized consonants, with learners’ mean score higher than the other group: \( MD = 1.47, SE = .17, p < .05, 95\% \text{ CI } [1.12, 1.82] \).

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>( \eta^2 )</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sounds (pharyngealized CVs varying in vowel)</td>
<td>2</td>
<td>1.96</td>
<td>5.55</td>
<td>.01</td>
<td>.13</td>
<td>.84</td>
</tr>
<tr>
<td>Sounds * Group</td>
<td>2</td>
<td>1.11</td>
<td>3.15</td>
<td>.049</td>
<td>.08</td>
<td>.59</td>
</tr>
<tr>
<td>Error (phonemes)</td>
<td>76</td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Moreover, the mean differences in participants’ production of the three pharyngealized CV sets (pharyngealized consonants followed by either /aː, iː/ or /uː/) were significant, except the difference between /iː/ and /uː/ CV sets. To illustrate, the production of CV sequences that included /aː/ was significantly different from CVs with /iː/ or with /uː/. However, the production of CV sequences with /iː/ was not significantly different from the production of CV sequences with /uː/: \( MD = .14, SE = .13, p > .05, 95\% \text{ CI } [-.12, .41] \). In fact, despite the non-significance of the mean difference, participants’ production of pharyngealized consonants followed by /iː/ was more accurate than their production of the same consonants with /uː/. Appendix H8 provides detailed results of LSD’s comparisons between the three sets of pharyngealized CVs.
Arabic non-learners’ production of pharyngealized phonemes was less susceptible to vowel manipulation than learners’ production. That is, the mean differences in the formers’ scores for pharyngealized consonants with three vowel types were less discrepant than the mean differences in learners’ scores. Appendix H9 summarizes these differences in mean scores. Non-learners’ mean scores were generally lower than learners’ mean scores for the three sound subcategories under investigation. When the pharyngealized phonemes were accompanied by the vowel /u:/, both groups gained the lowest score (Figure 4.6).

![Figure 4.6. Groups differences in the production of pharyngealized consonants with different vowel combinations](image)

Summary of group differences in producing pharyngealized CVs with different vowels.

In brief, split-plot ANOVA was conducted to examine the differences in participants’ production scores for pharyngealized consonants with different vowel combinations, and to test the significance of these differences when interacting with the difference in speakers’ Arabic proficiency. It was found that both Arabic learners and non-learners exhibited different production accuracies when changing the vowels that accompanied pharyngealized consonants. These differences approached statistical significance. Also, although the production of these
sound categories was different for both groups, Arabic proficiency resulted in only 8% of these production differences (as indicated by the partial eta squared).

4.2.6.2 Group differences in the production of pharyngeal CVs with different vowels.

This test was concerned with Arabic learners’ and non-learners’ production of pharyngeal consonants when changing the adjacent vowel. Similar to the previous test on pharyngealized consonants, I computed the repeated-measures (CV sets) in the following way:

- Pharyngeal sound with /aː/ = (/hraː/ + /ʕaː/) / 2 = Mean
- Pharyngeal sound with /iː/ = (/hiː/ + /ʕiː/) / 2 = Mean
- Pharyngeal sound with /uː/ = (/huː/ + /ʕuː/) / 2 = Mean

First: Homogeneity tests for variances in participants’ scores for pharyngeal CVs with different vowels.

The variances in participants’ production scores for two of these sub-categories were equal; however, the variances in their scores for the first sub-category (pharyngeal consonants with /aː/ vowel) were significantly heterogeneous (Levene’s $F(1, 38) = 22.35, p < .05$). Therefore, the homogeneity of scores covariance (in these three sound sub-categories together) was rejected according to Box’s test ($M = 28.85, F(6, 10462.2) = 4.39, p < .05$). In contrast, Mauchly’s test indicated that these covariance matrices of the same CVs were circular ($M = .92, \chi^2(2) = 3.11, p > .05$). Hence, I reported the RM ANOVA results associated with the sphericity assumption.

Second: Omnibus tests of within-subjects and between-subjects main effects.

There were significant differences in participants’ mean scores for the three sound sub-categories under investigation ($F(2, 76) = 5.0, p < .05, \eta^2_p = .12$, power = .8). The differences in
their Arabic proficiency (between-subjects factor) were also significant at the same alpha level
\( (F(1, 38) = 210.02, p < .05, \eta_p^2 = .85, \text{power} = 1) \).

Third: Interaction between Arabic proficiency and the production of pharyngeal consonants with different vowels.

Learners’ and non-learners’ production of pharyngeal consonants differed as the vowel attached to the consonant changed. Both groups mean scores for pharyngeal consonants associated with /uː/ \((M = 3.34, SD = 1.47)\) were lower than their mean scores for the same consonants with /aː/ or /iː/ \((M = 3.79, SD = 1.29; \text{and } M = 3.44, SD = 1.44, \text{respectively})\). Generally, learners’ production of the three CV sets was more accurate than non-learners’ production. Appendix H10 presents detailed mean scores for the three pharyngeal sets. Above all, these differences between the two groups of participants were statistically non-significant, as split-plot ANOVA revealed: \( F(2, 76) = .43, p > .05, \eta_p^2 = .01, \text{power} = .12 \).

Fourth: Post-hoc test of group differences in the production of pharyngeal CVs with different vowels.

Fisher’s LSD showed that learners’ mean score for all three sets of sounds was significantly higher than non-learners’ mean score: \( MD = 2.35, SE = .16, p < .05, 95% \text{CI} [2.02, 2.67] \). The test also showed that, for both groups, the production difference between /aː/ and /iː/ sets as well as the difference between /aː/ and /uː/ sets were both significant. Nevertheless, this statistical significance was ruled out in the ANOVA test because the \( p \) value of the non-significant difference between /iː/ and /uː/ sets was very large so that it could rule out the significant \( p \) value of the other comparisons.

This post-hoc analysis for pharyngeal CVs with varying vowels yielded similar results to the test on pharyngealized CVs with varying vowels. For instance, the mean production score for
pharyngeal CVs with the vowel /aː/ was significantly higher than the mean score for the same sounds with /iː/ or /uː/. The same difference was observed among pharyngealized CV sets that differed in the adjacent vowel. Furthermore, similar to the results of pharyngealized CVs, the mean difference between pharyngeal CVs with /iː/ and pharyngeal with /uː/ was not statistically significant: $MD = .11, SE = .17, p > .05, 95\% CI [-.23, .44]$. Appendix H11 demonstrates detailed results obtained from Fisher’s LSD post-hoc test for the three pharyngeal sets. It was this non-significance between CV with /iː/ and CV with /uː/ that caused split-plot ANOVA to produce non-significant results.

In brief, as illustrated in Figure 4.7, there were differences between the two participant groups in terms of their production of pharyngeal consonants as the adjacent vowels changed (i.e. /ħaː, ʕaː/ vs. /hiː, ʕiː/ vs. /huː, ʕuː/). However, these differences were minimal and statistically not significant.

Figure 4.7. Groups’ differences in the production of pharyngeal consonants with different vowel combinations
4.3 Non-native Arabic speakers’ production patterns

In this section, I explored the most frequently occurring sounds perceived by Arabic L1 listeners and produced by English speakers in response to pharyngeal and pharyngealized stimuli. Since the sound production of the learners and non-learners was completely different (as the statistical tests revealed), I discussed the production patterns of each participant group separately. I highlighted the most frequently reported sounds in which the English speakers intended to shadow each consonant. In cases where the three cohorts /ħaː, hiː, huː/ induced different production responses, I discuss the responses for each cohort separately. If a stimulus yielded multiple, different responses, I highlighted only the ones that Arabic listeners reported most frequently, with reference to the percentage of occurrence of each response. The percentage of a response occurrence takes into consideration both Arabic perceivers’ judgements of the sound response. For example, if Perceiver 1 reported perceiving the stimulus /haː/ produced as /ʔ/ ten times by the non-learners, and Perceiver 2 reported the same response for the same stimulus 20 times, the percentage of /ʔ/ occurrence is 75% for that particular group of participant. The same principles applied to all target sounds produced by both learners and non-learners of Arabic.

4.3.1 Arabic non-learners’ sound responses to pharyngealized stimuli /tˤ, sˤ, dˤ, ðˤ/.

The pharyngealized stimuli induced the most diverse responses among the non-learners. A matrix of detailed responses for all pharyngealized cohorts is provided in Appendix I1. There were some variations among the four pharyngealized consonants /tˤ, sˤ, dˤ, ðˤ/ in terms of the number of frequent sound responses. For instance, while there were only three response sounds for /sˤ/, there was a total of 15 sound responses to /tˤ/ as perceived by both Arabic perceivers. Summary of non-learners’ sound shadowing is presented in Table 4.9.
4.3.1.1 Non-learners’ shadowing responses to /tˤ/.

There was a wide variety of sounds that the non-learners produced in response the three cohorts of /tˤ/; however, in about 24.2% of the cases they produced it correctly. The most recurrent wrong responses for this consonant were /t, d/ and /ʔ/ which occurred 16.7; 13.3; and 5, respectively. The responses for the three cohorts /tˤaː, tˤiː, tˤuː/ were similar to some extent; nonetheless, the most noticeable difference among them is that /tˤaː/ resulted in the least number of correct shadowing (17.5%) and was produced as /ð/ in 12.5% of the cases.

Table 4.9
Summary of non-learners’ responses to pharyngealized stimuli /tˤ, sˤ, dˤ, ðˤ/

<table>
<thead>
<tr>
<th>Sound</th>
<th>English L1 speakers’ sound responses (% of occurrence)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Same sound</td>
</tr>
<tr>
<td>/tˤ /</td>
<td>24.2</td>
</tr>
<tr>
<td>/sˤ /</td>
<td>47.5</td>
</tr>
<tr>
<td>/dˤ /</td>
<td>4.2</td>
</tr>
<tr>
<td>/ðˤ /</td>
<td>22.5</td>
</tr>
</tbody>
</table>

*Percentages of recurrent responses are presented in the same order in which the sound responses are presented.

4.3.1.2 Non-learners’ shadowing responses to /sˤ/.

The pharyngealized fricative /sˤ/ induced the least variant responses compared to other pharyngealized consonants. Non-learners’ responses to its three cohorts were similar to a great extent. That is, they produced all cohorts as either /sˤ/ or /s/, except in one case in which one Arabic listener perceived /sˤiː/ as /ziː/. For all cohorts, the percentages of /sˤ/ and /s/ reoccurrence...
were comparable—that is, whereas /sˤ/ responses occurred 47.5% of the times, /s/ was produced 51.7% of the times in response to the same stimuli.

**4.3.1.3 Non-learners’ shadowing responses to /dˤ/.

Contrary to /sˤ/ shadowing responses, the pharyngealized /dˤ/ resulted in the most deviant shadowing compared to the other pharyngealized consonants. The most dominant sound response for /dˤ/ was the plain /d/ which was perceived in 26.7% of the shadowing responses for all cohort exemplars (i.e. in response to /dˤaː, dˤiː, dˤuː/ stimuli in general).

However, the three cohorts yielded completely different shadowed sounds. For instance, non-learners never produced /dˤaː/ correctly or as /daː/; instead, they produced it as /baː/ (47.5%). Importantly, non-learners produced pharyngealized, yet wrong sounds in response to the stimulus /dˤaː/; namely, in 25% of cases, they articulated it as either /bˤaː/ or /vˤaː/ (i.e. each of these response sounds occurred 12.5% times).

With respect to the other cohorts, the /dˤiː/ and /dˤuː/ provoked the correct responses only 5 and 7.5%, respectively. A remarkable difference between these two cohort exemplars (i.e. /dˤiː, dˤuː/) was that the most frequently shadowed sound for /dˤuː/ was /d/ which appeared in more than half of the times. The exemplar /dˤiː/ yielded more variant shadowed sounds including /t/ which appeared about 17% of the times.

**4.3.1.4 Non-learners’ shadowing responses to /ðˤ/.

Across its three cohorts, the pharyngealized /ðˤ/ yielded pharyngealized sound responses which were dissimilar to the stimulus in most cases. The most frequently shadowed sound was a pharyngealized fricative /vˤ/ which does not exist in Arabic and which appeared in approximately half of the cases. At the cohort level, this novel shadowed sound /vˤ/ was repeated 62.5% for /ðˤaː/ exemplar; 55% for /ðˤuː/ stimulus; and 25% for /ðˤiː/ exemplar. The most prevalent
shadowed sounds for the three cohorts were voiceless fricatives /vˤ/; /ðˤ/; and /ð/, with varying percentages of reoccurrence.

4.3.2 Arabic non-learners’ sound responses to pharyngeal stimuli /h, ʕ/.

The pharyngeal consonants (and their cohorts) provoked less variant shadowing responses compared to the pharyngealized consonants (Table 4.10). Appendix I2 provides a comprehensive overview of the most dominant shadowed sounds that non-learners produced to imitate all pharyngeal cohorts.

4.3.2.1 Non-learners’ shadowing responses to /h/.

The non-learners group produced the pharyngeal consonant /h/ correctly in one fifth of the times, and the most frequent shadowed response was /h/ (55.8%). A remarkable difference among its three cohorts /ħaː, ħiː, ħuː/ was that, while /ħaː/ stimulus induced mainly two types of sound responses (/ħaː/ and /ha:/), the other cohorts /hiː/ and /huː/ resulted in diverse responses. Responses to these latter cohorts included /x, k, h/ and other less frequent responses.

Table 4.10
Summary of non-learners’ production of pharyngeal stimuli /h, ʕ/

<table>
<thead>
<tr>
<th>Sound</th>
<th>Arabic L1 listeners’ perception (% of occurrence)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct response</td>
<td>/ʔ/</td>
<td>/h/</td>
<td>Recurrent responses</td>
</tr>
<tr>
<td>/h/</td>
<td>20</td>
<td>0</td>
<td>55.8</td>
<td>/h;/l/</td>
</tr>
<tr>
<td>/ʕ/</td>
<td>55</td>
<td>37.5</td>
<td>1.7</td>
<td>/ʃ/</td>
</tr>
</tbody>
</table>

*Percentages of recurrent responses are presented in the same order in which the sound responses are presented.

4.3.2.2 Non-learners’ shadowing responses to /ʕ/.

The other pharyngeal consonant /ʕ/ had also less obvious differences between its cohorts /ʕaː, ʕiː, ʕuː/. Most of non-learners’ production of these sounds were either /ʕ/ or /ʔ/ (55 and
37.5%, respectively). Other responses that were neither /ʕ/ nor /ʔ/ were less common; for example, the most frequently produced sound (other than /ʕ/ and /ʔ/) was /r/ which appeared only 3.3% of all cases in this group.

### 4.3.3 Arabic learners’ sound responses to pharyngeal stimuli /h, ʕ/.

Both pharyngeal consonants /h, ʕ/, and their cohorts, were similar to each other in terms of the percentage of correct productions as well as the limitedness of the number of deviant responses. Precisely, Arabic learners produced /h/ 95% correctly, and /ʕ/ 92.3% correctly (Appendix I3). The non-pharyngeal /h/ (4.2%) was the most frequent wrong production observed for /h/ cohorts. Similarly, the non-pharyngeal /ʔ/ was found in 3.3% of the responses to pharyngeal /ʕ/.

### 4.3.4 Arabic learners’ sound responses to pharyngealized stimuli /tˤ, sˤ, dˤ, ðˤ/.

As shown in Table 4.1, a limited variability in learners’ production of pharyngealized consonants was found. Except for /dˤ/ exemplars, most sounds produced by the learners were either the correct pharyngealized sound or its plain counterpart; nonetheless, in some cases, the voicing counterpart of the non-pharyngealized consonant was observed (Appendix I4 includes detailed production patterns).

#### 4.3.4.1 Learners’ shadowing responses to /tˤ/.

Among the English learners of Arabic, the total percentage of correct productions of all /tˤ/ cohorts was about 93, with correct responses for each cohort ranging from 90 to 95%. Given this, in only 5 or 10% of the times, these participants produced a sound that diverged from the exemplar. These divergent sounds included the plain counterpart /t/ (2.5%); /q/ (1.7%); and /d/ (.8%). It is important to note that .8% is the least reoccurrence percentage given the study sample
size. Precisely, this percentage indicates that only one Arabic listener perceived the sound produced as such by one English speaker.

Table 4.11
Summary of learners’ responses to pharyngealized stimuli /tˤ, sˤ, dˤ, ðˤ/

<table>
<thead>
<tr>
<th>Sound</th>
<th>Arabic L1 listeners’ perception (% of occurrence)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Same sound</td>
</tr>
<tr>
<td>/tˤ/</td>
<td>93.3</td>
</tr>
<tr>
<td>/sˤ/</td>
<td>88.3</td>
</tr>
<tr>
<td>/dˤ/</td>
<td>26.7</td>
</tr>
<tr>
<td>/ðˤ/</td>
<td>63.3</td>
</tr>
</tbody>
</table>

Percentages of recurrent responses are presented in the same order in which the sound responses are presented.

4.3.4.2 Learners’ shadowing responses to /sˤ/.

Learners group produced the pharyngealized /sˤ/ as either /sˤ/ or /s/ only, and no divergent productions were observed. The learners group was similar to the non-learners in the sense that both groups’ shadowing of /sˤ/ fell in the same sound category (i.e. /sˤ/ or /s/). For the learners group, no obvious differences between the three consonant cohorts were found, except for the fact that /sˤa:/ resulted in 100% correct productions whereas /sˤi:/ and /sˤu:/ provoked 80 and 85% correct responses, respectively.

4.3.4.3 Learners’ shadowing responses to /dˤ/.

The most obvious pharyngealization deviations among Arabic learners were found in the production of /dˤ/. Learners articulated this consonant correctly in approximately one quarter of the cases. There were noticeable differences between the three cohorts in terms of the correct responses they induced: /dˤi:/ had the least number correct productions (only 5%), while /dˤa:/
and /dˤuː/ yielded 22.5 and 52.5% correct shadowing, respectively. The most dominant sounds produced by Arabic learners in response to all /dˤ/ cohorts, in general, were /tˤ/ (36.5%) and /ðˤ/ (15%), both of which were pharyngealized.

**4.3.4.4 Learners’ shadowing responses to /ðˤ/.

Compared to /dˤ/, about two thirds (exactly 63%) of learners responses to /ðˤ/ were correct despite the differences among its cohorts. The sound responses that diverged from /ðˤ/ were /ð/ (15%); /tˤ/ and /ðˤ/ (4.2% each); /θˤ/ (.8%); and /vˤ/ (3.3%), all of which were pharyngealized, except the plain counterpart of the target sound (i.e. /ð/).

**4.4 Chapter summary

**4.4.1 Summary of split-plot ANOVA findings for the four research questions.

The aim of the previously reported statistical findings was to find out differences in the production of pharyngeal and pharyngealized consonants (and their sub-categories) as induced by varying proficiency levels in Arabic. In order to achieve this goal, I conducted split-plot ANOVA after examining the homogeneity of the variance and covariance of participants’ production scores.

To reiterate, the shadowed sounds were produced by English L1 speakers, both learners and non-learners of Arabic. The production accuracy was determined in terms of Arabic L1 listeners’ perception of the divergence between the shadowed sound and the stimulus (i.e. the exemplar sound that a native Arabic speaker produced). Thus, the exemplar sounds were the benchmarks against which participants’ production ability was assessed using a one-to-five numerical scale.

It was found that differences in non-native Arabic speakers’ proficiency in the language yielded different production accuracies of the sounds under investigation. These differences were
statistically significant for certain sound categories, but not so for other categories. The following is a brief recapitulation of the tests on sound production differences as associated with differences in Arabic proficiency.

There was a dramatic, significant difference between pharyngealized and non-pharyngealized consonants production, with remarkable intra-group differences. Arabic proficiency seems to influence the variation between pharyngealized and non-pharyngealized sounds; however, the production of the plain counterparts was less susceptible to Arabic proficiency variation (Section 4.2.3.1).

Differences in Arabic proficiency yielded significantly different production variations for pharyngeal consonants and their counterparts. To illustrate, group differences in the production of non-pharyngeal consonants were not similar to the group differences in pharyngeal consonants. While the difference in proficiency level seemed to substantially influence the production of pharyngeal consonants, it was less so for non-pharyngeal sounds (Section 4.2.3.2).

There were significant group differences in the production of pharyngeal versus pharyngealized consonants. Compared to the production of pharyngealized sounds, the production of pharyngeal sounds was more susceptible to variation in Arabic proficiency (Section 4.2.4).

There were also significant differences between the two pharyngeal consonants /ʰ, ꝏ/ in terms of learners’ and non-learners’ production. The two groups showed different variations between /ʰ/ and / ꝏ/. Namely, Arabic learners production correctness of /ʰ/ was comparable to / ꝏ/ while the non-learners’ production of one sound differed from the other sound significantly (Section 4.2.5.1).
Similarly, the differences in the two groups’ production of the four pharyngealized consonants /tˤ, sˤ, dˤ, ðˤ/ were significant at .05 probability level. Arabic proficiency was associated with higher production scores for all of these sounds. Importantly, the within-category variation for the learners group was similar to the variation in non-learners’ production. This means that both groups showed similar production differences between the four sounds (Section 4.2.5.2).

The last research question addressed the production differences between pharyngeal (and pharyngealized) consonants when changing the adjacent vowel. Although there was a difference between learners’ and non-learners’ production of /haː, qaː/ vs. /hiː, qiː/ vs. /huː, quː/, this difference was not significant. When comparing these vowel effects on pharyngealized consonants, the difference between learners’ and non-learners production approached statistical significance ($p = .049$). Regardless of the (non)-significance of these differences, production accuracy for both groups dropped noticeably when the pharyngeal and pharyngealized consonants were followed by /uː/ (Section 4.2.6).

4.4.2 Summary of learners’ and non-learners’ shadowing responses for pharyngeal and pharyngealized stimuli.

For both participant groups, the pharyngeal consonants induced less variant sound responses than the pharyngealized consonants. However, the pharyngealized /sˤ/ was an exceptional case among the other pharyngealized consonants. That is, this target sound evoked the least number of deviant responses, and most of shadowed responses fell in the same category /sˤ, s/. The non-learners group produced a wider variety of sounds (for both pharyngeal and pharyngeal stimuli) than the learners group. For instance, learners articulated the target /tˤ/ either
correctly or as /t/ whereas the non-learners exhibited more responses (such as /tˤ, t, d, ð/ and /ʔ/) in their attempt to imitate the same target sound /tˤ/.

Non-learners’ deviations from the pharyngealized stimuli were remarkable. Namely, multiple pharyngealized non-Arabic consonants were produced in response to /tˤ, dˤ, ðˤ/. These divergent responses included /vˤ, bˤ, θˤ/. The pharyngealized /vˤ/, for example, was observed in more than half of non-learners’ responses to /ðˤ/ cohorts. Also, about 4.2% of non-learners’ responses to /dˤ/ cohorts were perceived by Arabic listeners as /bˤ/ which does not exist in Arabic phoneme inventory. Contrary to learners’ production, some uvular, pharyngealized consonants were recurrent in non-learners’ attempt to shadow the pharyngealized target sounds.
Chapter V

Discussion and implications

In the preceding chapter, a statistical analysis of the findings was reported for each of the research questions separately in addition to the analysis of the most common production responses. This chapter consolidates the findings of the previous chapter and discusses them in light of the speech learning model (SLM, Flege, 1995) and perceptual assimilation model (PAM, Best, 1995). The discussion also accounts for the differences in native phonological space of the NL (English) and the TL (Arabic), which forms the basis for PAM and SLM. Moreover, I explain some theoretical and practical implications of the study results that provide insights on the research and practice of learning Arabic phonology. I also highlight some limitations to the study at the data collection and analysis levels, and offer some recommendations for future research in the field.

5.1 Overview of the study

This study aimed at exploring the learnability of Arabic phonology by examining the production accuracy among non-native Arabic speakers who vary in their Arabic proficiency level. The target Arabic sounds of this study were the pharyngeal /h, ʕ/ and pharyngealized consonants /tˤ, sˤ, dˤ, ðˤ/ and the plain counterparts of both (i.e. /h, ʔ/ and /t, s, d, δ/, respectively). The study was also concerned with differences between different sound categories.

I conducted both split-plot ANOVA to test the differences between sound categories using the numerical data that Arabic L1 perceivers provided. The results were further elaborated
using a post-hoc test, Fisher’s LSD. All split-plot ANOVA tests yielded significant differences between Arabic learners’ and non-learners’ sound production, except the two tests of vowel effect which approached significance. In most cases, the learners’ group surpassed the non-learners in their sound production except in the case of non-pharyngeal sounds. However, the magnitude of the difference between the two groups varied according to the sound category that was being tested.

Moreover, using descriptive statistics, I examined the most recurrent sounds that participants produced in response to each sound class, as evidenced by Arabic listeners’ perception. In this chapter, I aligned these findings with the split-plot ANOVA output to enrich the discussion of the differences between sound categories.

5.2 Discussion of the results

This section consolidates the findings of sound responses (presented in section 4.3) with the output of ANOVA and multivariate tests. Whenever relevant, I interpreted the findings in light of Best’s (1995) PAM and Flege’s (1995) SLM. I delineated this comprehensive discussion according to the four research questions, and I also unfolded the interrelatedness of these questions. Furthermore, according to the findings, I outline some predictions about Arabic learners’ potential for learning the sounds under investigation.

5.2.1 Production of pharyngeal and pharyngealized consonants and their counterparts.

I segregated the comparison of pharyngeal and non-pharyngeal sounds production from the comparison of pharyngealized sounds and their counterparts. The reason for such segregation was based on my prediction that the ability to produce pharyngeal consonants would be different from the pharyngealized ones.
5.2.1.1 *Production of pharyngealized consonants and their counterparts.*

Consistent with the initial predictions, higher proficiency level in Arabic led to accurate production of pharyngealized consonants /tˤ, sˤ, dˤ, ðˤ/. However, although the plain counterparts (i.e. /t, s, d, ð/) exist in English phonemic inventory, non-learners’ production was unexpectedly less authentic than learners’ production. This might have resulted from a confounding variable. It is likely that the context, in which the plain consonants occurred, influenced Arabic perceivers’ assessment. That is, non-learners’ production inaccuracy of most target sounds (pharyngeal and pharyngealized sounds) might have *impressionistically* caused the Arabic perceivers to provide low rating scores for the plain sounds even if they did not diverge perceptually from the exemplars. In general, I found that exposure to Arabic caused greater improvement in the production of pharyngealized than plain consonants.

There are some possible explanations of the production differences between pharyngealized consonants and their counterparts. First, it can be justified from an articulatory perspective based on the determinants of the native phonological space. The difference can be also explained from a perceptual perspective based on PAM assumptions and SLM’s hypotheses. In what follows, I provide a thorough discussion of both explanations.

*First explanation of production differences between pharyngealized and plain sounds.*

This production difference might have stemmed from the differences between Arabic and English phonological spaces, in general, and their gestural dynamics, in particular. TL sounds whose gestural dynamics resemble those of the NL are easier to learn than those sounds whose gestures are quite deviant from the NL.

On the one hand, the phonetic gestures that are harnessed in the production of Arabic plain consonants are similar, to some extent, to the gestures used in producing English /t, s, d, ð/.
This similarity of phonetic gestures in both languages justifies participants’ high production accuracy of /t, s, d, ð/. Nonetheless, I cannot assume that they are completely similar because there is no clear evidence in literature that, for example, the timing of the release of the alveolar constriction in /u/ is similar in both languages.

On the other hand, the pharyngealized consonants which induced production difficulty differ from English sounds /t, s, d, ð/ in terms of the gestural constellations and the temporal features of their phonetic gestures. To illustrate, the gestural constellations of Arabic pharyngealized consonants are quite different from those of English /t, s, d, ð/. These Arabic sounds require a combination of multiple phonetic gestures including a coronal constriction (either alveolar for /tˤ, sˤ, dˤ/ or dental for /ðˤ/); a posterior stricture between the tongue root and the back wall of the pharynx; elevation of the hyoid bone and raising the larynx. This constellation of gestures is not harnessed in the production of English consonants /t, s, d, ð/ which require only the anterior constriction; therefore, English speakers encountered difficulty in integrating these gestures together.

Another source of difference between pharyngealized consonants and English sounds pertains to the temporal characteristics of the pharyngeal gestures in both languages. There is no evidence of the timing of pharyngeal gestures used in producing Arabic consonants and English pharyngeal /a/. My argument, therefore, emphasizes the relative timing of the pharyngeal gesture as related to the anterior constriction in Arabic pharyngealization. According to Laufer et. al., (1988), the tongue root starts retracting to the pharynx before the formation of the oral constriction and persists after the release of the oral stricture. In contrast, the timing of the formation and release of the pharyngeal stricture in English /a/ is not dependent on (or related to)
any anterior constriction. These timing differences can explain the difficulty that English non-learners of Arabic encountered in articulating Arabic pharyngealization.

The production patterns presented in Appendix E1 further support my argument. In cases of wrong response to the pharyngealized stimuli, non-learners’ articulation either matched the anterior or the pharyngeal stricture. This is evident in examples where participants produced the plain counterpart instead of the pharyngealized consonant, and in cases where they produced a pharyngealized consonant with different anterior constriction. For example, they responded with pharyngealized /bˤ, θˤ/ and /vˤ/ which do not exist in Arabic phoneme inventory. These findings (i.e. the misarticulation of either the anterior or posterior constriction) indicate that these participants were able to form a pharyngeal constriction but were unable to integrate it with the anterior gesture (i.e. gestural constellation). Therefore, it was not the pharyngeal gesture in itself that caused the production difficulty; rather, it was the coordination of multiple, simultaneous gestures that caused non-learners’ misarticulation of pharyngealized consonants. This inability to coordinate the right phonetic gestures simultaneously could diminish be increasing the speakers’ exposure to sufficient perceptual input and by adequate pronunciation training—as evidenced in the authenticity of pharyngealization production among the learners group.

Second explanation of production differences between pharyngealized and plain sounds.

The second explanation of the difference between learners’ and non-learners’ production of pharyngealized consonants and the plain ones is based on the assumptions of PAM and SLM. I explained non-learners’ production in light of PAM specifically (because this model is designated to naïve perception) and learners’ production in light of SLM which tackles ultimate sound attainment.
Learners’ production of the pharyngealized consonants and their plain counterparts was consistent with SLM’s second and third hypotheses (H2 and H3)—which state that when the phonetic dissimilarity between NL and TL sounds increases, the likelihood that learners form a new phonetic category for the TL sound also increases. Hence, since the learners were able to discern the perceptual dissimilarity between the pharyngealized consonants and the closest English sounds /t, s, d, ð/, they had established new phonetic categories that represent Arabic pharyngealized consonants. This category formation might have contributed to learners’ high accuracy in producing the pharyngealized sounds.

PAM’s category-goodness assimilation may justify non-learners’ production of these sound categories. This assimilation pattern accounts for cases in which the two TL segments (which constitute a contrastive pair) are phonetically similar to a NL phonetic category; however, one member of the TL pair is more discrepant from the NL category than the other member. Hence, based on the magnitude of difference between English sounds and each member of Arabic contrastive pairs (i.e. /tˤ, t/; /sˤ, s/; /dˤ, d/ and /ðˤ, ð/), I speculated that the plain Arabic members were less discrepant from the English sounds than the pharyngealized member. For example, although both members of the Arabic pair /tˤ, t/ correspond to a single English category (i.e. /t/), the pharyngealized segment /tˤ/ is more deviant from the English category than plain /t/.

Two findings support my speculation. First, non-learners were able to articulate the plain members of Arabic contrastive pairs /t, s, d, ð/, which may indicate that these sounds were good exemplars of the English counterparts /t, s, d, ð/. Second, more than a quarter (26.7%) of non-learners’ responses to pharyngealized exemplars were non-pharyngealized; therefore, it is likely that they assimilated Arabic pharyngealized consonants to the same native category (i.e. to English categories /t, s, d/ and /ð/). Given these two observations and in light of PAM’s
category-goodness pattern, non-learners might assimilate the two members of Arabic contrastive pairs (/tˤ, t/; /sˤ, s/; /dˤ, d/ and /ðˤ, ð/) to a single English category (/t/; /s/; /d/ and /ð/, respectively). Therefore, their ability to discriminate between the Arabic contrastive phonemes was moderate (according to PAM’s predictions), which resulted in low production accuracy for the pharyngealized consonants.

It is important to emphasize that this interpretation is based on indirect observations derived from the production test. In essence, the study did not test participants’ perceptual discrimination between pharyngealized and non-pharyngealized consonants.

*Sounds’ learnability: Pharyngealized and non-pharyngealized consonants.*

These findings suggest that, at advanced stages of Arabic learning, it is likely that learners demonstrate improved ability in articulating both sound categories. Importantly, unlike the early learning stages, the ability to produce pharyngealized sounds at advanced learning levels is comparable to the ability of articulating the plain ones.

**5.2.1.2 Production of pharyngeal consonants and their counterparts.**

As expected, exposure to Arabic led to improved ability to produce the pharyngeal consonants /ħ,ʕ/. However, non-learners surpassed the learners group in the production of non-pharyngeal sounds. This was contrary to the initial predictions which state that lack of exposure to the language would not influence the production of non-pharyngeal consonants. Given this, increased proficiency in Arabic led to more considerable improvement in producing pharyngeal consonants than non-pharyngeal ones. In the following discussion, I explain the reasons for the production differences between pharyngeal and non-pharyngeal consonants that both learners and non-learners exhibited.
Non-learners’ production of pharyngeal versus non-pharyngeal consonants.

Evidence from non-learners’ responses to these two Arabic categories indicated that, in a great number of instances, each pharyngeal consonant and its counterpart were assimilated to the same English sound. That is, both segments of the Arabic pair /ħ, h/ were produced as English /h/; and /ʕ, ʔ/ were produced as /a/. Nonetheless, the pharyngeal segments induced a greater production difficulty than their counterparts. In what follows, I provide three possible explanations of this observation: from auditory and articulatory perspectives and in light of PAM.

At the auditory level, Arabic pharyngeal consonants /ħ, ʕ/ exhibit more divergence from English /h, a/, respectively, than Arabic non-pharyngeal counterparts.

From an articulation perspective, these sounds differ in terms of the degree of constriction used in producing them. The underlying principle of this justification is that, at a given place of articulation, the degree of constriction—that is narrower than the degree of constriction harnessed in a language—does not fall within the language’s phonological space. Hence, the narrow pharyngeal approximation of Arabic pharyngeal consonants /ħ, ʕ/ does not belong to English phonological space to which participants were attuned.

For instance, the phonetic gestures of Arabic non-pharyngeal /h/ are quite similar to those of English /h/, whereas the narrow pharyngeal approximation of Arabic /h/ is not harnessed in English. Therefore, it was more difficult for English speakers to articulate the pharyngeal /h/ than its non-pharyngeal counterpart.

The same reasoning applies to the difference between the Arabic pair /ʕ, ʔ/ and English vowel /a/. To illustrate, Arabic pharyngeal consonant /ʕ/ requires a narrower constriction than Arabic glottal stop /ʔ/ (Esling, 1999), and this pharyngeal degree of constriction (in pharyngeal
/ʕ/) is not harnessed in the articulation of English /a/. Therefore, the pharyngeal sound /ʕ/ was more difficult to articulate than its counterpart.

In addition, the assimilation patterns of Arabic pharyngeal sounds and their counterparts exemplified PAM’s category-goodness assimilation type. Evidence from non-learners’ production responses for these two Arabic sound categories showed that both segments of the Arabic contrastive pair /ħ, h/ were assimilated to a single English category /h/. Likewise, the two Arabic contrastive segments /ʕ, ʔ/ were assimilated to a single English category /a/. However, in both cases (i.e. the case of /ħ, h/ as well as /ʕ, ʔ/), the pharyngeal segment was more divergent from the English category than the non-pharyngeal segment.

In fact, PAM cannot explain the production difference between pharyngeal and non-pharyngeal sounds because it predicts a “very good” discriminability between them. In contrast, evidence from participants’ responses indicated that both contrastive segments in the pairs /ħ, h/ and /ʕ, ʔ/ were assimilated to English /h/ and /a/, respectively. This suggests that, in contrast to PAM’s predictions, non-learners low performance might be due to their inability to discriminate between the contrastive segments. However, this speculative interpretation needs to be testified empirically using a sound discrimination task.

*Learners’ production of pharyngeal versus non-pharyngeal consonants.*

The findings on Arabic L2 learners’ production of pharyngeal consonants and their counterparts supported SLM’s hypotheses, particularly H2, 3, 5. Since the pharyngeal consonants /ħ, ʕ/ are phonetically distinct from English /h, a/, respectively, English learners of Arabic were able to discern these phonetic differences. Consequently, these learners could establish new phonetic categories for the pharyngeal sounds (according to SLM’s H2 and H3).
This possible category formation for the pharyngeal consonants contributed to learners’ articulation perfection of pharyngeal consonants.

Arabic non-pharyngeal consonants /h, ʔ/, on the other hand, bear an insignificant perceptual distinction from English /h/ and /a/. According to SLM’s fifth hypothesis, this phonetic similarity might have blocked category formation for the non-pharyngeal consonants (i.e. the equivalence classification phenomenon). In essence, since English learners of Arabic could not discern any perceptual difference between Arabic non-pharyngeal consonants and English /h, a/; therefore, they used the same English category to process the Arabic sounds. This can explain why Arabic non-pharyngeal consonants and English /h, a/ resembled each other in production.

*Sound learnability: Pharyngeal and non-pharyngeal consonants.*

Finally, according to the differences between the two participant groups, I concluded that improving a learner’s Arabic proficiency may increase their ability to produce pharyngeal consonants. However, only if they sustain the same perfection level in articulating non-pharyngeal consonants that they demonstrate at early learning stages, they might be able to produce the non-pharyngeal consonants authentically.

In general, the production difference between pharyngealized and non-pharyngealized sounds (Section 5.2.1.1) was more evident than the production difference between pharyngeal and non-pharyngeal sounds (Section 5.2.1.2), although both the non-pharyngealized and non-pharyngeal categories are common between Arabic and English. This was consistent with the findings of Question III which compared between the pharyngeal and pharyngealized consonants production.
5.2.2 Production of pharyngeal versus pharyngealized consonants.

Although exposure to Arabic language resulted in higher production accuracy for both pharyngeal and pharyngealized consonants, it did not influence the direction of difference between these two sound categories. This is evident from the fact that both learners and non-learners of Arabic produced the pharyngeal consonants more accurately than the pharyngealized sounds. However, the increase in Arabic proficiency led to an increased difference between these sound categories.

This finding was inconsistent with the initial prediction, which was based on the fact that pharyngeal consonants are more extreme and require a narrower approximation than the pharyngealized ones (Delattre, 1971; Laufer et al., 1988). These differences in articulatory settings for both categories, according to the initial prediction, would cause all participants to encounter greater difficulty in articulating pharyngeal consonants. There are two possible explanations for this counter evidence, each of which tackles one group of participants.

5.2.2.1 Explanations of production differences among the learners.

The production differences between pharyngeal and pharyngealized consonants among the learners group can be explained in relation to their language use. At the time of the study, it was quite noticeable that English was the most dominant language they used and that certain Arabic words were more frequently used than any other Arabic words in their daily communication with each other. These words consist of the Arabic pharyngeal segments /ḥ, ʕ/, which suggests that pharyngeal consonants were more dominantly used than the pharyngealized consonants. This was true for the majority of the learners group since it was a homogeneous group in the sense that most of them had the same cultural background and ethnicity affiliation. It might also be due to the fact that some of them \( n = 7 \) spoke Somali, which has pharyngeal
consonants, as an additional language. These confounding variables may explain the reason for these participants’ authentic articulation of pharyngeal consonants more than the pharyngealized ones; however, a corpus-based analysis of words and sounds frequency may provide a more solid evidence.

5.2.2.2 *Explanations of production differences among the learners and non-learners.*

Another reason is based on the articulatory differences between pharyngeal and pharyngealized consonants. English has a pharyngeal place of articulation for the vowel /a/; therefore, English and Arabic phonological spaces are similar, to some extent, in terms of their *spatial* invariant. However, the complex gestural constellation that is harnessed in articulating pharyngealized consonants in particular posed a difference in the dynamic aspect of the two phonological spaces. This complexity decreases in the case of pharyngeal consonants production which requires only one primary place of articulation. In contrast, it becomes more persistent in the case of pharyngealized consonants because they employ two parallel places of articulation. Therefore, it is possible that the increased complexity of pharyngealized consonants’ gestural dynamics (as compared to pharyngeal consonants’ gestures) caused non-native Arabic speakers to misarticulate them. In fact, the observed difference between these two sound categories in terms of their articulation correctness is consistent with my argument about the difficulty of producing pharyngealized versus non-pharyngealized consonants (Section 5.2.1.1).

5.2.2.3 *Sound learnability: Pharyngeal and pharyngealized consonants.*

Given the fact that both groups demonstrated the same direction of production difference between pharyngeal and pharyngealized consonants, I cannot assume that improving a speaker’s Arabic proficiency would result in a reversed difference between these two categories. Based on two findings—i.e. non-learners’ higher performance in pharyngeal consonants than in
pharyngealized ones; and learners’ more authentic production than non-learners’ production—I can conclude that improving a speaker’s Arabic proficiency would increase their ability to coordinate complex, simultaneous articulatory gestures in articulating the pharyngealized sounds.

5.2.3 Intra-category differences among pharyngealized as well as pharyngeal CVs.

According to the significant production difference between pharyngeal and pharyngealized consonants found in Question III, I divided the intra-category comparison into two tests. Each test addressed the intra-category differences among the phonemes of a single category.

5.2.3.1 Intra-category differences between the two pharyngeals /ħ/ and /ʕ/.

Although Arabic proficiency resulted in an authentic production of the two sounds, its effect on the production of /ħ/ was more evident than /ʕ/. Also, it led to a marginal production difference between the two sounds as opposed to the considerable production difference that was associated with the lack of proficiency.

The direction of the difference between /ħ/ and /ʕ/ contradicted the initial prediction. I expected that the frication noise of /ħ/ (which is absent in /ʕ/) would carry enough perceptual signal to allow participants to perceive it accurately and, thus, produce it with higher correctness than /ʕ/. However, there should definitely be another difference between these two consonants that caused the production difference. Perhaps, the counter evidenced difference between /ħ/ and /ʕ/ might have resulted from the difference in the acoustic cues that are more salient on /ʕ/ than /ħ/. Since /ʕ/ is a plosive consonant, the burst of air in its articulation may presumably aid its perception and, consequently, its production; however, there is no clear evidence for this assumption in literature. Alternatively, the observed production difference between /ħ/ and /ʕ/
might have resulted from the gestural differences between /h/ and its English counterpart versus the difference between /ʕ/ and its English counterpart.

*Explanation of production differences between /h/ and /ʕ/ from articulation perspective.*

It is likely that the inconsistent differences between each pharyngeal consonant and its English counterpart resulted in these different rates of assimilation. Evidence from participants’ production responses (Sections 4.3.2 and 4.3.3) supported this claim. Approximately 20% of participants’ responses to /ʕ/ were assimilated to English /a/ as opposed to 30% productions of English /h/ in response to /h/ exemplars.

In addition, Delattre’s (1971) empirical findings confirmed this argument. The voiceless pharyngeal /h/ has a narrower and lower pharyngeal stricture (and, thus, smaller pharyngeal cavity below this stricture) than the voiced pharyngeal /ʕ/. Their English counterparts, on the other hand, have wider approximations: a mid-approximation for the low vowel /a/ and a wide laryngeal constriction for /h/ (Esling, 1999). Consequently, in terms of the degree of constriction, the difference between /h/ and its counterpart /h/ is greater than the difference between /ʕ/ and /a/. Based on the definition of the native phonological space, these degrees of constriction for /h/ and /ʕ/, which are narrower than English pharyngeal approximation, cannot be incorporated within English phonological space. In addition, since the pharyngeal constriction for /h/ is farther from English phonological space than the degree of /ʕ/ constriction, it is possible that this resulted in more difficulty articulating /h/ than /ʕ/.

*SLM account of the production differences between /h/ and /ʕ/.*

Although learners’ production of /h/ was lower than /ʕ/, I cannot assume that the equivalence classification mechanism (SLM’s H5) intervened with phonetic category formation for /h/. The reason is that learners produced it with a considerable perfection despite the fact that
they scored higher in /ʕ/ production. Hence, SLM cannot explain the production difference between /ħ, ʕ/ since both are perceptually distinct from English sounds.

*Sound learnability: /ħ/ versus /ʕ/.*

Given the minimal articulation difference that was observed among the learners, with enough exposure to the language and extensive pronunciation training, the difference between the two pharyngeal sounds (in terms of their articulation accuracy) could diminish over time and the articulation accuracy would increase remarkably. Hence, differences in gestural specifications—which imposed differences in production accuracy between the two sounds among non-learners—would have no significant influence on production accuracy when improving the speaker’s Arabic proficiency.

5.2.3.2 Intra-category differences among the pharyngealized sounds.

In general, adequate exposure to Arabic increased the production accuracy of all pharyngealized consonants differently. That is, it influenced the production of some consonants more considerably than other consonants. For instance, its effect on improving the articulation of /tˤ/ was the most remarkable compared to other pharyngealized consonants. Arabic proficiency is associated with comparable production correctness for /tˤ/ and /sˤ/. It also increased the production difference between /dˤ/ and /ðˤ/. Compared to other pharyngealized consonants, /dˤ/ was the most vulnerable to production errors across all learning stages. In contrast, /sˤ/ was the most correctly produced sound regardless of speakers’ proficiency level. Above all, although exposure to the language improved the articulation of all these consonants, it did not change the direction of differences among them (Figure 4.5 shows that both groups demonstrated the same pattern of difference)
In fact, PAM cannot justify these differences among pharyngealized consonants since all of them are equally deviant from the NL’s sounds (i.e. /t, s, d, ð/) and, therefore, all of them represented the same assimilation pattern. The following sub-sections provide interpretations of these differences from several perspectives.

**Production of /sˤ/ as well as /tˤ/.

The reason for such differences might be related to the differences in the sounds’ phonological features. I argued earlier (Section 5.2.1.1) that the pharyngeal gesture by itself does not impose a great difficulty because participants were able to articulate pharyngeal consonants and pharyngealized non-Arabic sounds. Therefore, and because all pharyngealized consonants share relatively similar pharyngealization, it cannot be the pharyngeal feature that caused the production differences. It is, rather, other inherited features that induced these differences.

For instance, the [sibilant] and [-voiced] features for /sˤ/ aided the detection of the primary articulation. The secondary articulation (pharyngealization) was easily detectable and imitated in a great number of instances (Section 5.2.1.1). When integrating all these features in a single sound (i.e. /sˤ/), it became relatively easy for participants to recognize its articulatory configurations and shadow them in their sound production. Evidence from participants’ responses confirmed this explanation. Namely, all participants’ responses to /sˤ/ stimuli were either plain sibilant /s, z/ or pharyngealized sibilant /sˤ/. Similarly, the voiceless feature, that /tˤ/ bears, might have facilitated sound recognition and imitation.

**Production of /ðˤ/ and /tˤ/.

However, there is still a persistent question of why the voicelessness of /tˤ/ helped in articulating it more authentically than the voicelessness of /ðˤ/. In addition, the pharyngeal constriction for the fricative /ðˤ/ was wider than the constriction used for the plosive /tˤ/ (Laufer
et. al., 1988). These acoustic cues that are equally salient on /tˤ/ and /ðˤ/, and the articulatory advantage of /ðˤ/ over /tˤ/ cannot explain the higher production accuracy for /ðˤ/ than /tˤ/.

Alternatively, there might be other acoustic and articulatory features that are more salient and accessible on /ðˤ/ than /tˤ/, and it might be related to the manner of articulation. This leaves an open question for future research in perception and production differences between these sounds based on their acoustic and articulatory correlates.

However, there is another explanation of the relatively lower production performance for /ðˤ/ compared to /tˤ/. From a dialectological perspective, most Arabic dialects use /ðˤ/ and /dˤ/ interchangeably. Therefore, participants (particularly the learners group) might have confused these sounds, which caused a lower performance for /ðˤ/ (and /dˤ/) compared to /tˤ/. The findings from participants’ response patterns supported this interpretation. Approximately 4.17% of learners’ responses to /ðˤ/ exemplars were produced as /dˤ/, which was the second most recurrent, deviant response—i.e. the most frequent responses other than /ðˤ/ included /ð/ (15.8%) and /dˤ/.

Production of /dˤ/.

Finally, a confounding variable might have caused the low production performance for /dˤ/. In some instances, Arabic perceivers rated the shadowed sound as convergent to the stimulus and, at the same time, transcribed it differently (mostly as /b/). This indicates that they might have perceived /dˤ/ stimulus itself wrongly. In other instances, they were able to detect the divergence between /dˤ/ stimuli and the corresponding responses, in which cases they were able to recognize /dˤ/ correctly. In brief, a problem might have stemmed from /dˤ/ exemplars themselves or from Arabic listeners’ inability to perceive /dˤ/ as a stimulus and as a response. This could have possibly influenced rating participants’ production of /dˤ/ as low.
5.2.4 Vowel effect on pharyngeal and pharyngealized consonants’ production.

Since pharyngeal and pharyngealized consonants yielded significant production differences, I divided this question into two tests of the effect of vowel choice on each sound category separated from the other category.

5.2.4.1 Vowel effects on pharyngealized consonants’ production.

Three sets of pharyngealized consonants constituted the target sounds for this comparison: /tˤaː, sˤa:, dˤaː/ versus /tˤiː, sˤi:, dˤiː/ versus /tˤuː, sˤu:, dˤuː/. It is important to emphasize that these vowels (i.e. /aː, iː, uː/) have the same vowel length, yet they differ in vowel quality.

Arabic experience did not reverse the direction of differences between vowels. That is, all non-native speakers, regardless of their proficiency level, produced pharyngealized CVs most accurately when the adjacent vowel was /aː/ and least accurately when it was /uː/. However, Arabic proficiency contributed to improved production for the three CV sets differently. For instance, compared to the other sets, the CV set with /aː/ was the most remarkably influenced by speakers’ proficiency level. The other effect of Arabic proficiency on these CV sets is that it increased the differences between them substantially. For example, the lack of exposure to the language was associated with minimal difference ($MD = .02$) between /aː/ and /iː/ sets (as indicated in the results of a paired-samples t-test: $t(19) = .90, p > .05, MD = .02, SD = .98$); in contrast, after exposure this difference increased significantly.

Explanation of vowel effect on pharyngealized consonants’ production.

In general, the overall trend of difference between the three CV sets was in conformity with Jongman, Herd, Almasri, Sereno and Combest (2011). This study investigated differences in the acoustic effects of the coarticulation associated with pharyngealization. They found that, in
progressive coarticulation cases, the pharyngealization effect was persistent throughout the entire length of /a/ while it rapidly disappeared when the adjacent vowel was /i/ or /u/. Furthermore, they found that pharyngealization had a stronger effect on the coarticulated /a/ than on /i/ or /u/, as indicated in significantly higher F1 and lower F2 in /a/ compared to /i/ and /u/. It is likely that the persistence and markedness of pharyngealization on the coarticulated /a/ contributed to my participants’ higher performance for pharyngealized CVs with /a:/.

Alternatively, it is possible that Arabic L1 listeners’ ability to detect pharyngealization on /a:/ is better than the other vowels. Therefore, they rated the pharyngealized CVs with /a:/ higher than /i:/ and /u:/.

In brief, given the fact that /a:/ bears a more salient pharyngealization effect than /i:/ and /u:/, the high scores for CVs with /a:/ are due to either English or Arabic speakers’ detection of pharyngealized /a:/ rather than pharyngealized /i:, u:/.

With respect to the difference between pharyngealized CVs with /i:/ and CVs with /u:/, Jongman et. al. (2011) found a significantly stronger pharyngealization effect on F2 in the coarticulated /i/ than /u/. Jongman’s et. al. (2011) acoustic evidence supports my findings in two aspects: first, both groups’ performed better in producing pharyngealized CVs with /i:/ than /u:/ and, second, the small production difference between these two CV sets corresponds with Jongman’s et. al. (2011) finding that only F2 of the coarticulated /i:/ was influenced by pharyngealization. In other words, pharyngealization has a relatively small, yet significant, effect on the pharyngealized vowel /i:/ compared to /u:/; therefore, the difference in production accuracy between /i:/ and /u:/ was also small.

*Sound learnability: Pharyngealized consonants with different vowels.*

These findings suggest that when pharyngealized consonants are accompanied with /a:/, they become more easily learnable. The reason is that, this vowel bears more pharyngealization
effect than /i:/ and /u:. This prediction is in agreement with Ali and Daniloff (1974) which revealed that the presence of pharyngealization can be detected on the coarticulated vowel itself.

In addition, for pharyngeal and pharyngealized CVs alike, both participant groups had, to some extent, similar direction of difference between the three vowel sets. Therefore, Arabic proficiency level had no influence on producing these consonants differently as a result of changing the adjacent vowel.

5.2.4.2 Vowel effects on pharyngeal consonants’ production.

Similar to the pharyngealized consonants, this section targeted three sets of pharyngeal CVs: /haː, ʕaː/ versus /hiː, ʕiː/ versus /huː, ʕuː/. Increasing speakers’ proficiency level enhanced their production for all pharyngeal CV sets. Regardless of non-native speakers’ proficiency level in Arabic, changing the adjacent vowel influenced the production of pharyngeal CVs quite minimally. In addition, language exposure did not affect the direction of differences between the three CV sets, which is evident from the fact that, for both groups, the /a:/ set was the most authentically produced set whereas the /u:/ set was the least.

If Jongman’s et. al. (2011) findings can be generalized to pharyngeal consonants, they may account for the differences observed among the pharyngeal CV sets. In other words, Jongman et. al. (2011) found a small difference between /i:/ and /u:/ in terms of bearing the pharyngealization effect. If we arguably accept the assumption that pharyngeal consonants spread the pharyngeal effect to the adjacent vowel, the acoustic difference between /i:/ and /u:/ could be minimal; therefore, the production difference between them would be also minimal. Importantly, this is a speculative, anecdotal inference from Jongman’s et. al. (2011) findings and it was not empirically evidenced.
5.3 Implications for Arabic sound learning

This study entails theoretical implications to be considered in research conducted on sound learning as well as some practical implications to aim to inform current practices in Arabic sound learning and teaching. In the subsequent discussion, I align these implications with specific results that were empirically evidenced in this study.

5.3.1 Theoretical implications.

The study suggests that articulatory gesture is a basic primitive of sound learning. One of the findings that validate this claim is participants’ responses to pharyngealized stimuli in which they decomposed the pharyngealized consonants into two components: one was exemplified in the primary place of articulation of non-Arabic sounds, and another in pharyngealizing these sounds. This suggests that participants identified the pharyngeal gesture and that they integrated this gesture with an anterior gesture which approached the anterior gesture of the stimulus. My claim is in conformity with Best’s (1995) argument that phonetic gesture is the basic premise. If it were the acoustic cues that aid the perception-production loop, the production of /ħ/ would have been more accurate than /ʕ/ because the former (i.e. /ħ/) has more salient acoustic cues than /ʕ/. The findings, nevertheless, revealed that the production of /ʕ/ was more authentic than /ħ/. In fact, the exclusion of acoustic cues as basic primitive in sound learning needs to be attested empirically because, in essence, there might be other gestural differences between /ħ/ and /ʕ/ that influenced their production. In other words, my study promotes the inclusion of phonetic gesture as a primitive in sound learning, and it merely doubts the inclusion of acoustic cues as primitives because it is not based on robust evidence.
Furthermore, the study agrees with Flege’s (1995) SLM on the *linearity* of the relationship between phonetic similarity and sound learnability. However, my conclusion is different from SLM in terms of the determinants of phonetic similarity and the direction of the relation between phonetic similarity and sound learnability. SLM assessed phonetic similarity based on perceptual distinctions between the NL and TL sounds while I determine phonetic similarity according to the TL sound’s gestural specifications. My study results indicate that the increased complexity of a TL sound’s gestural constellations (compared to NL sounds) corresponds to increased production difficulty. This leads to the second difference between my conclusion and SLM. That is, SLM proposes a negative correlation between perceptual similarity and sound learnability (i.e. perceptually similar sounds are less easy to learn than dissimilar ones). In contrast, I suggest a positive correlation between the similarity (i.e. similarity between NL and TL) in gestural constellations and sound learnability. In other words, the TL sounds whose gestural constellations are more similar to NL sounds are easier to learn than dissimilar sounds.

Importantly, my study results cannot support the claim that phonetic similarity is determined according to the spatial proximity between NL and TL sounds. The reason is, the study yielded contradicting results regarding this determinant. For instance, pharyngeal consonants are farther from English phonological space than pharyngealized consonants; however, they are easier to articulate than the pharyngealized ones. In contrast, /h/ which is farther from English phonological space than /ʕ/ induces lower production accuracy. In addition, in these two cases (i.e. in the case of pharyngeal versus pharyngealized sounds, and in the case of /h/ versus /ʕ/), spatial proximity is not the only difference between the sounds under question.
5.3.2 Practical implications.

Throughout the discussion of the results (in Section 5.2), I offered several sound learnability conclusions which assist practitioners (i.e. teachers, learners, curriculum developers, etc.) of Arabic phonology learning in addressing areas of difficulty and developmental learning changes. Since some sounds are learned easier than others, this may help teachers and curriculum developers in sequencing the material presented to Arabic learners. In addition, I suggest multiple practical implications to be taken into consideration in learning Arabic phonology.

I recommend that practitioners give a special attention to articulatory training in learning and teaching Arabic phonology, particularly Arabic pharyngeal and pharyngealized consonants. Perhaps, it was the immediacy factor in the rapid-shadowing test that contributed to non-learners’ ability to produce these sounds which are quite distinct from English phonemes. Therefore, I suggest that learners of Arabic sounds, when employing rapid articulation, can possibly overcome the production difficulties imposed by the articulatory configurations. However, a single shadowing attempt is not likely to result in a complete mastery of the pharyngeal gesture (based on the findings that non-learners were unable to articulate pharyngeal consonants in some instances). Furthermore, given learners’ attunement to Arabic sounds, sufficient training and exposure to the language may accelerate and improve Arabic sounds learning.

I also recommend that in order to learn producing pharyngealized consonants, it is possibly more efficient for non-native Arabic learners to first learn the pharyngeal, simple gesture exemplified in pharyngeal consonants /ħ, ʕ/. This suggestion is founded on two observations: first, the high accuracy of pharyngeal consonants’ production compared to the pharyngealized ones which require an additional anterior place of articulation; and, secondly,
non-learners’ ability to configure the pharyngeal gesture and integrating it with the wrong anterior gesture which resulted in non-Arabic sounds (i.e. /θˤ, vˤ, bˤ/). These two findings suggest that participants were able to configure the pharyngeal gesture by itself; however, it was the *coordination* between this simple gesture and the right anterior place of articulation that led to misarticulating the pharyngealized consonants. I predict that at a later stage of learning, non-native speakers may learn to coordinate the pharyngeal gesture (which they have already learned) with other primary articulation in order to produce the right pharyngealized consonant.

Moreover, I suggest that affixing pharyngeal and pharyngealized consonants with a vowel that is capable of bearing a sufficient, salient pharyngealization feature may help learners to learn these sounds. Previous studies found evidence that the vowel adjacent to pharyngealized consonants aided the detection of pharyngealization; in addition, my study revealed that the adjacent vowels were not equal in helping learners to configure pharyngealization. Specifically, the vowel /aː/ (which bears more acoustic signals of pharyngealization, according to Jongman’s et. al., 2011) yielded more production accuracy than /iː/ and /uː/.

Finally, the theoretical implications can possibly inform subsequent research in non-native sound learning whereas the other implications provide guidance to the practice of learning (and teaching) non-native sounds. However, given the study population parameters, the findings and implications are not generalizable to all sound learning contexts. Precisely, the findings are restricted to non-native learners of Arabic whose L1 is Canadian English. This excludes native speakers of other languages and other English dialects. In addition, my findings are only applicable to adult learners since the study targeted only adult language learners who arguably exhibit learning capabilities different from young learners’ aptitude. Moreover, the findings on vowel manipulation cannot be generalized to other vowels that have different length. The study
employed only long vowels with varying vowel quality; therefore, I cannot assume that my findings are applicable on short vowels even if they have the same vowel quality. In essence, short vowels differ acoustically from long vowels in bearing pharyngealization cues (Jongman et. al., 2011).

5.4 Conclusions of the study

This thesis extends the existing research in non-native sound learning by exploring a language that has never been tested against theories of this discipline. It also forms a departure from recurrent trends in Arabic phonology research by incorporating recently emerging non-native sound learning models such as Best’s (1995) PAM and Flege’s (1995) SLM.

The main purpose of this study is to examine the learnability of certain Arabic consonantal segments, whose primary or secondary place of articulation is the pharynx, at different learning stages. ‘Learnability’ in this context refers to non-native speakers’ ability to articulate these sounds, a process that is stimulated by sound perception. The learning stages that are targeted in this study are the advanced stage in which speakers demonstrate high Arabic proficiency, and an earlier stage of learning in which speakers had not received Arabic input systematically. All non-native Arabic speakers whose L1 was English were tested in their articulation ability in a rapid-shadowing test. Their production was assessed according to Arabic L1 listeners’ perception—that is, these listeners transcribed the sounds that English speakers produced and rated the degree of similarity between the Arabic exemplars and these sounds. This measurement indicates the degree of production difficulty of a given sound, which I used to compare Arabic consonants to each other and to find out the effect of Arabic proficiency. In the following sub-sections, I discuss the production difficulties across different consonants and across groups.
5.5.1 Consonant production differences.

Despite group differences, there are significant differences among the target sounds. For instance, the pharyngealized consonants are the most difficult sounds to produce compared to their counterparts and the pharyngeals. Pharyngeal consonants’ production also differ from the non-pharyngeals’; however, the production difficulty differ according to speakers’ proficiency. That is, increased proficiency results in more authentic production of pharyngeal than non-pharyngeal; conversely, the lack of Arabic proficiency leads to more accurate production of the non-pharyngeals than the pharyngeals. These production difficulties are attributable to various reasons; for instance, in cases where proficient Arabic speakers are able to discern perceptual differences between NL and TL sounds, category formation aided their production accuracy. Also, other reasons include gestural differences between speakers’ NL and TL—particularly, differences in gestural constellations, temporal requirements and degrees of constriction.

In addition, some pharyngealized consonants exhibit substantial production difficulty than others. For example, /dˤ/ is the most difficult consonant to produce whereas /sˤ/ is the least difficult. Similarly, the two pharyngeal consonants are not equally learnable; in particular, /ʕ/ is easier to learn than /ħ/. Reasons such as gestural differences cannot explain these intra-category production differences; however, dialectological and phonological factors contribute to these production difficulties.

Several findings confirm my claim that the pharyngeal gesture in itself does not cause production difficulty even if it is not harnessed in NL’s consonants’ production. For example, given the fact that the target segments differ from their counterparts in their pharyngeal articulation, the purpose of the first research question which compares between them is to
discover whether it is the pharyngeal place of articulation that causes production difficulties. Results from this question reveal that the pharyngeal articulation in itself does not cause the misarticulation. In essence, non-native speakers were able to use the pharynx in consonantal sound production although their sounds did not always match the Arabic exemplars. In addition, the results of the second question further support this claim in the sense that pharyngeal consonants (which require only the pharyngeal gesture) are easier to produce than pharyngealized ones (which require multiple gestures). Furthermore, analysis of participants’ sound responses shows that participants are able to use the pharynx as a primary or secondary place of consonantal articulation. In some cases where they used it as a secondary articulation, they failed to form the right anterior constriction. These findings suggest that it is the complex gestural constellation that causes production difficulty rather than the pharyngeal articulation alone.

5.5.2 Group differences: Effects of speakers’ Arabic proficiency.

Arabic proficiency affects the production of all sound categories differently which means that some sounds are more vulnerable to changes in speakers’ proficiency than others. In all cases, improved proficiency corresponds to high production accuracy except in the case of non-pharyngeal consonants. Its effects vary according to the sound category under investigation; for instance, its effect on the production of pharyngeal and pharyngealized consonants is more evident than their counterparts. Also, although it has a considerable effect on both pharyngeal and pharyngealized consonants, its effect on the former is more dramatic than the latter. Arabic proficiency influences the production of the two pharyngeals differently; namely, the production of /ḥ/ is more vulnerable to differences in speakers’ proficiency level than /ʕ/. Similarly, not all pharyngealized consonants are equal in terms of Arabic proficiency effect on their production.
That is, the most profound effect of Arabic proficiency is on the production of /tˤ/, followed by /ðˤ/, dˤ/ and /sˤ/, respectively. When these pharyngealized consonants are followed by /aː/, they become more susceptible to differences in speakers’ Arabic proficiency than /iː/ and /uː/. Its effect is also noticeable among pharyngeal consonants which differ in the adjacent vowel. However, it is a reversed effect in the sense that while the most observable effects of Arabic proficiency is on pharyngealized CVs with /aː, iː, uː/, respectively; its effect on pharyngeal CVs is more profound when the adjacent vowel is /uː, iː, aː/, respectively. In general, the consonants that show more profound effect of proficiency level are the ones which impose greater production difficulty for the non-learners. Thus, they signal the areas of articulation difficulty, and comparing them to each other provides insights for practitioners of Arabic sound learning on how to sequence and present the learning material.

5.5 Limitations and recommendations for future research

There are, in fact, some limitations to the study at the data collection and analysis levels. Particularly, the Arabic learners group vary considerably in the age at which they started speaking the language and, consequently, the length of their Arabic speaking experience. The youngest age at which participants (n = 3) started speaking Arabic was 2 while the oldest was 22 (n = 1). Previous studies argued that the age of learning influences speakers’ learnability of the non-native language. Being that, there might be some intra-group differences in task performance due to the variations in individual Arabic experience. This leads to another study limitation; namely, the study did not completely adhere to the posttest-only control group design whose fundamental components are random assignment and controlling the treatment factor (i.e. Arabic experience in this case). It was not possible to assignment participants to groups randomly because, in essence, I had no control over their Arabic experience (i.e. the learners
group have already been exposed to the language before conducting the study). Although this violated the theoretical assumptions of this experimental design, it promoted the ecological validity of the study as participants learned the language in real-life settings. Moreover, the multilingualism of most participants might have influenced their sound production ability. Among the languages they reported were French, Spanish and Somali all of which have at least one pharyngeal sound (Delattre, 1971). I speculate that participants might have formed a phonetic category that represents these non-native pharyngeal sounds and, subsequently, processed the Arabic stimuli according to these established phonetic categories. However, there is no study in literature that testified whether processing non-native sounds is influenced by other non-native phonetic categories.

In addition to these limitations to data collection, there is a limitation at the data analysis level. In the literature of Arabic phonology, there was no index of quantitative analysis of acoustic differences between English sounds and the target Arabic sounds (i.e. pharyngeal and pharyngealized consonants). Neither were there sufficient paradigms of perceptual information of these sounds derived from empirical findings. These data, according to Strange (2007), should be the basis for determining learners’ potential for perceiving and producing non-native sounds. My analyses, therefore, were based on articulatory (and sometimes acoustic) specifications of Arabic sounds separately from English sounds. In other words, the analyses were not grounded on evidenced comparisons of the acoustic, perceptual, articulatory properties of Arabic and English sounds.

With these limitations in mind, I recommend a range of research topics by either considering some methodological improvements to my study or extending the research on Arabic phonology learnability. Methodological improvements, for example, include controlling the
quantity and quality of Arabic input that the experimental group is exposed to prior to testing their sound production. Along with random assignment and sampling, this ensures inter-group equivalence more reliably. By varying input quantity, the researcher can obtain more than one experimental group, besides the control group. This factorial design (Cozby, 1992) may help in observing Arabic sound learnability across varying proficiency levels, instead of examining it between two distinct levels only.

Research on Arabic sound learning can be extended by testifying whether non-native Arabic speakers’ multilingualism has an effect on their potential of learning Arabic sounds, and aligning these findings with data obtained from English monolinguals. This may offer insightful theoretical implications on whether multilingual speakers’ ability to produce Arabic pharyngeal sounds is aided by other non-native category formation. Additionally, I suggest enriching the field of Arabic phonology research by providing empirically evidenced indexes of the acoustic differences between Arabic pharyngeal and pharyngealized consonants and their counterparts. This will pave the ground for subsequent research that testifies PAM and SLM against Arabic phonology by helping researchers drawing valid conclusions. Moreover, since /ðˤ/ and /dˤ/ are vulnerable to dialectical differences, it is insightful to consider dialectological variables besides other linguistic variables in examining non-native speakers’ production of these sounds. Finally, in order to attest my claim about determining sound learnability according to its spatial proximity, I recommend examining the perception and production of other Arabic sounds which vary in their spatial proximity to English sounds’ place of articulation. In fact, the field of Arabic phonology has not been extensively researched; however, these recommendations are derived from the methodological limitations to my study as well as the arguable conclusions I have drawn.
References


Appendix A

Appendix A1

Arabic consonantal phonemic chart

<table>
<thead>
<tr>
<th></th>
<th>Labial</th>
<th>Labio-dental</th>
<th>Interdental</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pharyngeal</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>b ب</td>
<td>d د</td>
<td>t ت</td>
<td>ء ء</td>
<td>k ك</td>
<td>ئ ئ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affricate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3 ج</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>f ف</td>
<td>θ ث</td>
<td>ð ذ</td>
<td>ؤ ؤ</td>
<td>x خ</td>
<td>ئ ئ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m م</td>
<td>n ن</td>
<td></td>
<td></td>
<td>ء ء</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td></td>
<td></td>
<td>ل ل</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flap</td>
<td></td>
<td></td>
<td>r ر</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi vowel</td>
<td>w و</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y ي</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* The sounds on the first line of each cell are voiceless, and the ones in the first line are voiced.
Appendix A (continued)

Appendix A2

Hierarchy of the target sounds (Arabic pharyngeal and pharyngealized consonants) and their plain counterparts.
Appendix B

Language Questionnaire

To the participant:

This is an anonymous questionnaire that is intended to assess your language background. The responses provided in this questionnaire will be used for research purposes only, and the data will be kept anonymous and confidential.

Participant information:

Code: __________________________
Age: __________________________
Gender: _________________________
Email: _________________________

Please read carefully and answer the following questions.

1. First language(s): __________________________

2. Please list all the languages you have learned regardless of your proficiency level. State at which age you have learned the respective language(s). Describe how you learned them (check all that apply). If you learned more than 4 languages, please list them at the back of this sheet.

<table>
<thead>
<tr>
<th>#</th>
<th>Language</th>
<th>Age of learning</th>
<th>Learning source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>from birth</td>
<td>4-12 year old</td>
<td>□ Formal classes  □ Social interaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>more than 18-year old</td>
<td>□ Independent reading</td>
</tr>
<tr>
<td></td>
<td>□ from birth</td>
<td>4-12 year old</td>
<td>□ Independent reading</td>
</tr>
<tr>
<td></td>
<td>□ 13-18 year old</td>
<td>more than 18-year old</td>
<td>□ Exposure to media</td>
</tr>
<tr>
<td></td>
<td>□ from birth</td>
<td>more than 18-year old</td>
<td>□ Other (write) ......</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>from birth</td>
<td>4-12 year old</td>
<td>□ Formal classes  □ Social interaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>more than 18-year old</td>
<td>□ Independent reading</td>
</tr>
<tr>
<td></td>
<td>□ 13-18 year old</td>
<td>more than 18-year old</td>
<td>□ Exposure to media</td>
</tr>
<tr>
<td></td>
<td>□ from birth</td>
<td>more than 18-year old</td>
<td>□ Other (write) ......</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>from birth</td>
<td>4-12 year old</td>
<td>□ Formal classes  □ Social interaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>more than 18-year old</td>
<td>□ Independent reading</td>
</tr>
<tr>
<td></td>
<td>□ 13-18 year old</td>
<td>more than 18-year old</td>
<td>□ Exposure to media</td>
</tr>
<tr>
<td></td>
<td>□ from birth</td>
<td>more than 18-year old</td>
<td>□ Other (write) ......</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>from birth</td>
<td>4-12 year old</td>
<td>□ Formal classes  □ Social interaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>more than 18-year old</td>
<td>□ Independent reading</td>
</tr>
<tr>
<td></td>
<td>□ 13-18 year old</td>
<td>more than 18-year old</td>
<td>□ Exposure to media</td>
</tr>
<tr>
<td></td>
<td>□ from birth</td>
<td>more than 18-year old</td>
<td>□ Other (write) ......</td>
</tr>
</tbody>
</table>
3. In this section, please assess your proficiency level in the following language areas. Assessment scale is beginner, medium, and advanced.

<table>
<thead>
<tr>
<th>Language</th>
<th>Speaking</th>
<th>Listening</th>
<th>Reading</th>
<th>Writing</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Exposure to aural input in Arabic:
   A. Have you had any exposure to the Arabic language?
      ☐ Yes  ☐ No

      If yes, please answer B and C below:
   B. Please indicate all types of exposure to Arabic Language.
      ☐ Media
      ☐ Formal classes
      ☐ Informal social interaction
      ☐ Other (identify): _______________________

   C. How often have you been exposed to the above-mentioned resources?
      ☐ Everyday  ☐ Twice a week
      ☐ Once a week  ☐ Once a month

5. Oral communication in Arabic:
   A. Have you ever practiced speaking in Arabic?
      ☐ Yes  ☐ No

      If yes, please answer B and C below:
   B. Age at which age you have practiced speaking Arabic:
      from: ___________________  to: ___________________

   C. How often:
      ☐ Everyday  ☐ Twice a week
      ☐ Once a week  ☐ Once a month

6. Please indicate the easiest language aspects you have found in learning Arabic:
   ☐ Arabic sounds  ☐ Grammar  ☐ Orthography
   ☐ Vocabulary  ☐ Other (identify): _______________________
7. Please indicate the most difficult language aspects you have found in learning Arabic (if any):
   - [ ] Arabic sounds
   - [ ] Grammar
   - [ ] Orthography
   - [ ] Vocabulary
   - [ ] Other (identify): ________________

8. If you have any comments or further information, please write them below.
   __________________________________________
   __________________________________________
   __________________________________________
   __________________________________________
# Appendix C

Matrix of the stimuli illustrating the three cohorts of each consonant and the random order in which they were presented to participants

<table>
<thead>
<tr>
<th>Serial#</th>
<th>Phonological value</th>
<th>Arabic Orthography</th>
<th>Random#</th>
<th>Random Phonological value</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-1</td>
<td>/taː/</td>
<td>ت</td>
<td>R-1</td>
<td>/waː/</td>
</tr>
<tr>
<td>T-2</td>
<td>/tiː/</td>
<td>ط</td>
<td>R-2</td>
<td>/tˤiː/</td>
</tr>
<tr>
<td>T-3</td>
<td>/tuː/</td>
<td>م</td>
<td>R-3</td>
<td>/haː/</td>
</tr>
<tr>
<td>T-4</td>
<td>/tˤaː/</td>
<td>ر</td>
<td>R-4</td>
<td>/quː/</td>
</tr>
<tr>
<td>T-5</td>
<td>/tˤiː/</td>
<td>ن</td>
<td>R-5</td>
<td>/yˤiː/</td>
</tr>
<tr>
<td>T-6</td>
<td>/tˤuː/</td>
<td>ط</td>
<td>R-6</td>
<td>/ʔaː/</td>
</tr>
<tr>
<td>T-7</td>
<td>/ðaː/</td>
<td>س</td>
<td>R-7</td>
<td>/zuː/</td>
</tr>
<tr>
<td>T-8</td>
<td>/ðiː/</td>
<td>ظ</td>
<td>R-8</td>
<td>/sˤaː/</td>
</tr>
<tr>
<td>T-9</td>
<td>/ðuː/</td>
<td>ظ</td>
<td>R-9</td>
<td>/ʔiː/</td>
</tr>
<tr>
<td>T-10</td>
<td>/ʔaː/</td>
<td>س</td>
<td>R-10</td>
<td>/kuː/</td>
</tr>
<tr>
<td>T-11</td>
<td>/ʔiː/</td>
<td>ط</td>
<td>R-11</td>
<td>/siː/</td>
</tr>
<tr>
<td>T-12</td>
<td>/ʔuː/</td>
<td>م</td>
<td>R-12</td>
<td>/dˤaː/</td>
</tr>
<tr>
<td>T-13</td>
<td>/saː/</td>
<td>ت</td>
<td>R-13</td>
<td>/kiː/</td>
</tr>
<tr>
<td>T-14</td>
<td>/siː/</td>
<td>م</td>
<td>R-14</td>
<td>/ʔuː/</td>
</tr>
<tr>
<td>T-15</td>
<td>/suː/</td>
<td>ط</td>
<td>R-15</td>
<td>/ʔiː/</td>
</tr>
<tr>
<td>T-16</td>
<td>/sˤaː/</td>
<td>س</td>
<td>R-16</td>
<td>/ʔuː/</td>
</tr>
<tr>
<td>T-17</td>
<td>/sˤiː/</td>
<td>ط</td>
<td>R-17</td>
<td>/tuː/</td>
</tr>
<tr>
<td>T-18</td>
<td>/sˤuː/</td>
<td>م</td>
<td>R-18</td>
<td>/ʔiː/</td>
</tr>
<tr>
<td>T-19</td>
<td>/daː/</td>
<td>ت</td>
<td>R-19</td>
<td>/ʔuː/</td>
</tr>
<tr>
<td>T-20</td>
<td>/diː/</td>
<td>م</td>
<td>R-20</td>
<td>/baː/</td>
</tr>
<tr>
<td>T-21</td>
<td>/duː/</td>
<td>م</td>
<td>R-21</td>
<td>/dˤiː/</td>
</tr>
<tr>
<td>T-22</td>
<td>/dˤaː/</td>
<td>ط</td>
<td>R-22</td>
<td>/yuː/</td>
</tr>
<tr>
<td>T-23</td>
<td>/dˤiː/</td>
<td>م</td>
<td>R-23</td>
<td>/ʔaː/</td>
</tr>
<tr>
<td>T-24</td>
<td>/dˤuː/</td>
<td>ط</td>
<td>R-24</td>
<td>/qˤiː/</td>
</tr>
<tr>
<td>T-25</td>
<td>/haː/</td>
<td>ح</td>
<td>R-25</td>
<td>/ziː/</td>
</tr>
<tr>
<td>T-26</td>
<td>/hiː/</td>
<td>ح</td>
<td>R-26</td>
<td>/ʔaː/</td>
</tr>
<tr>
<td>T-27</td>
<td>/hu:/</td>
<td>R-27</td>
<td>/du:/</td>
<td></td>
</tr>
<tr>
<td>T-28</td>
<td>/kā:/</td>
<td>R-28</td>
<td>/wi:/</td>
<td></td>
</tr>
<tr>
<td>T-29</td>
<td>/ši:/</td>
<td>R-29</td>
<td>/tˤuː:/</td>
<td></td>
</tr>
<tr>
<td>T-30</td>
<td>/šu:/</td>
<td>R-30</td>
<td>/sa:/</td>
<td></td>
</tr>
<tr>
<td>T-31</td>
<td>/haː/</td>
<td>R-31</td>
<td>/hiː:/</td>
<td></td>
</tr>
<tr>
<td>T-32</td>
<td>/hiː/</td>
<td>R-32</td>
<td>/yaː:/</td>
<td></td>
</tr>
<tr>
<td>T-33</td>
<td>/huː/</td>
<td>R-33</td>
<td>/dˤuː:/</td>
<td></td>
</tr>
<tr>
<td>T-34</td>
<td>/ʔaː/</td>
<td>R-34</td>
<td>/ðaː:/</td>
<td></td>
</tr>
<tr>
<td>T-35</td>
<td>/ʔiː/</td>
<td>R-35</td>
<td>/huː:/</td>
<td></td>
</tr>
<tr>
<td>T-36</td>
<td>/ʔuː/</td>
<td>R-36</td>
<td>/sˤiː:/</td>
<td></td>
</tr>
<tr>
<td>F-37</td>
<td>/waː/</td>
<td>R-37</td>
<td>/zaː:/</td>
<td></td>
</tr>
<tr>
<td>F-38</td>
<td>/wiː/</td>
<td>R-38</td>
<td>/sˤiː:/</td>
<td></td>
</tr>
<tr>
<td>F-39</td>
<td>/wuː/</td>
<td>R-39</td>
<td>/daː:/</td>
<td></td>
</tr>
<tr>
<td>F-40</td>
<td>/yaː/</td>
<td>R-40</td>
<td>/bˤiː:/</td>
<td></td>
</tr>
<tr>
<td>F-41</td>
<td>/yiː/</td>
<td>R-41</td>
<td>/tˤaː:/</td>
<td></td>
</tr>
<tr>
<td>F-42</td>
<td>/yuː/</td>
<td>R-42</td>
<td>/ʔaː:/</td>
<td></td>
</tr>
<tr>
<td>F-43</td>
<td>/ʒaː/</td>
<td>R-43</td>
<td>/bˤuː:/</td>
<td></td>
</tr>
<tr>
<td>F-44</td>
<td>/ʒiː/</td>
<td>R-44</td>
<td>/sˤiː:/</td>
<td></td>
</tr>
<tr>
<td>F-45</td>
<td>/ʒuː/</td>
<td>R-45</td>
<td>/wuː:/</td>
<td></td>
</tr>
<tr>
<td>F-46</td>
<td>/qaː/</td>
<td>R-46</td>
<td>/suː:/</td>
<td></td>
</tr>
<tr>
<td>F-47</td>
<td>/qiː/</td>
<td>R-47</td>
<td>/hɑː:/</td>
<td></td>
</tr>
<tr>
<td>F-48</td>
<td>/quː/</td>
<td>R-48</td>
<td>/diː:/</td>
<td></td>
</tr>
<tr>
<td>F-49</td>
<td>/baː/</td>
<td>R-49</td>
<td>/qaː:/</td>
<td></td>
</tr>
<tr>
<td>F-50</td>
<td>/biː/</td>
<td>R-50</td>
<td>/sˤuː:/</td>
<td></td>
</tr>
<tr>
<td>F-51</td>
<td>/buː/</td>
<td>R-51</td>
<td>/taː:/</td>
<td></td>
</tr>
<tr>
<td>F-52</td>
<td>/zaː/</td>
<td>R-52</td>
<td>/hiː:/</td>
<td></td>
</tr>
<tr>
<td>F-53</td>
<td>/ziː/</td>
<td>R-53</td>
<td>/kaː:/</td>
<td></td>
</tr>
<tr>
<td>F-54</td>
<td>/zuː/</td>
<td>R-54</td>
<td>/ðuː:/</td>
<td></td>
</tr>
</tbody>
</table>
Appendix D

Inter-rater reliability test (Kappa) results for measuring the internal consistency between the two Arabic perceivers’ assessments of the shadowed sounds

<table>
<thead>
<tr>
<th>Shadowed Sound</th>
<th>Kappa value</th>
<th>Strength of agreement$^a$</th>
<th>Sig.</th>
<th>Cases (N)</th>
<th>Valid</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>/tˤaː/</td>
<td>.69</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/tˤiː/</td>
<td>.63</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/tˤuː/</td>
<td>.62</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/sˤaː/</td>
<td>.51</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/sˤiː/</td>
<td>.55</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/sˤuː/</td>
<td>.14</td>
<td>Slight</td>
<td>.17</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/dˤaː/</td>
<td>.48</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/dˤiː/</td>
<td>.27</td>
<td>Fair</td>
<td>.02</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/dˤuː/</td>
<td>.51</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/ðˤaː/</td>
<td>.64</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/ðˤiː/</td>
<td>.65</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/ðˤuː/</td>
<td>.45</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/ʕaː/</td>
<td>.57</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/ʕiː/</td>
<td>.51</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/ʕuː/</td>
<td>.46</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/ħaː/</td>
<td>.73</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/ħiː/</td>
<td>.70</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>/ħuː/</td>
<td>.63</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
### NON-NATIVE SOUND PRODUCTION

<table>
<thead>
<tr>
<th>Sound</th>
<th>Agreement</th>
<th>Strength</th>
<th>Score</th>
<th>Total</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ha:/</td>
<td>.57</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/hi:/</td>
<td>.49</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/hu:/</td>
<td>.40</td>
<td>Fair</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/ʔa:/</td>
<td>.56</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/ʔi:/</td>
<td>.66</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/ʔu:/</td>
<td>.78</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/ta:/</td>
<td>.88</td>
<td>Almost perfect</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/ti:/</td>
<td>.69</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/tu:/</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>N/A</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/sa:/</td>
<td>.38</td>
<td>Fair</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/si:/</td>
<td>.46</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/su:/</td>
<td>.16</td>
<td>Slight</td>
<td>.18</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/da:/</td>
<td>.41</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/di:/</td>
<td>.67</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/du:/</td>
<td>.50</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/ʔa:/</td>
<td>.63</td>
<td>Substantial</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/ʔi:/</td>
<td>.53</td>
<td>Moderate</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>/ʔu:/</td>
<td>.83</td>
<td>Almost perfect</td>
<td>.00</td>
<td>40</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Strength of agreement was determined according to Landis and Koch (1977).

<sup>b</sup>No statistics were computed because both perceivers’ scores were constant—i.e. both Arabic perceivers gave a score of 3 to all participants.
Appendix E

Research ethics approval from Carleton University’s Research Ethics Board (REB)

Ethics Clearance Form

This is to certify that the Carleton University Research Ethics Board has examined the application for ethical clearance. The REB found the research project to meet appropriate ethical standards as outlined in the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, 2nd edition and the Carleton University Policies and Procedures for the Ethical Conduct of Research.

X New clearance
☐ Renewal of original clearance

Original date of clearance:

Date of clearance: 19 March 2013
Researchers: Sawsan Alwabari, Master’s student
Department: Linguistics and Language Studies
Supervisor: Prof. Lev Blumenfeld, Linguistics
Supervisor: Prof. Erik Anonby, French
Project number: 13-1226
Title of project: Non-native Arabic speakers’ acquisition of pharyngeal and pharyngealized: Implications for the pedagogy of L2 phonology

Clearance expires: 31 May 2014

All researchers are governed by the following conditions:

Annual Status Report: You are required to submit an Annual Status Report to either renew clearance or close the file. Failure to submit the Annual Status Report will result in the immediate suspension of the project. Funded projects will have accounts suspended until the report is submitted and approved.

Changes to the project: Any changes to the project must be submitted to the Carleton University Research Ethics Board for approval. All changes must be approved prior to the continuance of the research.

Adverse events: Should any participant suffer adversely from their participation in the project you are required to report the matter to the Carleton University Research Ethics Board. You must submit a written record of the event and indicate what steps you have taken to resolve the situation.

Suspension or termination of clearance: Failure to conduct the research in accordance with the principles of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, 2nd edition and the Carleton University Policies and Procedures for the Ethical Conduct of Research may result in the suspension or termination of the research project.

Andy Adler, Chair
Carleton University Research Ethics Board

Louise Heslop, Vice-Chair
Carleton University Research Ethics Board
Appendix F

Samples of research recruitment advertisement: recruitment cards and posters

**Participation in a master’s research on sound**

**perception and production**

Adult native-English speakers who have NEVER learned Arabic or Hebrew are being sought out to participate in the study.

Time commitment is minimal (about 30 min)

**Participants are offered $10**

For participation and inquiries, please contact the researcher at sawsan.alwabari@carleton.ca

The project received the ethics approval from Carleton’s REB in March 19, 2013 (ethics@carleton.ca)

**Participation in a master’s research on sound**

**perception and production**

Adult native-Arabic speakers are being sought out to participate in the study.

**Participants are offered $10/30 minutess**

For participation and inquiries, please contact the researcher at sawsan.alwabari@carleton.ca

The project received the ethics approval from Carleton’s REB in March 19, 2013 (ethics@carleton.ca)
Appendix F (continued)

Research Participation

Participants required for research in speech perception and production

In order to participate in the project, you must be an English native speaker who has NEVER learned Arabic of Hebrew.

Participants must not have experienced any kind of speech or perception impairment.

Participants are offered $10 for taking part in this project.

Time commitment is minimal (about 30 minutes).

For further information on this participation opportunity, please contact us.

Email: sawsan.alwabari@carleton.ca

This research has been approved by Carleton University Ethics Board on March 19, 2013. For ethics inquiries, please contact the Research Ethics Board at ethics@carleton.ca
Appendix G

The consent form that participants signed in compliance with research ethics protocols

Carleton University

Learning Project Participant’s Informed Consent Form

Please read carefully. This informed consent form is to ensure that you understand the purpose of this learning project and the nature of your participation. This form must provide sufficient information so that you are in an informed position to decide whether you wish to participate in this learning project. There are no penalties of any sort, regardless of your decision. Ethics approval was issued in March 19, 2013.

<table>
<thead>
<tr>
<th>Course learning project is associated with:</th>
<th>ALDS 5909: MA Thesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name, affiliation, and means of contacting the research supervisors:</td>
<td>Co-supervisors:</td>
</tr>
<tr>
<td></td>
<td>Prof. Erik Anonby Email: <a href="mailto:erik_anonby@carleton.ca">erik_anonby@carleton.ca</a></td>
</tr>
<tr>
<td></td>
<td>Prof. Lev Blumenfeld Email: <a href="mailto:lev_blumenfeld@carleton.ca">lev_blumenfeld@carleton.ca</a></td>
</tr>
</tbody>
</table>

Student Conducting Learning Project

<table>
<thead>
<tr>
<th>Name</th>
<th>Contact Information</th>
<th>Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sawsan Alwabari</td>
<td><a href="mailto:sawsan.alwabari@carleton.ca">sawsan.alwabari@carleton.ca</a></td>
<td>S.A.</td>
</tr>
</tbody>
</table>

Title of the learning project: Non-native Arabic speakers’ acquisition of pharyngeal and pharyngealized phonemes: Implications for the pedagogy of L2 phonology

Purpose of the learning project: The study investigates the mental representations that English speakers have when perceiving pharyngeal and pharyngealized phonemes. It also compares between these two sound categories in terms of their physiological/ articulatory correlates. The findings are intended to inform current practices in the pedagogy of L2 phonology.

Procedures: A participant will read CV sequences. Then, an acoustically processed recording of this participant will be played, and other participants will be asked to listen carefully, identify the phonemes and produce similar sounds.

Duration and location: Duration: 30-60 minutes

Location: Carleton University

Anonymity/Confidentiality Names and personal information will not be collected, and all data will be presented in aggregate form.

Potential benefits/potential discomfort/Probable risks: Participation entails no risks or harms of any kind—whether physical, psychological, social, economic or legal. Each participant is given a
| **Right to withdraw:** | Participants have the right to withdraw from the experiment after starting it without any penalty. They also have the right to cancel their participation within 2 days after completing the experiment. In this case, they are asked to send a cancellation email to the researcher at sawsan.alwabari@carleton.ca. In both cases of withdrawal, the data obtained from them will be destroyed to ensure confidentiality. |
| **Debriefing:** | Participants are asked to first read and sign this consent. Then, they are prompted to fill out a questionnaire before commencing the experiment. The experiment will follow the procedures outlined above. No debriefing is required after finishing the experiment. |

**Consent**

I have read the above description of the learning project and understand the conditions of my participation. My signature indicates that I agree to participate in this learning project.

__________________________  ____________________________  ____________________________
Participant’s Code  Participant’s Signature  Date
Appendix H

Appendix H1

Descriptive statistics of groups’ scores for pharyngealized and plain consonants

<table>
<thead>
<tr>
<th>Participants groups</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>CI</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>SE</td>
<td>CI</td>
<td></td>
</tr>
<tr>
<td>Pharyngealized</td>
<td>Learners</td>
<td>3.60</td>
<td>.58</td>
<td>.12</td>
<td>3.36</td>
</tr>
<tr>
<td></td>
<td>Non-learners</td>
<td>2.13</td>
<td>.52</td>
<td>.12</td>
<td>1.88</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.87</td>
<td>.92</td>
<td>.09</td>
<td>2.69</td>
</tr>
<tr>
<td>Non-pharyngealized</td>
<td>Learners</td>
<td>4.51</td>
<td>.49</td>
<td>.10</td>
<td>4.32</td>
</tr>
<tr>
<td></td>
<td>Non-learners</td>
<td>4.32</td>
<td>.35</td>
<td>.10</td>
<td>4.13</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>4.41</td>
<td>.43</td>
<td>.07</td>
<td>4.28</td>
</tr>
</tbody>
</table>

Note. SE = standard error; CI = confidence interval; LB = lower bound; UB = upper bound.

Appendix H2

Descriptive statistics of both groups’ scores for pharyngeal consonants and their counterparts

<table>
<thead>
<tr>
<th>Participants groups</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>CI</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>SE</td>
<td>CI</td>
<td></td>
</tr>
<tr>
<td>Pharyngeal</td>
<td>Learners</td>
<td>4.70</td>
<td>.39</td>
<td>.11</td>
<td>4.47</td>
</tr>
<tr>
<td></td>
<td>Non-learners</td>
<td>2.35</td>
<td>.61</td>
<td>.11</td>
<td>2.12</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>3.52</td>
<td>1.29</td>
<td>.08</td>
<td>3.36</td>
</tr>
<tr>
<td>Non-pharyngeal</td>
<td>Learners</td>
<td>4.25</td>
<td>.55</td>
<td>.12</td>
<td>4.00</td>
</tr>
<tr>
<td></td>
<td>Non-learners</td>
<td>4.54</td>
<td>.56</td>
<td>.12</td>
<td>4.29</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>4.40</td>
<td>.57</td>
<td>.09</td>
<td>4.22</td>
</tr>
</tbody>
</table>

Note. SE = standard error; CI = confidence interval; LB = lower bound; UB = upper bound.
Appendix H (Continued)

Appendix H3

Descriptive statistics of both groups’ scores for pharyngealized and pharyngeal consonants

<table>
<thead>
<tr>
<th></th>
<th>Learners</th>
<th>Non-learners</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharyngealized</td>
<td>Mean</td>
<td>SD</td>
<td>Std. Error</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LB</td>
</tr>
<tr>
<td>Learners</td>
<td>3.60</td>
<td>.58</td>
<td>.12</td>
</tr>
<tr>
<td>Non-learners</td>
<td>2.13</td>
<td>.52</td>
<td>.12</td>
</tr>
<tr>
<td>Total</td>
<td>2.87</td>
<td>.92</td>
<td>.09</td>
</tr>
</tbody>
</table>

|                   | Mean     | SD           | Std. Error | 95% CI | N   |
|-------------------|----------|--------------|           |        |     |
|                   |          |              | LB | UB |     |     |
| Learners          | 4.70     | .39          | .11 | 4.46 | 4.93 | 20  |
| Non-learners      | 2.35     | .61          | .11 | 2.12 | 2.58 | 20  |
| Total             | 3.52     | 1.29         | .08 | 3.36 | 3.69 | 40  |

Appendix H4

Descriptive statistics of both groups’ scores for pharyngeal phonemes /h, ʕ/

<table>
<thead>
<tr>
<th></th>
<th>Learners</th>
<th>Non-learners</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharyngeal /ʕ/</td>
<td>Mean</td>
<td>SD</td>
<td>Std. Error</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LB</td>
</tr>
<tr>
<td>Learners</td>
<td>4.69</td>
<td>.48</td>
<td>.17</td>
</tr>
<tr>
<td>Non-learners</td>
<td>2.88</td>
<td>.95</td>
<td>.17</td>
</tr>
<tr>
<td>Total</td>
<td>3.79</td>
<td>1.18</td>
<td>.12</td>
</tr>
</tbody>
</table>

|                  | Mean     | SD           | Std. Error | 95% CI | N   |
|------------------|----------|--------------|           |        |     |
|                  |          |              | LB | UB |     |     |
| Learners         | 4.70     | .46          | .14 | 4.42 | 4.98 | 20  |
| Non-learners     | 1.82     | .76          | .14 | 1.53 | 2.10 | 20  |
| Total            | 3.26     | 1.59         | .10 | 3.06 | 3.46 | 40  |
Appendix H (Continued)

Appendix H5
Multivariate tests output for the within-subjects factor ([tˤ, sˤ, dˤ, ðˤ]) and its interaction with the independent variable

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
<th>η²</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharyngealized phonemes [tˤ, sˤ, dˤ, ðˤ]</td>
<td>61.32</td>
<td>3.00</td>
<td>36.00</td>
<td>.00</td>
<td>.84</td>
<td>1.00</td>
</tr>
<tr>
<td>Phonemes * Group</td>
<td>8.38</td>
<td>3.00</td>
<td>36.00</td>
<td>.00</td>
<td>.41</td>
<td>.99</td>
</tr>
</tbody>
</table>

*Note.* These results were the same for all multivariate tests: Pillai's Trace, Wilks' Lambda, Hotelling's Trace, and Roy's Largest Root.

Appendix H6
Post-hoc test (LSD) comparing participants’ production of /tˤ/ vs. /sˤ/ vs. /dˤ/ vs. /ðˤ/ |

<table>
<thead>
<tr>
<th>(I)</th>
<th>(J)</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/tˤ/</td>
<td>-.78</td>
<td>.15</td>
<td>.00</td>
<td>-.109</td>
<td>-.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>/sˤ/</td>
<td>1.07</td>
<td>.17</td>
<td>.00</td>
<td>.72</td>
<td>1.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>/dˤ/</td>
<td>.26</td>
<td>.18</td>
<td>.17</td>
<td>-.11</td>
<td>.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>/ðˤ/</td>
<td>.78</td>
<td>.15</td>
<td>.00</td>
<td>.47</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td>/sˤ/</td>
<td>/tˤ/</td>
<td>.81</td>
<td>.15</td>
<td>.00</td>
<td>.50</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>/sˤ/</td>
<td>/dˤ/</td>
<td>-1.07</td>
<td>.17</td>
<td>.00</td>
<td>-1.42</td>
<td>-.72</td>
<td></td>
</tr>
<tr>
<td>/sˤ/</td>
<td>/ðˤ/</td>
<td>-1.85</td>
<td>.13</td>
<td>.00</td>
<td>-2.12</td>
<td>-1.58</td>
<td></td>
</tr>
<tr>
<td>/dˤ/</td>
<td>/tˤ/</td>
<td>-1.85</td>
<td>.15</td>
<td>.00</td>
<td>-1.12</td>
<td>-.50</td>
<td></td>
</tr>
<tr>
<td>/dˤ/</td>
<td>/sˤ/</td>
<td>-.26</td>
<td>.18</td>
<td>.17</td>
<td>-.63</td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>/dˤ/</td>
<td>/ðˤ/</td>
<td>-1.04</td>
<td>.15</td>
<td>.00</td>
<td>-1.35</td>
<td>-.73</td>
<td></td>
</tr>
<tr>
<td>/ðˤ/</td>
<td>/tˤ/</td>
<td>-.81</td>
<td>.15</td>
<td>.00</td>
<td>.50</td>
<td>1.12</td>
<td></td>
</tr>
</tbody>
</table>
Appendix H (Continued)

Appendix H7

Descriptive statistics of both groups’ scores for pharyngealed consonants /tˤ, sˤ, dˤ, ðˤ/  

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Std. Error</th>
<th>95% CI</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LB</td>
<td>UB</td>
</tr>
<tr>
<td>/tˤ/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learners</td>
<td>4.20</td>
<td>.69</td>
<td>.21</td>
<td>3.78</td>
<td>4.62</td>
</tr>
<tr>
<td>Non-learners</td>
<td>1.81</td>
<td>1.11</td>
<td>.21</td>
<td>1.39</td>
<td>2.23</td>
</tr>
<tr>
<td>Total</td>
<td>3.00</td>
<td>1.52</td>
<td>.15</td>
<td>2.71</td>
<td>3.30</td>
</tr>
<tr>
<td>/sˤ/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learners</td>
<td>4.24</td>
<td>.60</td>
<td>.13</td>
<td>3.97</td>
<td>4.51</td>
</tr>
<tr>
<td>Non-learners</td>
<td>3.33</td>
<td>.60</td>
<td>.13</td>
<td>3.05</td>
<td>3.60</td>
</tr>
<tr>
<td>Total</td>
<td>3.78</td>
<td>.75</td>
<td>.10</td>
<td>3.59</td>
<td>3.98</td>
</tr>
<tr>
<td>/dˤ/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learners</td>
<td>2.44</td>
<td>1.00</td>
<td>.17</td>
<td>2.09</td>
<td>2.79</td>
</tr>
<tr>
<td>Non-learners</td>
<td>1.43</td>
<td>.47</td>
<td>.17</td>
<td>1.08</td>
<td>1.79</td>
</tr>
<tr>
<td>Total</td>
<td>1.94</td>
<td>.92</td>
<td>.12</td>
<td>1.69</td>
<td>2.19</td>
</tr>
<tr>
<td>/ðˤ/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learners</td>
<td>3.53</td>
<td>1.22</td>
<td>.21</td>
<td>3.11</td>
<td>3.96</td>
</tr>
<tr>
<td>Non-learners</td>
<td>1.96</td>
<td>.55</td>
<td>.21</td>
<td>1.53</td>
<td>2.39</td>
</tr>
<tr>
<td>Total</td>
<td>2.75</td>
<td>1.23</td>
<td>.15</td>
<td>2.44</td>
<td>3.05</td>
</tr>
</tbody>
</table>
Appendix H8
Post-hoc test (LSD) comparing participants’ production of pharyngealized consonants with varying vowels

<table>
<thead>
<tr>
<th>(I)</th>
<th>(J)</th>
<th>MD (I-J)</th>
<th>SE</th>
<th>Sig.</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phz + /aː/</td>
<td>.29</td>
<td>.13</td>
<td>.03</td>
<td>.03 .55</td>
</tr>
<tr>
<td></td>
<td>Phz + /iː/</td>
<td>.43</td>
<td>.14</td>
<td>.00</td>
<td>.15 .72</td>
</tr>
<tr>
<td></td>
<td>Phz + /uː/</td>
<td>-.29</td>
<td>.13</td>
<td>.03</td>
<td>-.55 -.03</td>
</tr>
<tr>
<td></td>
<td>Phz + /aː/</td>
<td>.14</td>
<td>.13</td>
<td>.27</td>
<td>-.12 .41</td>
</tr>
<tr>
<td></td>
<td>Phz + /iː/</td>
<td>-.43</td>
<td>.14</td>
<td>.00</td>
<td>-.72 -.15</td>
</tr>
<tr>
<td></td>
<td>Phz + /uː/</td>
<td>-.14</td>
<td>.13</td>
<td>.27</td>
<td>-.41 .12</td>
</tr>
</tbody>
</table>

Appendix H9
Descriptive statistics of both groups’ production scores for pharyngealized CVs with different vowels

<table>
<thead>
<tr>
<th>Participants groups</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharyngealized + /aː/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learners</td>
<td>4.04</td>
<td>.72</td>
<td>20</td>
</tr>
<tr>
<td>Non-learners</td>
<td>2.18</td>
<td>.64</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>3.11</td>
<td>1.15</td>
<td>40</td>
</tr>
<tr>
<td>Pharyngealized + /iː/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learners</td>
<td>3.48</td>
<td>.63</td>
<td>20</td>
</tr>
<tr>
<td>Non-learners</td>
<td>2.16</td>
<td>.90</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>2.82</td>
<td>1.02</td>
<td>40</td>
</tr>
<tr>
<td>Pharyngealized + /uː/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learners</td>
<td>3.30</td>
<td>.90</td>
<td>20</td>
</tr>
<tr>
<td>Non-learners</td>
<td>2.05</td>
<td>.52</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>2.68</td>
<td>.96</td>
<td>40</td>
</tr>
</tbody>
</table>
Appendix H (Continued)

Appendix H10

Descriptive statistics of both groups’ production scores for the pharyngeal CVs with different vowels: /ħaː/, /ʕaː/; /ħiː/, /ʕiː/ and /ħuː/, /ʕuː/:

<table>
<thead>
<tr>
<th>Participants’ groups</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharyngeal + /aː/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learners</td>
<td>4.89</td>
<td>.29</td>
<td>20</td>
</tr>
<tr>
<td>Non-learners</td>
<td>2.69</td>
<td>.87</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>3.79</td>
<td>1.29</td>
<td>40</td>
</tr>
<tr>
<td>Pharyngeal + /iː/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learners</td>
<td>4.63</td>
<td>.83</td>
<td>20</td>
</tr>
<tr>
<td>Non-learners</td>
<td>2.26</td>
<td>.79</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>3.44</td>
<td>1.44</td>
<td>40</td>
</tr>
<tr>
<td>Pharyngeal + /uː/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learners</td>
<td>4.58</td>
<td>.56</td>
<td>20</td>
</tr>
<tr>
<td>Non-learners</td>
<td>2.10</td>
<td>.93</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>3.34</td>
<td>1.47</td>
<td>40</td>
</tr>
</tbody>
</table>

Appendix H11

Post-hoc test (LSD) comparing participants’ production of pharyngeal CVs varying in adjacent vowels:

<table>
<thead>
<tr>
<th>(I)</th>
<th>(J)</th>
<th>MD</th>
<th>SE</th>
<th>Sig.</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharyngeal + /aː/</td>
<td>Pharyngeal + /iː/</td>
<td>.34</td>
<td>.13</td>
<td>.01</td>
<td>.08</td>
</tr>
<tr>
<td>Pharyngeal + /aː/</td>
<td>Pharyngeal + /uː/</td>
<td>.45</td>
<td>.15</td>
<td>.00</td>
<td>.15</td>
</tr>
<tr>
<td>Pharyngeal + /iː/</td>
<td>Pharyngeal + /aː/</td>
<td>-.34</td>
<td>.13</td>
<td>.01</td>
<td>-.61</td>
</tr>
<tr>
<td>Pharyngeal + /iː/</td>
<td>Pharyngeal + /uː/</td>
<td>.11</td>
<td>.17</td>
<td>.53</td>
<td>-.23</td>
</tr>
<tr>
<td>Pharyngeal + /uː/</td>
<td>Pharyngeal + /aː/</td>
<td>-.45</td>
<td>.15</td>
<td>.00</td>
<td>-.75</td>
</tr>
<tr>
<td>Pharyngeal + /uː/</td>
<td>Pharyngeal + /iː/</td>
<td>-.11</td>
<td>.17</td>
<td>.53</td>
<td>-.44</td>
</tr>
</tbody>
</table>
## Appendix I

### Appendix I1

Matrix of Arabic non-learners’ production of pharyngealized consonant cohorts /tˤ, sˤ, dˤ, ðˤ/

<table>
<thead>
<tr>
<th>Sound</th>
<th>Arabic L1 listeners’ perception (% of occurrence)</th>
<th>Recurrent</th>
<th>% of recurrent responses^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>/tˤaː/</td>
<td>/tˤ/ 17.5</td>
<td>/t/ 20</td>
<td>/ð/ 13.3</td>
</tr>
<tr>
<td>/tˤiː/</td>
<td>/tˤ/ 27.5</td>
<td>/t/ 10</td>
<td>/t/ 15</td>
</tr>
<tr>
<td>/tˤuː/</td>
<td>/tˤ/ 27.5</td>
<td>/t/ 20</td>
<td>/t/ 20</td>
</tr>
<tr>
<td>/tˤ/</td>
<td>/tˤ/ 24.2</td>
<td>/t/ 16.7</td>
<td>/t/ 16.7</td>
</tr>
<tr>
<td>/sˤaː/</td>
<td>/sˤ/ 50</td>
<td>/s/ 50</td>
<td>/sˤ/ 50</td>
</tr>
<tr>
<td>/sˤiː/</td>
<td>/sˤ/ 72.5</td>
<td>/s/ 25</td>
<td>/sˤ/ 72.5</td>
</tr>
<tr>
<td>/sˤuː/</td>
<td>/sˤ/ 20</td>
<td>/s/ 80</td>
<td>/sˤ/ 20</td>
</tr>
<tr>
<td>/sˤ/</td>
<td>/sˤ/ 47.5</td>
<td>/s/ 51.7</td>
<td>/sˤ/ 47.5</td>
</tr>
<tr>
<td>/dˤaː/</td>
<td>/dˤ/ 0</td>
<td>/d/ 0</td>
<td>/dˤ/ 0</td>
</tr>
<tr>
<td>/dˤiː/</td>
<td>/dˤ/ 5</td>
<td>/d/ 15</td>
<td>/dˤ/ 5</td>
</tr>
<tr>
<td>/dˤuː/</td>
<td>/dˤ/ 7.5</td>
<td>/d/ 65</td>
<td>/dˤ/ 7.5</td>
</tr>
<tr>
<td>/dˤ/</td>
<td>/dˤ/ 4.2</td>
<td>/d/ 26.7</td>
<td>/dˤ/ 4.2</td>
</tr>
<tr>
<td>/ðˤaː/</td>
<td>/ðˤ/ 35</td>
<td>/ð/ 0</td>
<td>/ðˤ/ 35</td>
</tr>
<tr>
<td>/ðˤiː/</td>
<td>/ðˤ/ 15</td>
<td>/ð/ 25</td>
<td>/ðˤ/ 15</td>
</tr>
<tr>
<td>/ðˤuː/</td>
<td>/ðˤ/ 17.5</td>
<td>/ð/ 10</td>
<td>/ðˤ/ 17.5</td>
</tr>
<tr>
<td>/ðˤ/</td>
<td>/ðˤ/ 22.5</td>
<td>/ð/ 11.7</td>
<td>/ðˤ/ 22.5</td>
</tr>
</tbody>
</table>

^aPercentages of recurrent responses are presented in the same order in which the sound responses are presented.
### Appendix I2

Matrix of Arabic non-learners’ production of pharyngeal consonant cohorts /ħ,ʕ/

<table>
<thead>
<tr>
<th>Sound</th>
<th>Arabic L1 listeners’ perception (% of occurrence)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/ħ/</td>
</tr>
<tr>
<td>/ha:/</td>
<td>35</td>
</tr>
<tr>
<td>/hi:/</td>
<td>12.5</td>
</tr>
<tr>
<td>/hu:/</td>
<td>12.5</td>
</tr>
<tr>
<td>/h/</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>/ʕa:/</td>
<td>5</td>
</tr>
<tr>
<td>/ʕi:/</td>
<td>50</td>
</tr>
<tr>
<td>/ʕu:/</td>
<td>55</td>
</tr>
<tr>
<td>/ʕ/</td>
<td>55</td>
</tr>
</tbody>
</table>

<sup>a</sup>Percentages of recurrent responses are presented in the same order in which the sound responses are presented.

### Appendix I3

Matrix of Arabic learners’ production of pharyngeal consonant cohorts /ħ,ʕ/

<table>
<thead>
<tr>
<th>Sound</th>
<th>Arabic L1 listeners’ perception (% of occurrence)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/ħ/</td>
</tr>
<tr>
<td>/ha:/</td>
<td>100</td>
</tr>
<tr>
<td>/hi:/</td>
<td>97.5</td>
</tr>
<tr>
<td>/hu:/</td>
<td>87.5</td>
</tr>
<tr>
<td>/h/</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>/ʕa:/</td>
<td>97.5</td>
</tr>
<tr>
<td>/ʕi:/</td>
<td>87.5</td>
</tr>
<tr>
<td>/ʕu:/</td>
<td>92.5</td>
</tr>
<tr>
<td>/ʕ/</td>
<td>92.3</td>
</tr>
</tbody>
</table>

<sup>a</sup>Percentages of recurrent responses are presented in the same order in which the sound responses are presented.
Appendix I (continued)

Appendix I4
Matrix of Arabic learners’ production of pharyngealized consonant cohorts /tˤ, sˤ, dˤ, ðˤ/

<table>
<thead>
<tr>
<th>Sound</th>
<th>Arabic L1 listeners’ perception (% of occurrence)</th>
<th>Recurrent</th>
<th>% of recurrent responses (^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/tˤaː/</td>
<td>/tˤ/ 95 0 0</td>
<td>/q/</td>
<td>5</td>
</tr>
<tr>
<td>/tˤiː/</td>
<td>/tˤ/ 95 5 0</td>
<td>/tˤ/</td>
<td>95</td>
</tr>
<tr>
<td>/tˤuː/</td>
<td>/tˤ/ 90 2.5 2.5</td>
<td>/q/</td>
<td>2.5</td>
</tr>
<tr>
<td>/tˤ/</td>
<td>/tˤ/ 93.3 2.5 .8</td>
<td>/q/</td>
<td>2.5</td>
</tr>
<tr>
<td>/sˤaː/</td>
<td>/sˤ/</td>
<td>/z/</td>
<td>Recurrent</td>
</tr>
<tr>
<td>/sˤiː/</td>
<td>/sˤ/ 100 0 0</td>
<td>/sˤ/</td>
<td>100</td>
</tr>
<tr>
<td>/sˤuː/</td>
<td>/sˤ/ 80 20 0</td>
<td>/sˤ/</td>
<td>80</td>
</tr>
<tr>
<td>/sˤ/</td>
<td>/sˤ/ 85 15 0</td>
<td>/sˤ/</td>
<td>85</td>
</tr>
<tr>
<td>/sˤ/</td>
<td>/sˤ/ 88.3 11.7 0</td>
<td>/sˤ/</td>
<td>88.3</td>
</tr>
<tr>
<td>/dˤaː/</td>
<td>/dˤ/ 22.5 0 0</td>
<td>/θˤ/</td>
<td>40</td>
</tr>
<tr>
<td>/dˤiː/</td>
<td>/dˤ/ 5 0 7.5</td>
<td>/tˤ/; /θˤ/</td>
<td>85</td>
</tr>
<tr>
<td>/dˤuː/</td>
<td>/dˤ/ 52.5 12.5 0</td>
<td>/tˤ/; /θˤ/</td>
<td>25; 5</td>
</tr>
<tr>
<td>/dˤ/</td>
<td>/dˤ/ 26.7 4.2 2.5</td>
<td>/tˤ/; /θˤ/</td>
<td>36.7; 15</td>
</tr>
<tr>
<td>/ðˤaː/</td>
<td>/ðˤ/ 85 2.5 2.5</td>
<td>/dˤ/</td>
<td>5</td>
</tr>
<tr>
<td>/ðˤiː/</td>
<td>/ðˤ/ 67.5 12.5 0</td>
<td>/θˤ/; /tˤ/</td>
<td>12.5; 10</td>
</tr>
<tr>
<td>/ðˤuː/</td>
<td>/ðˤ/ 37.5 32.5 0</td>
<td>/θˤ/; /vˤ/</td>
<td>32.5; 7.5</td>
</tr>
<tr>
<td>/ðˤ/</td>
<td>/ðˤ/ 63.3 15.8 .8</td>
<td>/θˤ/</td>
<td>15</td>
</tr>
</tbody>
</table>

\(^a\)Percentages of recurrent responses are presented in the same order in which the sound responses are presented.