The Perception and Production of Arabic Lexical Stress by Learners of Arabic:

A Usage-Based Account

By

Cheng-Wei Lin

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Near Eastern Studies)
in The University of Michigan
2018

Doctoral Committee:

Professor Mohammad T. Alhawary, Chair
Professor Patrice S. Beddor
Professor Nick C. Ellis
Professor Emeritus Raji M. Rammuny

Cheng-Wei Lin

cwlin@umich.edu

ORCID iD: 0000-0001-7295-9120

Acknowledgements

It has been a long journey, and I would like to thank all the people who have helped me and encouraged me these years.

First of all, I would like to thank the committee members for their feedback and time dedicated to my dissertation. I would like to especially thank my advisor Professor Mohammad Alhawary, who opened to me the different fields of Arabic linguistics, and has always motivated me while intellectually challenged me with his thorough feedback on my work to help me become a better researcher. I would like to thank Professor Patrice Beddor, who taught the first theoretical linguistics class I have ever taken formally, and has provided me inspiration and feedback since. I would also like to thank Professor Nick Ellis, who has inspired me of the paradigm examined in this dissertation, and has provided substantial advice in experimental testing. Lastly, I would like to thank Professor Raji Rammuny, who has taught me how to be a better Arabic teacher and who has always been supportive of my research.

Secondly, I would like to thank the people who made the experiment in this dissertation possible. I would like to thank the faculty members at National Chengchi University and Georgetown University for their help. I would like to especially thank Professor Huey-Tsyr Jeng and Professor Felicitas Opwis for giving me the permission to conduct my research there. I also sincerely thank all the learners of Arabic who participated in the experiment. If it were not for them, nothing would come out of this dissertation.

Thirdly, I would like to thank the people around me for their understanding and support. I would like to thank my cohorts from the department of Near Eastern Studies, especially for Lizz Huntley for her support and help with the dissertation. I would like to thank the members of the Phondi discussion group and the Psycholinguistics discussion group at U of M for allowing me to present my work and for their feedback on this study. My colleagues at Michigan State University must be thanked as well for their constant support and understanding, which makes juggling between teaching Arabic and completing my dissertation possible.

I would like to thank Y for her constant support and company through the toughest times. Lastly, I am deeply indebted to my parents for the years that I have been intermittently absent. I would like to thank their constant support of me and their belief in me. Without them, none of this could have been possible. This dissertation is dedicated to them.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF FIGURES viii

LIST OF TABLES x

LIST OF APPENDICES xii

ABSTRACT xiii

Chapter 1

Introduction 1

1.1 Arabic L2 Stress Perception and Production Studies 2

1.2 Objectives of the Dissertation 2

1.3 Organization of the Dissertation 4

Chapter 2

Background: Perception and Production of L2 Stress 6

2.1 Lexical Stress in Arabic 7

2.1.1 The Phonology of Stress 7

2.1.2 The Phonetics of Lexical Stress in Arabic 12

2.2 Studies in L2 Stress Perception and Production 13

2.2.1 L2 Stress Perception Studies 14

2.2.2 L2 Stress Production Studies 23

2.2.3 L1 and L2 Stress and Perception and Production and Arabic 26

2.2.4 Conclusions from L2 Stress Perception and Production Studies 28

2.3 Studies on Frequency Effect In Language Processing and Learning 29

2.3.1 Frequency Effect on L1 Learning and Processing 29

2.3.2 A Usage-Based Approach to Second Language Acquisition 32

2.4 Frequency and Stress Perception and Production 33

2.5 Conclusion ... 35

Chapter 3

Methodology ... 37

3.1. Experiment Design ... 38

3.1.1 Warm-Up Activity: Speed Reading ... 39

3.1.2 Stress Identification Task ... 40

3.1.3. Lexical Decision Task ... 43

3.1.4. Stress Production Task ... 45

3.1.5. Post-Experiment Questionnaire ... 49

3.1.6. Organization of the Experiment ... 53

3.2. Stimulus Design ... 57

3.2.1 Estimating Frequency Distribution of Stress Patterns in Arabic ... 62

3.2.2 Data Extraction ... 64

3.2.3 Frequency Distribution of Stress Pattern in Arabic ... 70

3.2.4 Correlation between the Sources ... 82

3.2.5 Stimulus Organization ... 91

3.2.6 Stimulus Construction ... 97

3.3. Participants ... 106

3.2.1 Recruitment Criteria ... 107

3.2.2 Recruitment Procedure ... 113

3.2.3 Recruitment Results ... 113

3.4. Data Analysis ... 115

3.5 Summary ... 119

Chapter 4

Results ... 123

4.1 Observation of Classroom Instruction ... 124

4.2 Warm-Up Task: Speed Reading ... 131

4.3 Stress Identification Task ... 134

4.3.1 Native Language and Performance ... 135

4.3.2 Proficiency and Performance ... 138

4.3.3 Stress Pattern and Performance 142
4.3.4 Relative Frequency and Performance 146
4.3.5 Summary of Results 153
4.4 Lexical Decision Task 155
4.4.1 Native Language and Performance 155
4.4.2 Proficiency and Performance 158
4.4.3 Stress Pattern and Performance 164
4.4.4 Relative Frequency and Performance 165
4.4.5 Summary of Results 168
4.5 Stress Production Task 171
4.5.1 Native Language and Performance 172
4.5.2 Proficiency and Performance 176
4.5.3 Stress Pattern and Performance 180
4.5.4 Relative Frequency and Performance 181
4.5.5 Summary of Results 186
4.6 Sub-Group Comparisons – Input Variance 189
4.7 Sub-Group Comparisons – Type of Nonsense Word 192
4.8 External Factors from the Questionnaire 193
4.8.1 Instruction and Textbook 195
4.8.2 Self-Awareness of Stress 198
4.8.3 Extracurricular Engagement 200
4.8.4 Explicit Knowledge 201
4.8.5 External Factors and Experiment Results 203
4.9 Summary of Results 206
4.9.1 Native Language 206
4.9.2 Proficiency 207
4.9.3 Relative Frequency 208
4.9.4 External Factors 209

Chapter 5

Discussion and Conclusion 211

5.1 Discussion 213
5.1.1 L1 Transfer 213
5.1.2 The Effect of Proficiency 220
5.1.3 Frequency Effect 222
5.1.4 Other External Factors 224
5.2 Limitations of the Study 227
5.3 Conclusion and Pedagogical Implications 229

APPENDICES 234
BIBLIOGRAPHY 283

LIST OF FIGURES

Figure 3.1 Layout of the Control Panel......39
Figure 3.2 Sample of Stimuli Presentation in the Stress Identification Task......42
Figure 3.3 Procedure of Stress Identification and Lexical Decision Task......44
Figure 3.4 Procedure of Stress Production Task......46
Figure 3.5 Sample Stimulus for the Stress Production Task......48
Figure 3.6 Sequencing of Questions in the Post-Experiment Questionnaire......50
Figure 3.7 Sequencing of Tasks in the Experiment......54
Figure 3.8 Data Processing Procedure for Frequency Dictionary......65
Figure 3.9 Correlation between Log Frequency and Rank of All Stress Patterns......71
Figure 3.10 Correlation between Log Frequency and Syllable Number of All Stress Patterns......72
Figure 3.11 Frequency Distribution of Bisyllabic Stress Patterns......73
Figure 3.12 Frequency Distribution of Trisyllabic Stress Patterns......73
Figure 3.13 Frequency Distribution of Quadrisyllabic Stress Patterns......74
Figure 3.14 Frequency Distribution of Syllabic Structure of Stress (All Stress Patterns)......75
Figure 3.15 Frequency Distribution of Syllabic Structure of Stress Grouped by Syllable Number......77
Figure 3.16 Frequency Distribution of Stress Position (All Stress Patterns)......78
Figure 3.17 Frequency Distribution of Stress Position Grouped by Syllable Number......79
Figure 3.18 Conditional Probability of Stress for All Stress Patterns......81
Figure 3.19 Correlation between Frequency Dictionary and Hans Wehr Dictionary......85
Figure 3.20 Correlation between Token-Type Frequency Ratio and Frequency Rank......87
Figure 3.21 Correlation between Frequency Dictionary and Al-Kitaab Textbooks......89
Figure 3.22 Stimuli Configuration......104
Figure 3.23 Stress Typology Model......107
Figure 3.24 Sample Analysis for Stress Production Responses......116
Figure 4.1 Boxplot for Duration of Completion for Speed Reading According to L1......132
Figure 4.2 Boxplot for Duration of Completion for Speed Reading Drill According to L1 and Proficiency......133
Figure 4.3. Boxplot for the Response Accuracy for the Stress Identification Task According to L1......136

Figure 4.4. Boxplot for the Reaction Time Grouped by Native Language for the Stress Identification Task....137
Figure 4.5. Boxplot for the Response Accuracy for the Stress Identification Task According to Participants' L1 and Proficiency 139
Figure 4.6. Boxplot for the Reaction Time Grouped by Participants' L1 and Proficiency for the Stress Identification Task 140
Figure 4.7. Boxplot for the Response Accuracy (Real Words) Grouped by Participants' L1 and Proficiency for the Stress Identification Task....142
Figure 4.8. Boxplot for the Response Accuracy for the Lexical Decision Taskaccordingly to Participants' L1 156
Figure 4.9. Boxplot for the Reaction Time Grouped by Native Language for the Lexical Decision Task 157
Figure 4.10. Boxplot for the Correctness Rate Grouped by Participants' L1 and Proficiency for the Lexical Decision Task....158
Figure 4.11. Boxplot for the Reaction Time Grouped by Participants' L1 and Proficiency for the Lexical Decision Task....160
Figure 4.12. Boxplot for the Response Accuracy for the Lexical Decision Task According to L1 and Lexical Status of Stimuli....161
Figure 4.13. Boxplot for the *d*-prime Score for the Lexical Decision Task According to L1 and Proficiency 163
Figure 4.14. Boxplot for the Response Accuracy for the Stress Production Task According to Native Language....173
Figure 4.15. Boxplot for the Reaction Time for the Stress Production Task According to Native Language 175
Figure 4.16. Boxplot for the Correctness of Response Grouped by Participants' L1 and Proficiency for the Stress Production Task (Trial 1) 176
Figure 4.17. Boxplot for the Reaction Time Grouped by Participants' L1 and Proficiency for the Stress Production Task (Trial 1)....179
Figure 4.18 Scores for Instruction and Textbook According to L1 and Proficiency....197
Figure 4.19 Scores for Self-Evaluation Grouped by L1 and Proficiency 199
Figure 4.20 Scores for Extracurricular Engagement Grouped by L1 and Proficiency 201
Figure 4.21 Explicit Knowledge Score According to L1 and Proficiency 202

LIST OF TABLES

Table 3.1 Schema of Variable Manipulation 95
Table 3.2 Type of Co-Occurrence of Root Consonants and Its Conditional Probability (Partial)........ 101
Table 3.3 Sample Block of Nonsense Words Derived from Real Word........ 102
Table 3.4 Sample Block of Nonsense Words Derived from Unattested Root [Fqz]........ 103
Table 3.5 Composition of Participants........ 114
Table 3.6 Correlates for A Hypothetical Production 118
Table 3.7 General Difference in Correlates between Stressed and Unstressed Syllables........ 119
Table 4.1. Stress Patterns Included in the Experiment........ 143
Table 4.2. Summary Table for Main Effect for Stress Pattern 2........ 144
Table 4.3. Summary Table for Multiple Comparisons for Stress Pattern 2 145
Table 4.4 Groups of Stress Patterns with Contrastive Relative Frequency 147
Table 4.5 Summary of Frequency Effect on Correctness of Response for Stress Identification Task (Partial) 148
Table 4.6 Summary of Frequency Effect on Reaction Time for Stress Identification Task (Partial)........ 150
Table 4.7 Summary Table for Findings of Stress Identification Task 154
Table 4.8 Summary of Frequency Effect on Correctness of Response for Lexical Decision Task (Partial) 166
Table 4.9 Summary Table for the Findings from Lexical Decision Task........ 168
Table 4.10 Summary of Frequency Effect on Response Accuracy for Stress Production Task (Partial)........ 182
Table 4.11 Summary of Frequency Effect on Reaction Time for Stress Production Task for Trial 1 (Partial) 184
Table 4.12 Summary Table for Findings in Stress Production Task 187
Table 4.13 Summary Table for Participants’ Responses to Question 1........ 195
Table 4.14 Perceived Stress Rule Introduced by the Arabic Teacher........ 196
Table 4.15 Subgroup Comparison for the Most-Mentioned Stress Rules 196
Table A.3.7 Conditional Probability of Arabic Roots........ 245
Table A.3.8.1 Nonsense Words Derived from Unattested Root and Template: 250
Table A.3.8.2 Nonsense Words Derived from Real Words:........ 251
Table A.4.1.1 Summary Table for Response Accuracy (Stress Identification Task) 254

Table A.4.1.2 Summary Table for Reaction Time (Stress Identification Task) 256
Table A.4.2.1. Summary for Response Accuracy (Stress Identification) 259
Table A.4.2.2. Summary for Reaction Time (Stress Identification) 260
Table A.4.3.1 Summary Table for Response Accuracy (Lexical Decision Task) 261
Table A.4.3.2 Summary Table for Reaction Time (Lexical Decision Task) 263
Table A.4.4.1. Summary for Response Accuracy (Lexical Decision) 264
Table A.4.4.2. Summary for Reaction Time (Lexical Decision) 265
Table A.4.5.1 Summary Table for Response Accuracy (Stress Production Task) 266
Table A.4.5.2 Summary Table for Response Accuracy (Stress Production Task) 268
Table A.4.6.1. Summary for Response Accuracy (Stress Production) 270
Table A.4.6.2. Summary for Reaction Time (Stress Production) 271
Table A.4.7.1. Reaction Time Comparisons with and without Graduate Students.................................. 272
Table A.4.7.2. Response Accuracy Comparisons with and without Graduate Students........................... 273
Table A.4.8.1 Comparisons between Types of Nonsense Words (Stress Identification Task)................ 274
Table A.4.8.2 Comparisons between Types of Nonsense Words (Lexical Decision Task)...................... 275
Table A.4.8.3 Comparisons between Types of Nonsense Words (Stress Production Task)..................... 276
Table A.4.9.1 Degree of Agreement for Teaching Factors Questions:.. 277
Table A.4.9.2. Degree of Agreement for Personal Evaluation Questions: 277
Table A.4.9.3. Degree of Agreement for Extracurricular Engagement Questions: 278
Table A.4.9.4. Degree of Correctness for Quiz on Stress Rules Questions:.. 278
Table A.4.11.1 Correlation between External Factor Scores (Native Arabic Group) 280
Table A.4.11.2 Correlation between External Factor Scores (English-Speaking Group)......................... 281
Table A.4.11.3 Correlation between External Factor Scores (Chinese-Speaking Group)........................ 282

LIST OF APPENDICES

Appendix 3.1 Task Description for the Warm-Up Activity 234
Appendix 3.2 Wordlists Used for the Warm-Up Activity.......... 235
Appendix 3.3 Task Description for the Stress Identification Task and Lexical Decision 236
Appendix 3.4 Task Description for the Stress Production Task 238
Appendix 3.5 Questionsf for Post-Experiment Questionnaire.......... 239
Appendix 3.6 Introduction to Lexical Stress in English 242
Appendix 3.7 Type of Co-Occurrence of Root Consonants and Its Conditional Probability.......... 245
Appendix 3.8 Stimuli.......... 250
Appendix 3.9 Samples of Recruitment Flyer in English and Chinese.......... 253
Appendix 4.1 Summary Tables for Multiple Comparisons for Stress Identification 254
Appendix 4.2 Summary Tables of Frequency Effect for Stress Identification Task.......... 259
Appendix 4.3 Summary Tables for Multiple Comparisons for Lexical Decision.......... 261
Appendix 4.4 Summary Tables of Frequency Effect for Lexical Decision Task 264
Appendix 4.5 Summary Tables for Multiple Comparisons for Stress Production (Trial 1).......... 266
Appendix 4.6 Summary Tables of Frequency Effect for Stress Production Task (Trial1) 270
Appendix 4.7 Comparisons between without and with Graduate Students 272
Appendix 4.8 Comparisons between Type of Nonsense Words.......... 274
Appendix 4.9 Questionnaire Results 277
Appendix 4.10 Sample of Stress Rules Provided by the Participants in the Questionnaire.......... 279
Appendix 4.11 Correlation between External Factor Scores and Experiment Results.......... 280

Abstract

Studies in second language (L2) stress perception and production over the past few decades have focused on the role of the native language (L1) of the L2 learner and how it systematically influences their performance in stress perception and production. However, these studies have not adequately explored and incorporated an important factor: input frequency of stress patterns (henceforth frequency), a factor that has been widely explored in other disciplines and has been found to be crucially relevant to language processing and learning. To bridge this gap in the literature, this study examines the effect of frequency, in addition to the role played by the learners' L1, on the perception and production of primary lexical stress in Arabic by L2 learners of Arabic in an experimental environment. To this end, a stress perception and production experiment was conducted on first- and second-year L1 English and L1 Chinese learners of Arabic as well as L1 Arabic speakers as controls. In the experiment, the participants completed a stress production task, a stress identification task, and a lexical decision task, where they were asked to produce stimuli that were nonsense words with frequency-biased stress patterns, to listen then identify the position of the stressed syllable in these frequency-biased stimuli, and to determine whether the stimuli were real Arabic words or not, respectively.

The results indicate a more evident frequency effect in the stress production task, where it had a local effect on learners' performance on stimuli. Specifically, they were significantly quicker and more accurate in

producing the stimuli when the stimuli had a frequent stress pattern whereas they were slower and less accurate when the stimuli had infrequent stress patterns.

In contrast, the results show a more global effect of participants' L1 on their perception and production of stress, as typological differences were found in the performance in the stress perception and production tasks. L1 speakers of Arabic consistently had slower and less accurate performance than the L2 learners in the stress identification task. The L1 Chinese participants had systematically more fluent and accurate production than their L1 English counterparts, which is argued to be contributed by the L2 Chinese participants' better utilization of correlates for stress.

These findings are taken to be in partial support for the role of frequency in stress perception and production, as significant differences were found in contexts with larger frequency contrast but not ones with moderate-to-small frequency contrast, and the fact that the performance of the participants was, to a large extent, conditioned by the preferences for acoustic cues and prosodic characteristics of their L1. However, frequency of input should not be disregarded, since it did capture aspects of learners' performance that were not conditioned by their L1. Future studies should build upon the method implemented in the present study to further explore the role of frequency in L2 learners in higher proficiency as well.

Pedagogically, the findings from the present study provided several implications for current Arabic curricular development and teaching practices, including raising the awareness of Arabic instructors of lexical stress and the importance of lexical stress for L2 teaching, developing teaching materials, and reflecting on current curricula practices that simultaneously engage multiple varieties exhibiting different stress systems.

Chapter 1

Introduction

Among the many skills that second language learners need to acquire is learning suprasegmental features, particularly learning how to stress a word in the target language. The ability to locate and produce stress on the correct syllable of the word is an integral part of good pronunciation, which promotes better communication, comprehension and overall oral proficiency. Failing to do so will result in what is generally perceived as having a "foreign accent", which might cause misunderstanding, a false perception of low proficiency and eventually lead to communication breakdowns. Such situations are never advantageous for second language learners, and particularly for those learning Arabic as a second language.

However, acquiring a second language is rather different from acquiring a first language, especially in terms of the amount and the type of language exposure that learners encounter. As opposed to first language acquisition, which normally takes place in the learner's social surroundings, second language acquisition normally occurs in the second language classroom with limited exposure time inside and outside of the classroom. The starting point for learners is distinctly different as well: first language learners do not have any previous language background, while second language learners approach their second language having already successfully acquired another language (i.e. their native language). These two differentiating factors between L1 and L2 learners (i.e., learners' native language and the language input) have been the center of investigation in the second language acquisition literature.

1.1 Arabic L2 Stress Perception and Production Studies

The L2 stress perception and production literature primarily echoes only the second of these two factors with most studies emphasizing the role played by the L1 of the learners and how that L1 could influence the outcome of their acquisition of L2 stress (often referred to as L1 transfer effect). Studies on L2 stress perception and production have mainly been contributed by studies in the field of English as the second language. Even though Arabic has been an increasingly important language due to its political, cultural and social relevance and is regarded as a critical language in the United States, studies on Arabic L2 stress perception and production are woefully lacking. Furthermore, the effect of language input with regards to the type and the amount/frequency of the input is generally absent or not controlled for in the L2 stress perception and production literature. As a result, not much is known about the extent to which the effect of language input could inform the perception and production of stress by L2 learners. Furthermore, failing to incorporate input factors poses challenges to the findings yielded from these L2 stress perception and production studies. The amount and frequency of input has been widely and increasingly explored in the field of psychology and psycholinguistics and has been identified to be shown to influence various aspects of language processing and learning.

1.2 Objectives of the Dissertation

To bridge this gap in both the second language acquisition literature in general and Arabic second language acquisition for stress perception and production in particular, the present study set out to incorporate the two aforementioned factors for L2 studies (i.e. the effect from the native language of the learners and the input that they are exposed to) in the study design and to examine how these two factors could inform stress perception and production. Precisely, I conducted (1) a corpus analysis attempting to approximate the type and the amount/frequency of stress in the input that the learners encounter when learning Arabic as a second language, (2) a stress perception and production experiment that evaluates the accuracy and fluency of L2 learners of Arabic in perceiving and producing stress, complemented by (3) a short exploratory questionnaire to probe other external, non-experimental factors that might contribute to learners' performance as well as their individual variability in performance. With these measures, the present study addresses the following questions with regards to L2 stress perception and production:

- RQ 1. Does the native language of the learners influence their performance in stress perception and production, and, if so, how?
- RQ 2. Does the proficiency of the learners influence their performance in stress perception and production, and, if so, how?
- RQ 3. Does the frequency of relevant structures for stress (i.e., position of the stressed syllable, syllabic structure of the stressed syllable, and stress pattern of the word) influence the participants' performance on stress perception and production, and, if so, how?
- RQ 4. Does frequency have a different effect on stress perception as compared to stress production?
- RQ 5. Does frequency have a different effect on learners as compared to native speakers?
- RQ 6. Do other external factors (specifically explicit knowledge of lexical stress, instruction, self-awareness and extracurricular engagement with the target

language) influence learners' performance on stress perception and production, and, if so, how?

The results of the experiment show evidence of a L1 transfer effect and an effect of input frequency, but to varying degrees. Unsurprisingly, L2 stress perception and production is to a large extent conditioned by the learners' native language, in that the accuracy of stress perception is increased by the presence of stress in their native language and is decreased by no lexical stress in the L1. Stress production is also found to be largely conditioned by the learners' L1, but in a less intuitive way. Learners with stress in their native language were found to have be less accurate in their production of stress while learners without stress in their native language were found to have generally better accuracy, which in turn suggests that learners' native language could promote some aspect of L2 stress perception and production but could impede some others at the same time. With regards to the input frequency effect, the experiment only found a local effect of language input on performance, in that learners exhibited better performance when the target structure is significantly more frequent in the target language and worse performance when the target structure is infrequent.

The findings from the experiment generally align with the conclusions drawn by previous L2 stress perception and production studies, but novel to the field is the local effect found for language input in the present study, which highlights the role of input frequency that has been long neglected in the stress perception and production literature.

1.3 Organization of the Dissertation

The dissertation is composed of the following chapters. The present chapter, i.e. Chapter 1, provides an overview of the research questions and a brief summary of the dissertation.

Chapter 2 provides a discussion of the phonological and phonetic properties of lexical stress in Arabic, a review of previous studies in L2 stress perception and production, and a summary of the tenets of usage-based approach to second language acquisition. Chapter 3 presents the study design of stress perception and production and the considerations taken in account when constructing stimuli, recruiting participants and analyzing data. Chapter 4 presents the results of the experiment and the implications with regards to the variables examined in the present study. Chapter 5 discusses the results of the experiment and how they relate to findings in the L2 stress perception and production literature and in the second language acquisition literature, and concludes with pedagogical implications.

Chapter 2

Background: Perception and Production of L2 Stress

Studies that investigate second language (henceforth L2) stress perception and production have gained momentum over the past few decades. Many studies have explored the topic from different perspectives, mainly focusing on how various factors could inform the degree to which speakers correctly perceive and produce stress. Factors of interest include internal factors such as the native language of the speaker (L1 transfer), the phonological properties of stress (such as syllabic structure and syllable weight), the acoustic properties of stress and the interaction of all these factors.

Although efforts have been made to expand the scope of studies to a broader range of learners of different native languages, a full understanding of stress perception and production has yet to emerge. An expanded scope would require more investigation, but past studies reported inconsistent results and had some limitation in terms of stimuli design since the majority of previous L2 stress perception and production studies have used real words as stimuli without controlling for the familiarity effect. Besides, factors considered to be relevant for language learning and processing in the field of psycholinguistics, such as frequency of input, do not seem to have been discussed in the literature. Before even delving into these studies, however, there are fundamental questions which need to be addressed. Questions of primary importance to this study are: what constitutes stress in a language in general and in Arabic in particular; and what are the phonological and acoustic characteristics of stress in Arabic?

By addressing these questions, the objective of this chapter is to better contextualize the present study within the literature of stress perception and production. The remainder of the chapter is organized as follows: Section 2.1 discusses the phonological and phonetic properties of lexical stress in Arabic. This is followed by a discussion of the findings from relevant previous stress perception and production studies in Section 2.2. Finally, a summary of the tenets of usage-based approach to second language acquisition is provided in Section 2.3.

2.1 Lexical Stress in Arabic

2.1.1 The Phonology of Stress

When a word has lexical stress, it means that one syllable of the word is perceived to be more prominent than other syllables in the word. A stressed syllable is also able to bear pitch movement which tends to have a significant perceptual load. Although it is not always the case, a stressed syllable tends to have relatively long duration and high intensity (Kager, 1996) or is simply louder (Ladefoged & Johnson, 2010). It has also been pointed out that the distinctive cues for a stressed syllable are its F0 contour and duration (Fry, 1958). Phonologically speaking, the general properties of stress are that (1) each word has at least one such "prominent" element (the cumulative property), (2) the location of stress is usually near the beginning or the end of a word (the demarcative property) and (3) stress placement is sensitive to the length of the syllable (weight-sensitivity).

Lexical stress in Arabic seems to strongly conform to these three properties, i.e., the cumulative property, the demarcative property and the weight-sensitivity property. Phonologically speaking, each Arabic content word has at least one stressed syllable. The distribution of the stressed syllable is fairly close to the edges of the word. Furthermore, stress assignment in Arabic is highly influenced by the weight of the surrounding syllables in the word. In Arabic, there are three types of syllable, which are the light syllable (CV), the heavy syllable (CVV, CVC) and the superheavy syllable (CVVC, CVCC, CVVCC). In the following discussion, I will show how different varieties of Arabic react to syllables of different weight, which in turn results in different algorithms for stress assignment. The algorithm (Angoujard, 1990; Versteegh, 1997) for stress assignment in Modern Standard Arabic is given in (1).

(1) A. Stress the final syllable if it is superheavy (CVVC, CVCC, or CVVCC):

[ki.'taːb]	[CV.'VVC]	"a book"
[mus.ta.'ʕidd]	[CVC.CV.'CVCC]	"to be ready"
[mu.'dˤaːdd]	[CV.#CVVCC]	"counter"

B. Otherwise, stress the penultimate syllable if it is heavy (CVV or CVC)

['kaː.tib]	['CVV.CVC]	"writer"
['yak.tub]	['CVC.CVC]	"he writes"

C. Otherwise, stress the antepenultimate syllable

['ka.ta.ba]	['CV.CV.CV]	"he wrote"

However, this is not the only algorithm for stress assignment in Arabic. Interestingly, different dialects of Arabic could have different algorithms for stress assignment, which will result in different stress patterns. The algorithm given in (1) has been pointed out to be identical to the algorithms used for stress assignment for Palestinian (Kenstowicz, 1983; Younes, 1995) and Damascene Arabic (Cowell, 1964; McCarthy, 1979). It seems to be the case that Levantine Arabic, at least Palestinian and Damascene dialects, have a rather homogenous stress system. However, the algorithm for stress assignment in Egyptian is rather different from the MSA and Levantine ones. Following previous analyses (Hayes, 1995; McCarthy & Prince, 1990; Mitchell, 1975) of Egyptian (Cairene) stress patterns, Watson (2007) revised and proposed the algorithm in (2).

(2) A. Stress the final syllable if it is superheavy (CVVC or CVCC, CVVCC) or CVV;

[ki.'taːb]	[CV.'CVVC]	“a book”
[ka.'tabt]	[CV.'CVCC]	“I wrote”
[ta.'maːss]	[CV.'CVVCC]	“mutual contact”
**[ra.'muː]	[CV.'CVV]	“they threw”

**cf., ['ra.muː] in Levantine/MSA

B. Otherwise, stress the penultimate syllable if it is heavy (CVV or CVC)

['mak.tab]	['CVC.CVC]	”office”
['kaː.tib]	['CVV.CVC]	“writer”

C. Otherwise, stress the antepenultimate or penultimate syllable, whichever is separated by an even number of syllables from the closest preceding heavy syllable, or – if there is no such syllable – from the beginning of the word.

**[mu.dar.'ri.sa]   [CV.CVC.'CV.CV]   "a female teacher"

**c.f., [mu.'dar.ri.sa] in Levantine/MSA

[ka.ta.'ba.tu]   [CV.CV.'CV.CV]   "she wrote it"

c.f., [ka.'ta.ba.tu] "he wrote it" in Levantine/MSA

Arabic also has geminate consonants; i.e., consonants that are doubled. Phonologically speaking, geminate consonants are treated as two identical and consecutive consonants the first of which is located in the coda position of the preceding syllable and the second in the onset position of the following syllable. For example, the word ['dar:ra.sa] "to teach" has a geminate [r] where the first [r] is the coda of [dar] and the second [r] is the onset of the second syllable [ra]. This syllabification helps apply the stress assignment algorithm to words with geminate consonant and to be able to obtain accurate stress assignment as well. It is worth noting that gemination is also possible for coronal consonants in word-final position, as in [mu.'dˤaːdd] "counter" and [ta.'maːss] "mutual contact" from (1) and (2) respectively.

Arabic has also been reported to have a secondary stress (Mitchell, 1960). Words such as [madʒ.ˌdʒaː.'ni:j] would allow secondary stress on the penultimate syllable. It remains unclear whether secondary stress exists across all varieties of Arabic. For example, Watson (2007: 121) concluded that the Egyptian Arabic lacks secondary stress in the system, which is in contrast with San’ani Arabic. To control for such disparity between varieties of dialects, this study will consider primary stress only, as secondary stress is rather unstable and difficult to control.

Furthermore, morphology seems to cause stress to shift in some stress patterns. One of such morphological affixation is the suffix for plurality in Arabic. As given in (3), when these words are in singular form, stress falls either on penultimate or antepenultimate positions whereas as opposed to plural forms, stress falls on the final position. This is due to the plural suffix [u:n] that will make the final syllable superheavy [CVVC]. It can be concluded that non-first-person plural verbs and some plural nouns always have stress in word-final position, regardless of the stress position of the singular form, the number of syllable in the word and other syllabic contexts. Although this behavior completely conforms to the algorithm laid out in (1) and (2), it also suggests that plurality could be a strong cue for stress assignment, which in turn might potentially obscure other cues (e.g., vowel length, position, and syllable type) for stress placement. Therefore, to delineate a more controlled environment for research, the study will limit the discussion to the lexical stress of words that are in more controlled morphological contexts rather than ones which are strongly conditioned by the morphological affixations. That is to say, I will not consider words in plural form, but focus on words that have minimal affixation without resulting in stress shift. This is, however, by no means to deny the role of morphology in the stress system.

(3)

['yak.tub]	['CVC.CVC]	“he writes”
[yak.tu.'bu:n]	[CVC.CV.'CVVC]	“they write”

['tas.ta.miʕ]	['CVC.CV.CVC]	"you listen"
[tas.ta.mi.'ʕuun]	[CVC.CV.CV.'CVVC]	"you all listen"

[mu.'dar.ris]	[CV.'CVC.CVC]	"a male teacher"
[mu.dar.ri.'suun]	[CV.CVC.CV.'CVVC]	"male teachers"

2.1.2 The Phonetics o Lexical Stress in Arabic

Phonetically speaking, the realization of stress in Arabic could be roughly described as having overall higher pitch levels, longer duration, and greater loudness than unstressed syllables (e.g., Al-Ani, 1992). To better describe the correlate of stress in Arabic, Almbark (2014) provided a comparison between the production of stressed and unstressed syllables by speakers of Jordanian and Cairene Arabic, and a comparison between stress production by speakers of Arabic and British English. It was reported that there is a significant effect of stress on the duration, intensity and f0 of the vowel in both Cairene and Jordanian Arabic. In a comparison between stress in Arabic and English, speakers of Arabic used F0 to mark stress whereas speakers of English used F0 as the cue for phrase level accent. Additionally, vowel reduction seems to be used as a strong cue for absence of stress in English.

It was however suggested by Almbark (2014) that there might be cross-dialectal variation in the phonetic correlates of stress in Arabic (Zuraiq, 2005 for Ammani Arabic; Bouchhioua, 2008 for Tunisian Arabic). That is to say, the duration, intensity and f0 of the stressed syllable might be contingent on the type of dialect that the speaker speaks. This variability implies that it would be far-fetched to assume that there exists a set of phonetic correlates for stress in Arabic that is shared by all varieties of Arabic. Besides, it has been suggested (de Jong & Zawaydeh, 2002) that some of the studies (e.g., Chahal, 2001) might have confounded accent and stress in their research design, which makes the understanding about the phonetic properties of Arabic stress less concrete.

2.2 Studies in L2 Stress Perception and Production

As mentioned in the beginning of the chapter, previous studies in second language stress perception and production have been centered around the effect of the native language of the learners. This entails determining the extent to which speakers' native language could promote or impede their performance, generally referred to as L1 transfer, in both stress perception and production in terms of accuracy or realization acoustic cues. The target language of these second language stress perception and production studies has predominately been English over the past decades due to the demand and the general interest from researchers in the field of English as a Second Language. Recently, although the number of studies that focuses on languages other than English as the target language has gradually increased. English remains by a large margin the bare bones for second language stress perception and production studies. These studies where English is the target language involves an increasingly larger body of learners from different language backgrounds including: Arabic (Almbark, Bouchhioua, & Hellmuth, 2014; Anani, 1989; Bouchhioua, 2008; Guma, 2003; Sheikh, 1987), Japanese (Archibald, 1997; Ueyama, 2000), Mandarin Chinese (Chen, Robb, Gilbert, & Lerman, 2001; Lai, 2008; Wang, 2008a; Wang & Yoon, 2008; Yu & Andruski, 2011), Thai (Jangjamras, 2011) and Vietnamese (Nguyen, 2004).

Strictly speaking, these stress perception and production studies do not always examine stress perception and production altogether, as many of the studies only examined either the perception of stress or the production of stress. Different studies might oftentimes have different foci in terms of the factors in questions as well. For the clarity of presentation, the findings of perception studies and production studies will be addressed in different sections.

2.2.1 L2 Stress Perception Studies

There are three main strands of studies in the second language (henceforth L2) stress perception literature, focusing on different aspects of L1 transfer. The first strand of studies is more language-specific, and provides direct comparison between the attention of learners of a specific native language and the native speakers of the target language to the various acoustic cues for stress. Specifically, these studies examined whether learners could attend to the same acoustic correlates of stress (i.e., F0, duration and intensity) as the native speakers, or if they would attend to the relevant correlates active in their native language to perceive stress in the target language. Studies of this type normally focus on learners of a specific native language in the study design, instead of on learners of many different language backgrounds; a control group consisting of native speakers was almost always included in these studies. Additionally, learners of different levels of proficiency were more often included in this strand of studies than in studies which looked at applications of prosodic structure and in studies which focused on the effects of L1 typology, both of which will be explored later in this section.

For instance, Wang (2008) examined the preference for correlates of stress by two groups of Chinese L1 learners of English (ones majoring in English vs. ones that were not majoring in English) and a group of native speakers of English as a control. In a stress identification test and an oddity test, the study found that the learners and the native speakers utilized different correlates of stress to perceive stress. Specifically, Chinese-speaking learners of English relied highly on F0 to perceive stress while making little use of duration and intensity as a cue for stress. In other words, when F0 of the stimuli was manipulated to be absent, the Chinese L1 learners had difficulty perceiving stress in the stimuli. In contrast, the native speakers of English attended to the other two acoustic cues (i.e., duration and intensity) more than the learners, and accordingly were more capable of perceiving stress when the F0 cue was absent.

Lai and colleagues (2008) reported similar findings from two groups (beginning vs. advanced) of Chinese L1 learners of English. Adding to the previous findings, Lai and colleagues found that the completeness of the vowel influences learners' perception of stress as well, in that stress is better perceived when the stimuli have full vowels as opposed to half vowels. Additionally, beginning learners of English attended more to the duration of the vowel in order to perceive stress, whereas advanced learners of English relied more on F0 to detect stress. Compared to their non-native counterparts, the native speakers were found to use both correlates (i.e., duration and F0) in perceiving stress.

Similar findings were reported for learners with various L1 backgrounds. Jangjamras (2011) examined three groups of Thai-speaking learners of English (beginning, immediate, advanced) along with the control group in a stress identification task where the correlates of stress were manipulated for the stimuli. The experimental results indicated that the learners and the native speakers had similar preferences for the position of stress, in that both prefer stress to fall in word-initial position as measured by higher accuracy and faster reaction time. Although the participants did not attend to one particular correlate of stress for stimuli with word-initial stress, the learners were found to use all three correlates of stress (i.e. F0, duration and intensity) for stimuli with word-final stress, while the native speakers only utilized duration as a cue for stress. Learners were also reported to be less attentive to vowel reduction (i.e. a well-observed phenomenon for English stress) as a cue for stress than the native speakers.

Reporting on Vietnamese learners of English, Nguyen (2004) indicated that learners transferred prosodic properties from their native language into their second language perception of stress. The beginning learners relied heavily on two acoustic correlates (i.e. F0 and intensity) that are active in their native language to help them perceive stress and failed to utilize duration as a cue for stress in English. As the learners became more advanced in the target language, they were found to increasingly utilize duration as a cue for stress.

In the same vein, Alfano and colleagues (2007) examined three groups (no exposure vs. some light exposure vs. beginning learners) of Spanish L1 learners of Italian along with a control group in a stress identification experiment where F0 and duration of the stimuli were manipulated. The motivation for choosing Spanish-speaking learners of Italian is the phonological affinity between the two languages, especially prosody (albeit with slight difference). Alfano, Llisterri and Savy reported that the Spanish-speaking learners attended more to F0 and less to duration as a cue for stress as opposed to the control group. Due to inconsistent patterns of perception found amongst participants of the same level of proficiency, it was concluded that variance might not solely be informed by the native language of the learners, but rather by the acoustic properties of the stimuli itself. In a later study conducted by Alfano, Schwab, Savy, and Llisterri (2010), Italian and French L1 learners of Spanish with varying proficiency levels (no exposure vs. some light exposure vs. beginning learners for Italian learners; advanced French L1 learners of Spanish vs. participants with no exposure to Spanish) were examined in a stress identification experiment where the stimuli were manipulated in F0 and duration. The results showed that Italian-speaking learners had better accuracy in perceiving stress in penultimate and antepenultimate positions while were less accurate in word-final positions. Unlike the Italian-speaking learners, the French learners were better on word-final positions while doing worse for other stress positions, reflecting the preference for word-final position in French. The manipulation of F0 and duration did not individually inform the learners' perception; however, when the two cues were manipulated together, the performance of the participants seemed to be largely influenced, resulting in less accurate perception. The researchers concluded that a systematic transfer was not observed in the perception of the participants and that, therefore, the native language of the learners alone is not sufficient to account for the variance found in the experiment, calling for further investigation into other acoustic and psychological factors on second language perception as well.

The second strand of stress perception studies focuses more on other aspects of L1 transfer, especially on the transfer of phonological parameters. This strand of studies examined whether the learners would apply the preferred prosodic structure (e.g., syllabic structure, position and weight, etc.) of their native language to the target language, resulting in a different perception from the native speakers.

For instance, Archibald (1993) examined Hungarian- and Polish-speaking learners of English, where the learners all have invariable stress in their native languages. The study found that these learners had different preferences in terms of the position of stress. It was shown that both Hungarian and Polish learners of English tend to perceive stress to be at the beginning of the word. It was reported that the word-initial position seemed to be slightly more preferred by the Hungarian speakers than the Polish ones, even though a statistical significance was not reached. In a later study, Archibald (1997) reported a longitudinal observation of stress acquisition by three Chinese-speaking learners and one Japanese-speaking learner of English. The learners completed two tests (i.e., a pre-test and a post-test) over an approximately five-month time frame. In each test, the learners listened to some real words that belonged to seven different classes of properties for stress, and identified the stressed syllable by circling the script of the stimulus which was provided on a piece of paper. The results did not find a significant difference in learners' perception of stress between the two testing sessions. Additionally, neither the grammatical category nor the metrical structure of the stimuli (e.g., syllabic structure, syllable weight) seemed to influence the Chinese-speaking and the Japanese-speaking learners' perception of stress, as they seem to treat stress in English as a lexical phenomenon. In other words, these learners seem to treat stress in English on a word-by-word basis. Stress is associated with the lexical item, but not with the general constraints (e.g. syllabic structure, weight of the syllable) for stress in the language.

Face (2005) examined learners' acquisition of syllable weight as a predictor for stress in Spanish. Although stress in Spanish is generally categorized as an unpredictable property (see Altmann 2006), recent studies (e.g. Face, 2000) have found that native speakers of Spanish take syllable weight into consideration when perceiving the stressed syllable in a word. Face expanded on this finding and investigated whether learners of Spanish would also use syllable weight as a cue for stress perception as they advanced in proficiency. In a stress identification experiment, three groups of English L1 learners of Spanish with different proficiency levels were presented with bisyllabic and trisyllabic nonsense words synthesized with neutral correlates of stress; that is, the correlates of stress for each syllable of the stimuli were the same. These syllables, however, varied in weight, as some of the syllables were heavy syllables (e.g., of CVVC) and some were light (e.g., of CV). By exposing the participants to stress-neutral stimuli with varying syllable weight, Face examined whether participants were more likely to perceive stress on heavy syllables than light ones, as demonstrated by the participants in the previous experiment with the same experiment conditions (Face 2000). The experimental results indicated that learners of Spanish took syllable weight into account when perceiving stress in the target language, and they did so more often when they became more proficient in the target language. However, the researcher also suggested that a competing factor for the learners' perception of stress is the default preference for the penultimate position, as the English L1 learners of Spanish seem to correctly think that the default position for Spanish is on the penultimate, regardless of syllable weight. Additionally, the English L1 learners also showed a preference for the antepenultimate position in their perception, which, as suggested by Face, might be due to the preferred position of stress in the learners' native language, as English seems to prefer stress to occur early in the word.

Yu and Andruski (2011) investigated learners' knowledge of stress typicality in English. As shown in previous studies, English bisyllabic words show a strong preference towards a trochaic stress pattern when the word is a noun whereas an iambic stress pattern is preferred when the word is a verb. To examine whether learners acquire the knowledge for the preferred prosodic structure for bisyllabic words in English, a group of Chinese-speaking learners of English and a control group of native speakers of English participated in two grammatical decision tasks. In the first task, the participants were presented with bisyllabic English words embedded in two grammatical frames ("the ____" and "to _____") to elicit participants' preferred stress within respective frames. In the second task, the participants were presented with bisyllabic English nouns and verbs with typical (i.e., trochaic for nouns and iambic for verbs) and atypical (i.e., iambic for nouns and trochaic for verbs) stress patterns to elicit their judgement on the whether the stimuli were grammatical. The results indicated that both learners and native speakers of English demonstrated knowledge for stress typicality in English as they were more accurate and fluent in determining bisyllabic stress patterns that were grammatical given the grammatical frames (i.e., trochaic for "the ___" frame and iambic for "to ___" frame), while they demonstrated different preference for stress patterns when given the ungrammatical contexts. English speakers seemed to prefer trochaic structure for the "to ___" frame while learners of English seemed to prefer iambic stress pattern for the "the ___" frame. Yu and Andruski further suggested that other factors such as word frequency and learners' level of proficiency of the target language might have contributed to the learners' perception as well.

Garcia (2016) examined English-speaking learners of Portuguese and whether they could acquire the parameter of extrametricity in the target language. English and Portuguese both are weight-sensitive languages, but the two languages can exhibit different stress patterns for identical syllabic structures. The reason for different realizations of stress patterns is the fact that the final syllable of a word is treated differently in the two languages. Specifically, the final syllable is skipped (hence, extrametrical) for stress assignment in English while the final syllable is not skipped and is considered for stress in Portuguese. Another reason for different stress realization between the two language lies in the position of the word that is sensitive to weight effect. In Portuguese, heavy syllables (e.g. CVC) attract stress in the word-final position whereas in English it is on the penultimate position. Garcia aimed to capture the extent to which learners of Portuguese acquired the parameter of extrametricity and its interaction with word-internal and word-final codas in the target language in two judgement tasks. In the first task, learners of Portuguese along with a group of native speakers of Portuguese were presented with pairs of identical nonsense words with stress marked on either final or penultimate position in orthography. Based on the orthography, the participants were required to determine using a seven-point Likert scales whether it was more natural for the stress to fall on the final or penultimate position of the word. In the second task, the participants were presented with pairs of nonsense words with stress in identical position but with different weight (i.e., CVC vs. CV,) and coda conditions (such as /l/, /m/, /n/, /r/, and /s/). Similar to the first task, the participants judged the naturalness of the stimuli on a seven-point Likert scale as to whether it was more natural to have one or the other of the stimuli pair in Portuguese. The results indicated that learners of Portuguese and native speakers demonstrated a similar preference for the stress pattern, and the similarity increased as the proficiency of the learners increased, which in turn suggested that the learners re-set the extrametricity setting in their native language and adjust for the extrametricity setting in the target language. A similar pattern was observed for the effect of coda conditions, as learners with a higher level of proficiency and native speakers demonstrated a similar preference for the types of coda conditions; for example, both group of participants seem to treat stimuli word-final CVC with coda /s/ as a light syllable, which in turn should be unstressed.

The third strand of L2 stress perception studies expand from previous L1 transfer studies, and pursue a more holistic and typological account of L2 stress perception. Since previous direct comparison studies have found differences in performance between learners of different native languages, researchers began to examine the association between the difference in performance and the typological differences in the native languages of the learners. Following generative grammar, the various stress systems attested for human languages could be understood as the results of different settings for stress parameters. To allow for a more thorough examination of various stress parameters and how the settings of these parameters could inform learners' perception of stress, this type of study involves learners of a wider range of native languages than the L1 transfer studies mentioned in the earlier paragraphs of this section. However, studies of this kind often lack a stratified sample of participants (i.e. learners of different levels of proficiency).

Altmann (2006) is a good example for establishing the typology of second language stress perception. Following her previous investigation (Altmann and Vogel, 2002), Altmann recruited seven groups of advanced learners of English with distinct native languages. From these seven groups of learners, Altmann examined the extent to which the varying success of stress perception by learners of different native languages could be captured through two hierarchical models for the typology of stress: the Stress Deafness Model (Peperkamp & Dupoux 2002) and the Stress Typology Model (Altmann 2006), which will be further discussed in Section 3.2.2.1. Through a stress identification experiment, Altmann reported that learners of different native languages perceived stress differently: learners with predictable stress in their native language (e.g., French) had more difficulty perceiving stress than learners without word-level stress in their native language (e.g., Chinese) or learners with unpredictable stress in their native language (e.g., Spanish). The results better support the Stress Typology Model as the learners' performance in stress perception seemed to be informed by the setting of the top node of the model (the stress/non-stress language divide). Specifically, the positive setting (i.e. that the learners' native language is a stress language) of the top node seems to hinder their rate of success to perceive stress.

Another example of the typology of stress perception studies would be the series of studies conducted by Peperkamp and Dupoux on what is referred as "stress deafness". In this series of studies (Dupoux, Pallier, Sebastian, & Mehler, 1997; Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008; Peperkamp & Dupoux, 2002; Peperkamp, Dupoux, & Sebastián-Gallés, 1999; Peperkamp, Vendelin, & Dupoux, 2010), Peperkamp and Dupoux examined the extent to which speakers of different native languages (Spanish, French, Spanish-French bilingual, Finnish, Hungarian and Polish) could identify stress in a word. Strictly speaking, some of the stress deafness studies (e.g., (Dupoux et al., 2001; Peperkamp et al., 1999)) are not second language studies, as the participants were not learners of a specific language, and the stimuli presented in the experiment were nonsense words with fairly unmarked syllabic structures where the stressed syllable of the words were manipulated in contrastive positions (e.g., 'piki vs. pi'ki) rather than real words of a specific language. The general purpose of the experiment was to examine the perceptibility of stress by speakers of different native languages, and whether this perceptibility varied along with the native language of the participants. One of the most significant finding from these studies was the stress deafness of the French speakers, a language with highly predictable (final syllable) stress. Precisely, French speakers were reported to have more difficulties in perceiving stress, especially when the acoustic correlates were not kept constant or when the stimuli were produced by different speakers. Difficulties were also found among participants whose native languages have predictable stress: Finnish and Hungarian speakers had more difficulties, and less so for Polish speakers. The varying perceptibility of stress among speakers of different native language led the researcher to argue for a stress typology model, i.e. the Stress Deafness Model, suggesting a negative relation between the predictability of stress in a specific language and the aptitude of its speakers for perceiving stress. In other words, if the speaker speaks a language where stress is more predictable, it is more difficult for the speaker to discriminate stress between words that form a minimal pair with contrasting stress position.

2.2.2 L2 Stress Production Studies

Although there is an overlap between the stress perception and the stress production studies, the stress production studies have a rather strong focus on the direct comparison between the production by L2 learners and native speakers. These studies focus on direct comparison between the acoustic properties of stress production by L2 learners of a specific language and by native speakers of that language. Second language stress production studies normally involve learners with a specific language background and varying levels of proficiency, rather than learners of different language backgrounds with a similar level of proficiency.

For instance, Chen, Robb, Gilbert and Lerman (2001) investigated the L2 production of stress by Chinese L1 learners of English and English L1 speakers as controls. The results showed that while both groups of participants demonstrated their ability to produce stressed syllables with relatively higher values for all correlates (i.e., F0, intensity and duration) of stress, their production of unstressed syllables differed. Compared to the English L1 speakers, the Chinese L1 learners produced unstressed syllables with significantly higher F0 and intensity, which was argued to be due to an interference from the participants' L1. This interference is ascribed to the generally higher fluctuation of F0 in Mandarin Chinese, as studies (e.g., Shen, 1990) have shown that Chinese seems to have a higher degree of fluctuation of F0 on the syllable level, and when combined with sentential intonation, an even higher F0 is resulted.

In the same vein, the production of L2 English stress by speakers of different language backgrounds were reported as well. Some of these production studies overlapped with the perception studies discussed above. For instance, Jangjamras (2011) reported that Thai L1 speakers and English L1 speakers had both similarities and differences in realizing stress. When a word had word-initial stress, both Thai L1 and English L1 speakers tended to have higher F0 and higher intensity on the stressed syllable. On the other hand, when a word had word-final stress, higher F0 and intensity were found less often, and a longer duration was found on the stressed syllable. Interestingly, it was found that the judges who evaluated the production of stress by the two groups indicated that they used duration and intensity to determine stressed syllables produced by the Thai L1 speakers whereas they used F0 for the English L1 speakers. This finding suggested that Thai speakers had relatively high F0 for both stressed and unstressed syllables, and therefore the judges needed to rely on the other two correlates, i.e. intensity and duration, to determine whether a syllable was stressed. On the other hand, the judges relied on F0 for stress produced by English L1 speakers, suggesting that duration and intensity do not seem to be essential indicators for stress. Reporting on Vietnamese L1 learners of English, Nguyen (2004) showed that advanced learners were able to produce stress utilizing all three correlates of stress as the English L1 speakers, indicating that accurate stress realization was attainable by Vietnamese speakers. However, the beginning learners did not seem to fully utilize duration to contrast stressed and unstressed syllables and did not show any signs of vowel reduction in the stressed syllable, which was characterized by the production of stress by English L1 speakers. The lack of vowel reduction when producing the stressed syllable was reported for other learners as well, as Arabic L1 learners of English did not seem to have vowel reduction in their production of stress in English (Almbark et al., 2014). Reporting on Arabic L1 learners of English as well, Anani (1989) found that the Arabic L1 speakers seemed be heavily influenced by the syllabic structure of the word. More interestingly, they seemed to apply the algorithm for stress in Arabic to determine the position of stress in English words; that is, when given an English word with a syllabic structure that is also available in Arabic, the Arabic L1 learners stressed the word as they would in Arabic, which is argued to be a strong indication of the effect of L1 transfer.

As a typological study for L2 stress perception and production, Altmann (2006) also investigated the production of L2 English stress by learners from seven different language backgrounds. In a production experiment using English nonsense words, the results showed that learners who speak a non-stress language (e.g., Korean, Japanese and Chinese) displayed the least native-like performance and tended to place stress on the last syllable. Speakers of Spanish, like English, a language that has unpredictable stress, did not perform as well as expected; aspects of their production were similar to the production by the native speakers, while other aspects indicated a preference for stress on the last syllable of the word. Unexpectedly, speakers of languages that had predictable stress, such as Arabic, French and Turkish, were found to have the most target-like production of stress. In particular, French learners were able to correctly place stress in non-final position in words that would normally be stressed word-finally in French. Altmann took these results of the experiment to support a typology model that predicts the aptitude of speakers of a particular language for stress production based on the stress parameters of that language such as predictability, existence of stress, and weight-sensitivity of stress.

2.2.3 L1 And L2 Stress and Perception And Production and Arabic

Stress perception and production seems to be a topic which has received relatively less attention in the Arabic literature, as most studies concerning stress in Arabic have been L1 stress production studies, which focus on providing descriptions on the acoustic properties of stress in Arabic (Almbark et al., 2014; Anani, 1989; de Jong & Zawaydeh, 2002, 1999; Heliel, 1982). Other than a study that examined whether the misplacement of stress could influence speakers' ability to recognize a word in Arabic (Boudelaa & Meftah, 1996), no study, as far as I know, has investigated L1 perception of stress with Arabic being the target language. This lack of investigation of L1 stress perception and production extends to L2 stress perception and production where Arabic is the target language as well. As far as I know, published work on L2 Arabic stress perception and production is extremely limited, bordering on non-existence.

The closest link that can be found between Arabic and L2 stress perception and production is from the participation of Arabic L1 speakers in ESL studies. Although many studies (Almbark et al., 2014; Anani, 1989; Bouchhioua, 2008) have recruited Arabic L1 speakers, the target language of these studies has been English, a language that has an unpredictable stress system, unlike Arabic. Thus, the findings yielded from these English L2 stress studies, albeit produced by L1 speakers of Arabic, are less relevant to Arabic second language stress production.

In addition to the limited number of studies of L1 Arabic stress perception and production studies, lexical stress in Arabic does not seem to be adequately addressed and somehow absent in the acquisition of Arabic as a L1. It has been noted (Ryding, 2013) that although native speakers of Arabic can accurately model stress for any given Arabic word, they may not know how to explain why a certain syllable should be stressed. Anecdotally, based on the author’s previous interaction with the native speakers, some native speakers did express that stress is something that they do not think in Arabic , and something that is not very important. Combining these two observations, the general lack of awareness of stress and the impression of lexical stress as a somehow marginal linguistic feature seems to lead to the lack of interest in researching L1 stress perception and production in Arabic in particular and seems to be generally true for other languages as well.

2.2.4 Conclusions from L2 Stress Perception And Production Studies

Previous L2 stress perception and production studies have provided an extensive examination on the effect of L1 transfer, mostly on English as a second language. To examine how the prosodic systems of individual languages interact with the stress system in English, these English L2 studies recruited learners with various L1 backgrounds and experimentally tested their perception and production of L2 English stress. If the effect of L1 transfer was evident, the learners would be expected to carry over some of the prosodic features exhibited in their native language to their perception and production of stress.

However, these L2 stress perception and production studies did not always provide congruent findings regarding the effect of L1 transfer and the extent to which it is exhibited in the learners' perception and production of stress. The findings can be summed in three mains categories, as some studies found that learners applied the stress pattern active in their native language to that of the target language, while some found that the learners perceived and produced stress in a way that was very similar to that of native speakers, and still others found that the learners' perception and production of stress did not entirely belong to either their native language or the target language (e.g., Archibald, 1997).

The lack of a consistent clustering of findings might be due to different designs of the experiments implemented in these studies. As rightly pointed out by Altmann (2006), not all studies investigated perception and production together; instead, the majority of the L2 stress studies focus either on perception or on production, making it difficult to relate the findings from one study to another and from perception to production as well. Moreover, the target group of interest varies from study to study, again making it more difficult to relate the results. Consequently, the varying language backgrounds of the target group require specific choices of stimuli and accordingly influence the target structures examined in each individual study. Finally, the most evident drawback of some of the studies, which could have led to the incongruous findings, is the inclusion of real words as stimuli. The potential challenge for including only real words as stimuli is that participants might be more familiar with some of the real words than some other participants. Hence, any difference in performance observed in the experiment might simply result from the participants' familiarity with the stimuli, rather than from the other factors (i.e. the participants' L1) in question.

The incongruous findings from the L2 stress perception and production studies could also be due to the failure to incorporate of other relevant factors, especially those relevant to language processing, such as effects from recency, redundancy and input frequency. Among these three factors, the effect of input frequency (henceforth frequency effect) has been increasingly examined in the field of psycholinguistics for both L1 and L2 acquisition but none of the studies discussed above has incorporated input frequency into the study design. However, a few studies (e.g., Altmann, 2006; John Archibald, 1997) did note the potential influence from the frequency of stress patterns in the target language, which in turn motivates incorporating frequency as a variable in L2 stress perception and production investigation.

2.3 Studies on Frequency Effect in Language Processing and Learning

2.3.1 Frequency Effect on L1 Learning And Processing

The frequency effect in language learning and processing suggests a certain level of correlation (be it positive or negative) between the frequency of a linguistic element in the input and the accuracy and fluency of the speaker performing tasks associated with the linguistic element. Frequency effect has received increasingly attention in the fields of psychology and psycholinguistics, as the effect is ubiquitous for nearly all kinds of "linguistic elements". In a review of the frequency effect in language processing, Ellis (2002) summarized aspects of language learning and processing that are sensitive to frequency effect, including phonology (Treiman & Danis, 1988), phonotactics (Frisch, Large, Zawaydeh, Pisoni, & others, 2001), syntax (Martin, Church, & Patil, 1987) , morphosyntax (Brooks, Braine, Catalano, Brody, & Sudhalter, 1993), reading and spelling (Coltheart, Curtis, Atkins, & Haller, 1993), lexis (Forster & Chambers, 1973), formulaic language (Biber et al., 1999), to name a few.

For instance, although speakers are not explicitly taught the phonotactics of their L1, they all seem to be really successful in judging whether a word is native or not. This is true for both adults and young children who are still acquiring the language (Treiman & Danis, 1988). In a lexical decision experiment by Frisch et al., 2001, participants were asked to judge the well-formedness of nonsense words on a 7-point Likert scale. The phonotactics of stimuli were manipulated based on the statistical information of the phonotactics in English to reflect different levels of well-formedness. The results showed a surprisingly strong correlation between the level of well-formedness and the probability of the phonotactics of the nonsense word. This high correlation suggested that the participants had a frequency distribution of phonotactics that was very similar to the actual distribution of the phonotactics in English. This finding in turn appear to support statistical learning by learners who track relevant frequency information about the language that they are acquiring. In Frisch's study, it was the frequency information about consonantal constituents of the lexicon that the learners were tracking.

Studies have also showed that infants, who are at the very first stage of acquiring a language, also utilize the statistical information to learn the regularity of the language. As the very first attempt to investigate word segmentation by infants, Saffran, Aslin and Newport (1996) exposed twenty-four 8-month old infants to a two-minute recording that was composed of four unbroken tri-syllabic nonsense words. Acoustic cues were controlled for the experiment and the statistical information, where the frequency of the transitional syllable was the only information available in the stimuli to determine the word boundary. The results showed that, at the age of 8 months, the infants were able to learn and analyze the statistical information that was provided to them in the stimuli, and to utilize such information to determine the word boundary. In a later experiment, Saffran and collegues (1999) adopted a similar design using a different acoustic cue; i.e., the F0 of the word, to signify word boundary. A similar result was yielded that infants and adults were both able to determine word boundary by learning from the implicit statistical information. It was also concluded that infants can utilize a variety of statistical information, such as frequencies of occurrence, frequencies of co-occurrence, conditional probability and relevant statistics (Romberg & Saffran, 2010) to accomplish various tasks that are essential to L1 learning, such as word segmentation (e.g., Goodsitt, Morgan, & Kuhl, 1993), lexical categorization (e.g., Gómez & Gerken, 2000) and form-meaning mapping (e.g.,Smith & Yu, 2008). It has also been shown that infants acquire a great deal of linguistic knowledge, such as vowel space, consonant categorization and phonotactics, through analyzing the frequency distribution of linguistic elements that are available to them (Jusczyk, 1997) .

Word processing is also found to be frequency-sensitive across different modalities. In the written language, words of high frequency have been shown to have the advantage of being named more quickly (Balota & Chumbley, 1984), being recognized to be either a real word or a nonsense word more quickly (Foster, 1976) and being spelled out more correctly (Barry & Seymour, 1988). In the spoken language, words with higher frequency were reported to have strong advantages for both recognition-related tasks, such as perception, naming and reading, and production-related tasks, such as speaking, typing and writing. Such effect was found among children and adults in both L1 and L2.

These studies have highlighted the role of frequency, which influences language processing and learning on several levels. Frequency also shapes the learners' linguistic capacity, and further informs one's performance and preference in both L1 and L2.

2.3.2 A Usage-Based Approach to Second Language Acquisition

Recent studies which have shown a strong link between frequency and language processing and learning have led to the emergence of the "usage-based approach" in language learning. Synonymous with frequency learning or statistical learning, the new approach assumes that learning is bottom-up, statistical, and frequency-informed. In terms of second language acquisition, target structures and idiosyncratic patterns of the target language are claimed to emerge from usage or exposure to input. The second language learner tracks the frequency of the linguistic input and extracts regularities about the language from the abstractions of frequent linguistic elements. As the learner repeats this process of tracking and abstracting characteristics of the language from the frequency distribution of linguistic elements over the course of lifetime or learning, the language emerges as one that has approximated the regularities of the input language.

For example, Ellis and colleagues (2014a) implemented two production experiments to investigate frequency effect on the production of Verb Argument Constructions (VACs) by native speakers of English. In the first experiment, the participants were asked to fill the verb slot in 40 VAC frames, such as "he ____ cross the…" with whatever verb that first comes to mind. In the second experiment, the participants were given a minute to fill in the slot with as many verbs as possible. The participants' responses were merged and correlated with the frequency of the same VACs in the British National Corpus. The results showed that the verb choices for the VACs frames were determined by the frequency, contingency and semantic prototypicality of the verb observed in the VACs. In a later investigation by Ellis and colleagues (2014b), the same tasks were assigned to L2 speakers of English, who were asked to provide the verb that best suits the VAC frame in task and to give as many verbs that fit the frames as possible in the other. The results showed that advanced L2 speakers behave much like native speakers of English, whose verb choices for the VACs frames were also conditioned by frequency, contingency and semantic prototypicality.

2.4 Frequency and Stress Perception and Production

Although the findings of the abovementioned studies seem to suggest that frequency influences several aspects of language processing and learning, there is little discussion of the frequency effect in the perception and production of lexical stress in L1. This observation also holds for L2 studies. That the topic draws little attention in L1 acquisition studies is understandable, since native speakers have rather consistent performance on either perception or production of stress in real words. It is rarely the case that they would mishear the stress or place the stress on the wrong syllable due to low frequency.

However, when presented with nonsense words of varying frequency, the consistent performance exhibited by native speakers for real words can be otherwise different. That is, they could perceive or produce stress less accurately and fluently when dealing with words that are not attested in the language. In a rare study investigating frequency effect on the stress production in Russian, Jouravlev and Lupker (2015) presented a probabilistic Bayesian model for stress assignment in Russian, a language with complex and relatively unpredictable stress.

The model considered the frequency of both lexical and non-lexical parameters and their resulting posterior probability to compute the position that is most likely to be stressed any given word. With this model, Jouravlev and Lupker tested the model's prediction of the stress pattern (to be either iambic or trochaic) for bisyllabic nonsense words in Russian. The probability of the stress pattern predicted was later correlated with the proportion of production in the corresponding stress pattern by native speakers of Russian, and a significant positive correlation was found. Additionally, a negative correlation was found between the probability of the predicted stress pattern and the latency (i.e. reaction time) of production. These findings show that the model was capable of predicting stress from a purely probabilistic and frequency-informed approach, and the accuracy and latency of speakers' production seems to be influenced by the frequency of the parameters as well, which in turn highlights the role of frequency in L1 stress production.

Finally, little research has been conducted on L2 stress while taking frequency effect into account. A notable exception is Post da Silveira (2011) who investigated whether the high frequency of penultimate stress position exhibited in both the participants' L1 Brazilian Portuguese and their L2 English could facilitate or impede their production of stress in English. A stress production task was conducted using stimuli consisting of bisyllabic, trisyllabic, and quadrisyllabic real words, with varying stress positions being in the first, second, or third syllable of the word. The experiment was conducted twice with a six-month interval, employing 16 participants with varying proficiency levels. The results did not reveal any difference in accuracy of stress production between the two periods, suggesting lack of any proficiency effect. However, participants were found to place stress on the penultimate position significantly more often than on any other position in the word. It was also found that learners resorted to the preferred position of stress (penultimate) in their L1 when dealing with words that had stress-bearing suffixes. Post da Silveira (2011) concluded that the preference for penultimate position comes from the effects of frequency from both L1 and L2. The study had a number of methodological limitations, including the use of only real words, lack of control for the frequency of stress patterns, and presence of a confounding variable where the penultimate is preferred in both correct and incorrect production making it, therefore, impossible to tease apart L1 from L2 effects.

The two studies discussed in this section are methodologically informative, as the method implemented in both studies, i.e. combining corpus analysis and experimental testing, seems to be the general norm for the testing of frequency effects on language learning. Specifically, the frequency of stress-related cues can be obtained from the corpus analysis and can inform the construction of stimuli used in the experiment. Accordingly, the empirical data yielded from experimental testing could be further related to the frequency of the stress-related cues implemented in the stimuli.

2.5 Conclusion

The studies reviewed in this chapter provide several insights into the phonological and phonetic properties of stress in Arabic as well as the tenets of frequency effect and how frequency could influence L1 and L2 language learning and processing.

Previous L2 stress perception and production studies have shown mixed results regarding the effect of L1 transfer, and do not consistently determine the extent to which the native language of the learners influences their perception and production of stress. These mixed findings could be due limitations already discussed, which include (1) frequent failure to investigate perception and production of stress in the same experiment, (2) lack of stratified data for a specific group of learners with different level of proficiency, as each individual study mainly focuses on one group of learners of interest, which in turn resulted in (3) the lack of a cohesive stimuli design and targeted structures and parameters of stress in the target language, and lastly but most importantly, (4) biases introduced by the use of real words in the stimuli. Furthermore, the frequency effect, a factor that is widely reported to influence various aspects of language processing and learning, seems to be generally neglected in the L2 stress perception and production literature. Additionally, past attempts to relate frequency to stress perception and production also seem to also highlight the role of frequency.

Learning from the limitations of previous L2 stress perception and production studies, it becomes clear that, in order to provide a more comprehensive examination of L2 stress perception and production, the following must be taken into consideration. Firstly, a study design that includes tasks examining the perception and production of stress must be implemented such that the results from stress perception and stress production are mutually informative. Secondly, the stimuli constructed for the experiment should not contain only real words; in fact, the inclusion of nonsense words should be preferred. Lastly, in designing the stimuli, not only do the segmental features need to be carefully and systematically contrasted and manipulated, but the frequency of the stress pattern as well other related cues also need to be considered and controlled. The present study attempts to take all these factors into account (see Chapter 3).

Chapter 3

Methodology

In Chapter 1, several research questions were proposed. These questions are re-stated below as follows.

RQ 1. Does the native language of the learners influence their performance in stress perception and production, and, if so, how?

RQ 2. Does the proficiency of the learners influence their performance in stress perception and production, and, if so, how?

RQ 3. Does the frequency of relevant structures for stress (i.e., position of the stressed syllable, syllabic structure of the stressed syllable, and stress pattern of the word) influence the participants' performance on stress perception and production, and, if so, how?

RQ 4. Does frequency have a different effect on stress perception as compared to stress production?

RQ 5. Does frequency have a different effect on learners as compared to native speakers?

RQ 6. Do other external factors (specifically explicit knowledge of lexical stress, instruction, self-awareness and extracurricular engagement with the target language) influence learners' performance on stress perception and production, and, if so, how?

Specifically, this study investigates whether and how internal factors (such as native language and proficiency of the learner), input factors (such as length, stress pattern and frequency of the word) and external factors (such as the explicit knowledge of stress, instruction, self-awareness and extracurricular engagement) influence the perception and production of stress by L2 learners of Arabic.

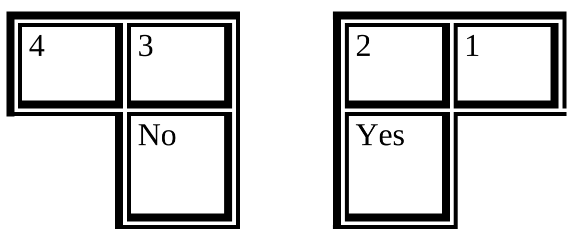
To examine the research questions of the study, a stress perception and production experiment was conducted. The experiment was piloted in late 2015 and conducted in early 2016. This chapter discusses the design and the rationale of the experiment. The chapter is organized as follows: Section 3.1 presents the experimental design and rationale for the tasks. Section 3.2 describes the stimulus design and the rationale for stimulus construction. Section 3.3 discusses the rationale for participants recruitment criteria. Section 3.4 presents the data analysis method, followed by a summary for the chapter in Section 3.5.

3.1. Experiment Design

The objective of the experiment is to evaluate the performance of learners of Arabic in stress perception and production. The common approach to evaluating speech perception and production is to access their fluency and accuracy. Fluency of perception and production can be understood as how quickly learners perceive and produce the target stimuli, whereas accuracy represents the distance between the learners' production of the target structure (i.e., proximity to native speakers' production) or disparity between learners' and native speakers' perception.

To achieve this goal, I conducted an experiment consisting of a warm-up activity, three main tasks that assessed participants' performance in stress perception and production, and a post-experiment questionnaire to investigate possibly relevant external factors that cannot be assessed elsewhere in the experiment. The entire experiment was developed using PsychoPy, a computer program commonly used for developing testing interface for experiments in the field of psychology and linguistics. The same program was used to administer the experiment with a 13-inch Windows laptop computer. The participants used a number pad as the control panel to navigate and complete tasks in the experiment via keypresses on the panel. The number pad was customized as the layout presented in Figure 3.1. The numbers and the yes/no key all start from right to left, which is counter-intuitive to English speakers, but is catered to Arabic orthography which is also written from right to left. All tasks in the experiment consisted of a practice phase and a testing phase. Detailed description of the tasks as well as the rationale behind the tasks are presented in the following sections.

Figure 3.1 Layout of the Control Panel



3.1.1 Warm-Up Activity: Speed Reading

As a warm-up activity, I used a speed reading drill that assessed how fast each individual participant could finish reading aloud a given list of Arabic words. The drill started with the practice phase, where the participants were prompted with the description of the drill on the computer screen. By pressing any key on the control panel, the participants were presented with a list of 16 real Arabic words on the screen with 8 words in a row for a total of two rows and were simultaneously prompted with a beep which was to signify the participants to start reading the words aloud. Once the participants finished reading, they were required to press any key on the control panel again to conclude the practice phase and were prompted to the testing phase of the drill. The testing phase followed the exact procedure as in the practice phase with a different list of real Arabic words. The wordlist used in the testing phases was longer than the one in the practice phase, consisting of a total of 48 real Arabic words with 8 words in a row for a total of 6 rows. The production of the participants in both phases was auditorily recorded, and the time that participants took to complete the drills was automatically calculated. The prompts used in both phases are included in Appendix 3.1. in both English and Chinese (i.e., the L1s of the two test groups of participants), whereas the wordlists used in the drill are presented in Appendix 3.2. The organization of the stimuli is discussed in more detail in Section 3.2.4.2.

Just as in the language classroom, where it is common for the instructor to start with a lighter activity first to put the students into the mode of using the target language, the first objective of the warm-up activity is to ease the participants into the mode of using Arabic and prepare them for the upcoming more intensive tasks. Secondly, a simple activity can provide an overall estimate of whether the participants of the same level of proficiency have comparable performance.

3.1.2 Stress Identification Task

The first main task of the experiment was the stress identification task, which aimed to examine learners' perception of lexical stress in Arabic. The task required the participants to listen to recordings of real and nonsense Arabic words and to identify the position of the stressed syllable in the words.

In the practice phase, the participants were prompted with the description of the task on the computer screen, as presented in Appendix 3.3. By pressing any key on the control panel, the participants were prompted to a "stand-by" page, signaling the participants to press any key to proceed to the task. Once the participants pressed any key, the participants heard an audio recording of a real word produced by a native speaker of Arabic from Saudi Arabia. Along with the recording, the orthography of the word was presented to the participants at the same time on the computer screen. After listening to the word, the participants determined which syllable of the word was stressed by pressing the corresponding number on the first row of the control panel, where the number "1" stood for the first syllable of the word, the number "2" for the second syllable and so on. The participants were asked to do so as quickly as possible. Immediately after the response, the participants were prompted with a lexical decision task which is discussed in more detail in Section 3.1.3. After responding to the lexical decision task, the participants were prompted with a one-second buffer screen on the computer screen, serving as the reminder for the number of the remaining trials and a transition page to the next stimulus. The next stimulus was automatically provided after the buffer screen lapsed.

The testing phase of the tasks follows the exact procedure, only with a different set of stimuli. Unlike the practice phase, the stimuli used in the testing phase could be either real or nonsense Arabic words. Additionally, the testing phase allowed a self-paced break for the participants during the task. A stand-by page appeared when the participants complete a block of stimuli where the participants can resume the tasks whenever they felt prepared. It is worth noting that the first block and the last block of stimuli in the testing phase were also buffers. Each buffer block contained five real words taken from the practice phase and responses to the stimuli in the buffer blocks were excluded from analysis. The purpose of the buffer blocks was similar to that of the warm-up activity at the beginning of the experiment, which was to raise the awareness and their attention to the task, and to avoid fatigue effect on participants' performance towards the end of the tasks.

Because the participants might not know the syllabification of the stimuli solely based on the orthography, the orthography of the word was syllabified and numbered to better inform the participants of the relative position of the syllables in the word. A sample of pre-syllabified and numbered orthography of a stimulus is given in Figure 3.2. Another consideration for the presentation of the stimuli was the limitation of vowel information. As shown in Figure 3.2, vowel information is concealed in the orthography – no short vowel markings are available and the length of the vowel is concealed. Even though the word is originally نتيجة "result" [na.ˈtiː.dʒa], with the presentation in Figure 3.2, the participants will only know that the stimulus has three syllable with [n] being the onset of the first syllable, [t] for the second syllable and [dʒ] for the third.

Figure 3.2 Sample of Stimuli Presentation in the Stress Identification Task

ــجة	ــتــ	نــ
(3)	(2)	(1)

The motivation for limiting vowel information for the participants is that Arabic orthography is rather "revealing" for the pronunciation of a word. That is, unlike the spelling of a word in English which does not always match with its pronunciation, the orthography in Arabic completely reflects the pronunciation of a word once the vowel markings are made available. It is likely that the participants will solely rely on the visual stimuli to extract the pronunciation of the stimulus and deduce the position of the stressed syllable without actually attending to the aural stimuli. Another reason for concealing vowel information is to avoid scenarios where learners impressionistically but falsely, however, prefer a certain type of vowel quality or vowel length in the script of stimuli without attending to the acoustic cues provided by the stimuli. To circumvent these adverse scenarios, full orthographical information and especially vowel information of the stimuli was kept to a minimum.

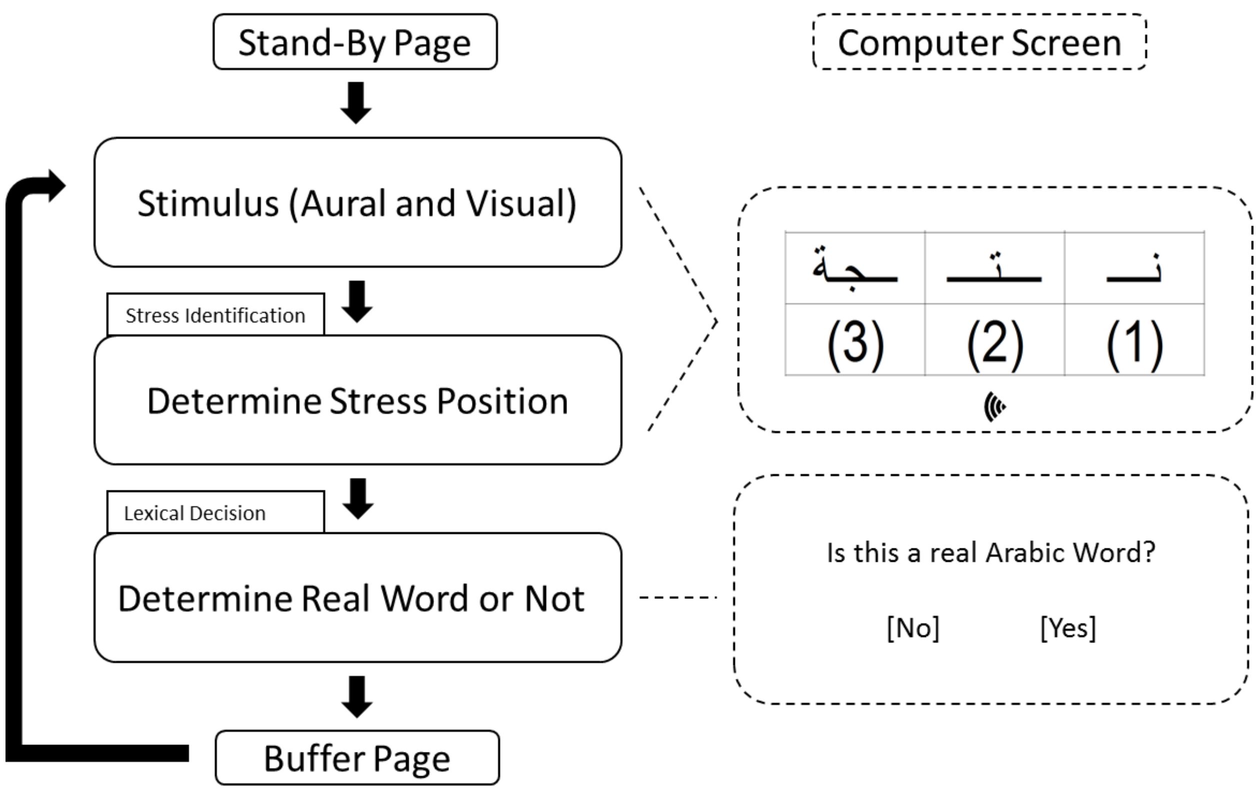
In this task, the fluency and the accuracy of participants' responses to the stimuli were evaluated. The fluency of response in this task was measured as reaction time, or the amount of time that the participants took to respond to the stimuli. As for accuracy of response, it was determined by whether the syllable selected by the participants matches the attested position for the stressed syllable in the language, i.e. one that is determined by the stress algorithm specified previously in Chapter 2. These measures provided different dimensions to assess participants' performance in stress perception: the performance of the participants was deemed to be poorer if it takes them longer time to "figure out" the stress position, even correctly. Similarly, independent of response time, performance was considered to be poorer if the participants did not identify the position of the stressed syllable correctly.

3.1.3. Lexical Decision Task

The second main task was the lexical decision, which targeted participants' lexical knowledge of Arabic lexicon and whether this knowledge could be influenced by the various variables investigated in this study (i.e. participants' native language, their proficiency and frequency of stress patterns). The task required the participants to determine whether the stimuli presented to them in the stress identification task were real Arabic words or not.

As mentioned in the previous section, the lexical decision task was integrated into the stress identification task and the procedure of the task is presented in the diagram in Figure 3.3. For each stimulus in the stress identification task, after determining the stress position of the stimulus, the participants were immediately prompted with a yes/no question on the computer screen, which asked whether the stimulus that they just heard was a real Arabic word or not. The participants were required to respond to the question as soon as possible by pressing on either the "Yes" button or the "No" button on the control panel. After the response, the participants were prompted with the remaining number of the tokens in the task. The same procedure was followed in both the practice and testing phase of the task.

Figure 3.3 Procedure of Stress Identification and Lexical Decision Task



The responses were evaluated for their fluency and accuracy as in the stress identification task. The fluency of response is the participants' reaction time measured as the time that the participants took to determine whether a given stimulus was a real Arabic word or not via keypress. The reaction time was automatically captured and calculated by the computer program as the duration between the timestamp of the keypress and the onset of the lexical decision screen. Response accuracy was the correction of the lexical decision (real or nonsense), which was also captured by the computer program with pre-loaded answer keys.

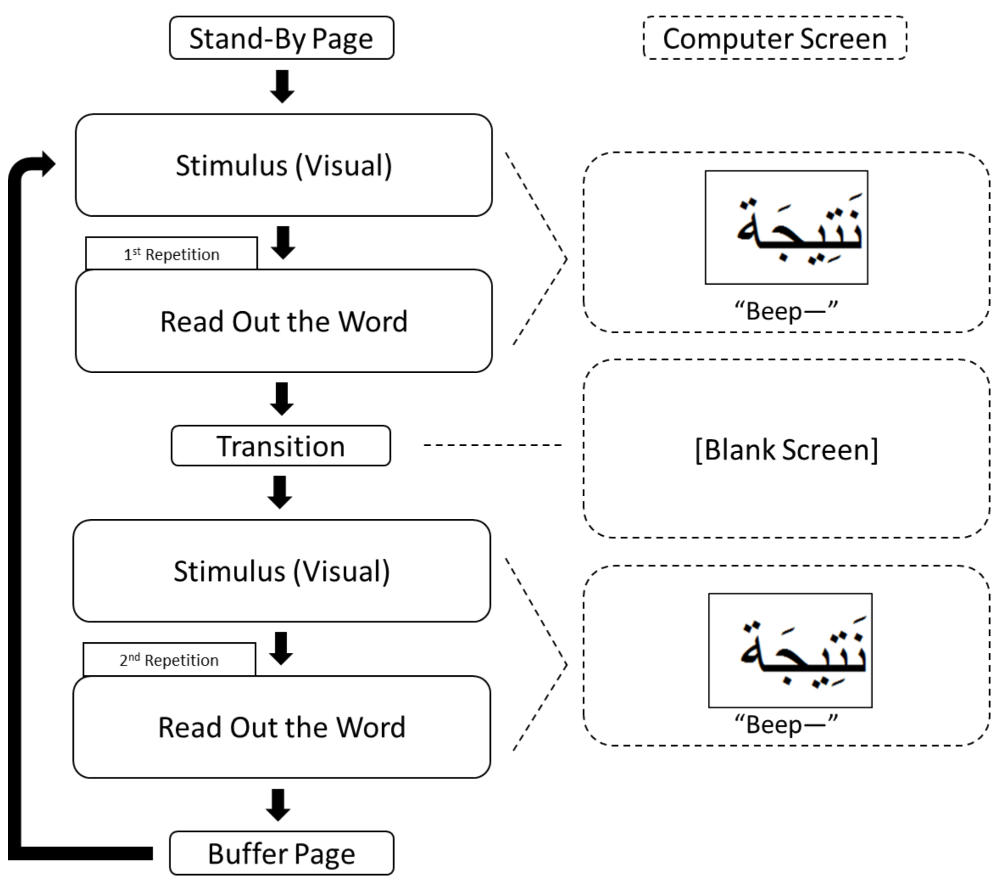
More importantly, the motivation for the lexical decision task was to use it as a filler task to conceal the main purpose of the experiment, which was to examine learners' performance on lexical stress in Arabic. By minimally diverting the participants' attention from the stress identification task with a task that was less relevant to lexical stress, i.e. the lexical decision task, the participants were less likely to notice the linguistic element that the experiment targets. Without such active notice of lexical stress, the participants were more likely to respond to the stimuli more naturally, and were less likely to spend unnecessarily longer time on each stimulus in order to strive for the correct answer.

3.1.4. Stress Production Task

The last of three main tasks was the stress production task, which aims to evaluate participants' production of lexical stress in Arabic. The task required the participants to read aloud each stimulus presented in the task. Unlike the warm-up activity, instead of reading aloud an entire list of words at once, the stress production task proceeded one word at a time and required the participants to produce each stimulus twice in succession.

The procedure is presented in the diagram in Figure 3.4. In the practice phase, the introduction page gives the instruction for navigating the task, as given in Appendix 3.4. With any keypress, the participants were prompted with the stand-by page on the computer screen, indicating that the participants can start the task whenever they are ready. Once the participants pressed any key to proceed to the task, they were prompted with a visual stimulus which was a real Arabic word in Arabic script on the computer screen accompanied by a beep. The beep prompted the participants to read out the script once as soon as possible within a 3-second time frame. Then the script disappeared for 0.5 second and reappeared on the screen with another beep, again prompting the participants to read out the script once within a 3-second frame. After the two repetitions, a buffer page briefly appeared for 0.5 second on the screen that reminded the participants of the remaining number of stimuli in the task and then transitioned to the next stimulus.

Figure 3.4 Procedure of Stress Production Task



The testing phase of the task followed the exact same procedure, but with a different set of stimuli. As in the stress identification task, the stimuli used in the testing phase consisted of both real Arabic words and nonsense words. Similarly, the testing phase allowed for a self-paced break for the participants during the task. A stand-by page appeared when the participants completed a block of stimuli where the participants could resume the tasks whenever they felt prepared. The first block and the last block of stimuli in the testing phase were buffers as well. Each buffer block contained five real words taken from the practice phase and responses to the stimuli in the buffer blocks were excluded from analysis. The purpose of the buffer blocks was, as in the stress identification task, to raise the awareness and their attention to the task, and to reduce possible fatigue towards the end of the task.

Unlike in the stress identification task, vowel information was available to the participants and no pre-syllabification was provided. As an example, Figure 3.5 is the same word نتيجة "result" [na.tiː.dʒa] as presented in Figure 3.2, but in the format for the stress production task. Compared to the presentation for the stress identification task, where the participants only knew that the stimulus is trisyllabic word with three onsets being [n], [t], [j] for the respective syllables, the stimulus now has all vocalic markings, where the participants knew that the vowel for the onset [n] is [a] in the first syllable, a long high vowel [i] for the onset [t] in the second syllable, and again the vowel [a] for the onset [dʒ] in the third syllable.

Figure 3.5 Sample Stimulus for the Stress Production Task

The motivation for providing full vowel information was to offer the participants as much information as possible to reduce the burden of reconstructing the vowels, thereby allowing them to focus more on other aspects of their language production, i.e. lexical stress. More importantly, the full vowel information disambiguated Arabic script and unified the pronunciation for the stimuli. Without any vocalic markings, Arabic scripts can be ambiguous, which in turn might result in varying pronunciations across the participants. For example, the Arabic script ولد is ambiguous and could be understood as different words which could result in more than one pronunciation, such as وَلَدَ [wa.la.da] “a child”, وُلِدَ [wu.li.da] “to be born”, or وَلَّدَ [wal.la.da] “to generate”. Varying pronunciation could also lead to a switch in stress position, such as سَجَّلَ [ˈsadʒ.dʒa.la] “to record” and سِجِلّ [si.ˈdʒill] “record”. Therefore, to control for potential biases from ambiguous orthography and to guarantee that the participants’ pronunciation of the stimuli matches the intended stimuli, all stimuli were fully vocalized.

The productions were recorded, and each repetition was individually evaluated for its fluency and accuracy. As before, fluency was understood as the reaction time of the response, that is, the amount of time that participants took to initiate the production. It was calculated from the timestamp of the onset of the production and the time stamp for the beep that appeared for each repetition. Responses were accurate if the participant placed stress on the target position in a word, which was determined by the stress algorithm. Unlike the two previous tasks in the experiment, since the computer program was not able to automatically evaluate the acoustic data produced by the participants, the data analysis of the production was conducted separately in another computer program, which is discussed in detail in Section 3.4.

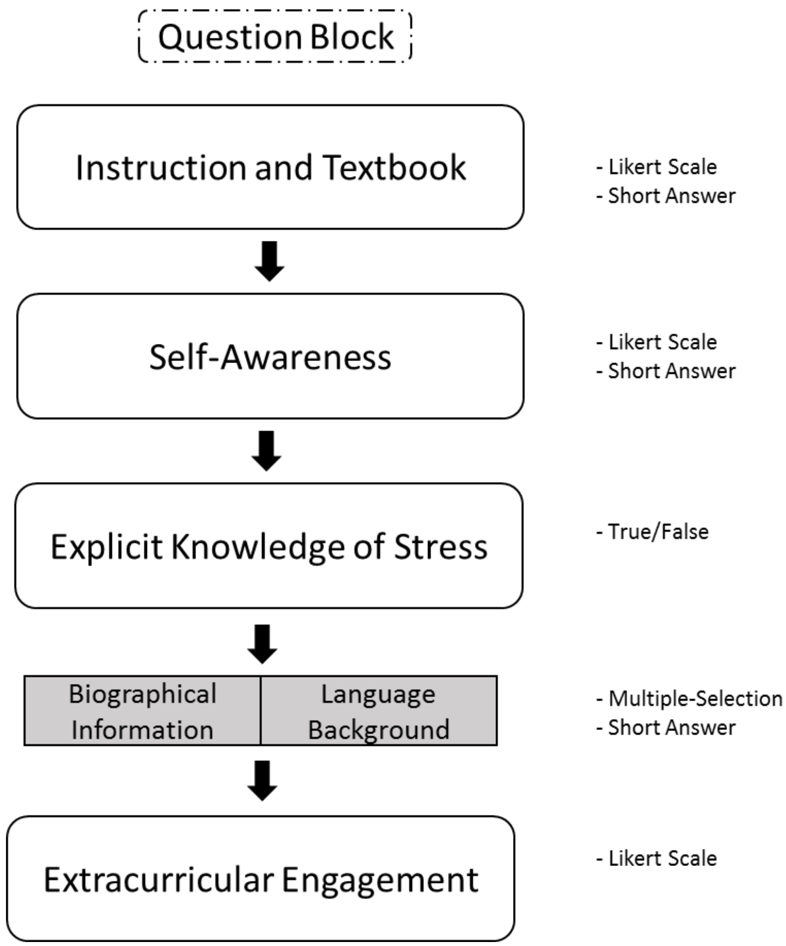
3.1.5. Post-Experiment Questionnaire

To assess on the effect of external factors that might not be directly assessible via the experiment and to answer RQ 6, which is re-stated below, a post-experiment questionnaire was conducted at the end of the experiment. The questionnaire was prepared and made available online using the Qualtrics online survey services. The questionnaire complemented the experimental data and aimed to obtain more detailed information about the participants that might help explain across participant variation.

RQ 6. Do other external factors (specifically explicit knowledge of lexical stress, instruction, self-awareness and extracurricular engagement with the target language) influence the learners' performance on stress perception and production and, if so, how?

Precisely, in addition to questions inquiring about participants' biographical information and language background , the questionnaire had four blocks of questions to assess four external factors: (1) participants' perception of the instruction that they have received from the instructor and the textbook for lexical stress in Arabic, (2) self-evaluation of their ability and performance on lexical stress in Arabic, (3) their explicit knowledge about the formal rules of Arabic stress, and (4) their extracurricular involvement with the target language. Most of the questions in the questionnaire were declarative sentences and the participants expressed their attitude towards these sentences using a five-point Likert Scale (Strongly Agree, Somewhat Agree, I Don't Know, Somewhat Disagree and Strongly Disagree). Additionally, some short open-ended questions were included to elicit the response in more detail. Lastly, a short quiz consisting of ten yes/no questions was implemented in the questionnaire to assess participants' explicit knowledge about the rules of lexical stress in Arabic. The sequencing of the questions as well as the eliciting method are illustrated in Figure 3.6.

Figure 3.6 Sequencing of Questions in the Post-Experiment Questionnaire



The questionnaire began with questions that probed the role of the instructor and the textbook used in the Arabic class in the instruction of lexical stress. Given that stress is an integral part of pronunciation, one would assume that the topic of stress would be extensively discussed by the instructor in the classroom or in the textbook. However, this does not seem to be generally true. For example, although being a rigid curriculum in the United States, the *Al-Kitaab* series does not mention the stress rules in Arabic in any of the three textbooks. In contrast, the *Al-Jadiid* series, a curriculum that is mainly used in China, has detailed explanation of the stress rules in Arabic from the first lesson of the textbook. Nonetheless, a very limited number of drills is available. This disparity in attention to lexical stress in Arabic could also be found in the classroom, as some instructors might be more attentive to learners' pronunciation of the stress, whereas some others might focus more on the fluency aspect of language production, i.e. the ability to produce or to comprehend uninterrupted speech.

Due to this disparity in instruction and textbook, these questions were included in order to provide a general assessment of the instruction that participants received from both the instructor and the textbook. The questions, as presented in Appendix 3.5.1, generally asked whether the participants perceived themselves to have received any explicit instruction on stress or any opportunity to practice stress in class from their instructors. Additionally, the questions also inquired about whether participants perceived themselves to have received instruction on stress from the textbook used in their Arabic class.

The second part of the questionnaire was composed of questions related to participants' self-evaluation, self-awareness and the degree of notice when dealing with tasks that are relevant to lexical stress in the target language. These questions, given in Appendix 3.5.2, asked whether participants perceive themselves to be capable of producing stress correctly and of evaluating stress production done by other people. In addition, the questions investigated the degree of stress awareness and their notice of stress during their production and whether they had made any effort outside of class to gain more knowledge about stress in the target language.

The third block of questions was a quiz targeting participants' explicit knowledge of the formal stress rules in Arabic. The purpose of the section was to examine to extent to which the participants could explicitly apply their knowledge about stress in a more formal context. Specifically, the quiz contained true and untrue statements, given in Appendix 3.5.3, regarding lexical stress in Arabic. Through these statements, the participants were required to access their explicit knowledge of the stress rules in Arabic and use this knowledge to determine the validity of the statement. These questions examined whether participants understood how the vocalic, consonantal and syllabic structure could influence the position of stress. Additionally, these questions also probed knowledge of possible and impossible position for stress.

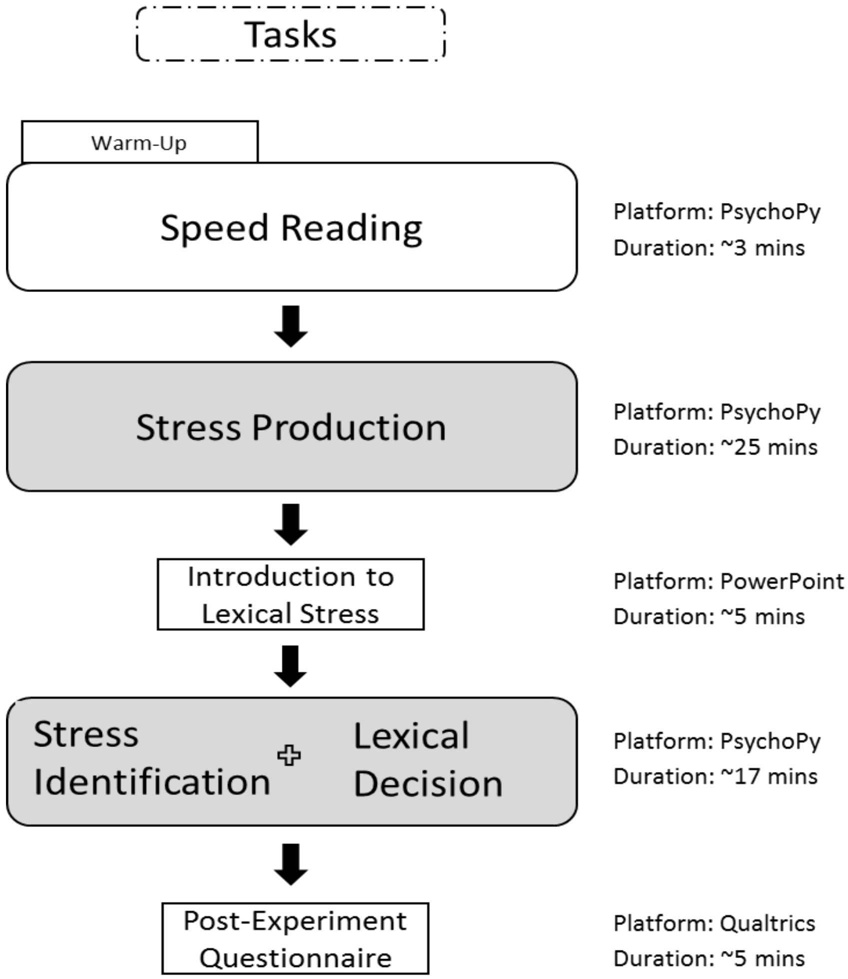
The fourth block of the questionnaire consisted of questions inquiring about the biographical information of participants as well their language background. The purpose of this section was to obtain more detailed information about the participants that might inform the inherent individual variance between participants within and across the same proficiency levels. As presented in Appendix 3.5.4, in addition to the basic biographical information, these questions probed deeper into the language that the participants used in different social contexts, as well as other foreign languages that the participants had studied before and the duration of the study. This section further investigated the participants' experience with Arabic in contexts other than the language classroom. Specifically, this section inquired about the travel experience and study-abroad experience to confirm whether the participants indeed met the requirement set for recruitment.

The last section of the questionnaire investigated the social engagement with the target language and the opportunity to use it outside of class (see Appendix 3.5.5). Since language learning does not only happen in the language classroom, the purpose of this section was to assess the opportunities as well as resources available to the participants outside of class. Participants were explicitly asked questions such as whether they had friends who are native speakers of Arabic, who might be a resource of learning outside of classroom. The participants were asked to evaluate the effort that they put into the four fundamental skills of language learning, i.e., speaking, reading, listening and writing.

3.1.6. Organization of The Experiment

The primary objective of sequencing of the experimental tasks was to minimize the possibility of participants' realization of the purpose of the experiment. In other words, the sequencing of the tasks needed to conceal the fact that the experiment is primarily concerned with the fluency and the accuracy of the participants' production and perception of stress. To achieve this goal, the proposed organization of the experiment is structured as in Figure 3.7.

Figure 3.7 Sequencing of Tasks in The Experiment



As shown in Figure 3.7, the experiment begins with the warm-up activity and is followed by the stress production task. The motivation for having the stress production task as the first main task was two-fold. To begin with, the stress production task sequenced better with the speed reading drill since both are of the same modality, which required the participants to read aloud Arabic scripts as quickly as possible. More importantly, it was less likely that the participants would realize that the stress production task targeted production of stress, because the description of the task did not have any mention of lexical stress. The same holds for the other tasks of the experiment. All that is required of the participants was to read aloud the Arabic script on the screen.

The stress production and embedded lexical decision task followed by the stress identification task as it is the only remaining task remaining given that the lexical decision tasks integrated into the stress production task. However, instead of proceeding directly to the stress identification task, some preamble was required to help the participants better understand what the tasks requires them to do. Precisely, since the stress identification task required the participants to determine the stressed syllable, it was essential for the experiment to ensure that the participants knew what the terminologies such as “syllable” or “stress” were and how they were realized in the language before proceeding to the task.

To achieve this goal, a brief introduction to stress that was developed in Microsoft PowerPoint was presented to the participants before the stress identification task, which consisted of an automated PowerPoint presentation accompanied with audio recording as the narrative for the presentation. The audio recording was done by a female native speaker of English, and subtitles in Chinese and English were prepared to accommodate participants who were not native speakers of English. The presentation as well as the narrative is presented in Appendix 3.6. As shown in Appendix 3.6, this introduction is by no means an introduction to the stress rules in the target language; instead, it uses English, a bridging language that all the participants know as either their first language or as a second language, to help the participants review the concept of lexical stress and the acoustic features associated with it. In the second half of the presentation, the narrative engages and interacts with the participants on questions such as “how many parts or syllables there are in this word?” or “which part or syllable of the word is stressed” while showing some English words as examples. These examples were pre-syllabified and the syllables were numbered to help the participants anchor the relative position of the syllable, which was exactly how the stimuli were presented in the upcoming stress identification task.

As discussed in earlier sections, all tasks in the experiment had two phases, in that they task started with the practice phase then was followed by the testing phase. Since all tasks are self-paced tasks, the participants had full control over the pace at which they proceeded with the task. Therefore, the experimenter was only present in the practice phase for each task to resolve any difficulties or questions during the practice phase, and to check whether participants followed the instruction of the task and proceeded correctly. Once the participants were familiar with the task and were proceeding to the testing phase of the task, the experimenter moved to a nearby room, and left the participants by themselves in the testing phase.

The experiment concluded with the post-experiment questionnaire, which utilized a different modality compared to the previous tasks and no longer required Arabic. The experimenter was present at this stage of the experiment to answer any questions and offer clarification regarding the survey questions if need be. After the participants completed the questionnaire, a very brief interview in the form of short exchanges was conducted to inquire about the participants' general impression of the experiment and what they thought the experiment was targeting. This short exchange was intended to see if the participants thought they knew what the purpose of the experiment was. The majority of the participants who completed the experiment did not know that the experiment was exclusively about lexical stress in Arabic, as most of them regarded the experiment to be a general assessment of their proficiency in the language. This impression of the experiment suggests that the sequencing as well as the filler task, i.e. the lexical decision, seem to be effective in terms of diverting participants' attention and concealing the main objective of the experiment.

In summary, the proposed framework of the experiment aimed to answer RQ 1 and 2, which examined the role of native language and proficiency in stress perception and production respectively. The stress identification and the stress production tasks assessed the participants' ability to locate the stressed syllable of a given auditory stimulus and to recreate it in oral production respectively. The lexical decision task, albeit mainly serving as a filler task, targeted the participants' explicit knowledge of the Arabic lexicon. The experiment concluded with a questionnaire, with questions that investigate the various external factors that might potentially contribute to participants' performance in stress perception and production. With this experimental framework, the remaining research questions that have not been addressed were ones that had to do with the effect of frequency on stress perception. These research questions are discussed in the next section in Section 3.2.

3.2. Stimulus Design

To investigate the effect of frequency, and to answer research questions 3 to 5, which are restated below, frequency must be integrated as a variable to the stimuli.

RQ 3. Does the frequency of relevant structures for stress (i.e., position of the stressed syllable, syllabic structure of the stressed syllable, and stress pattern of the word) influence the participants' performance on stress perception and production, and, if so, how?

RQ 4. Does frequency have a different effect on stress perception as compared to stress production?

RQ 5. Does frequency have a different effect on learners as compared to native speakers?

Simply put, integrating frequency as a variable of the stimuli makes it possible to examine how participants' performance, i.e. the fluency and accuracy of participants' production, is influenced by the frequency of the stimuli. Following previous investigations of frequency effect, it is predicted that participants will perform better (i.e., more accurate and more fluent) in the task when the stimuli are frequent.

To be able to address these questions, the concept of the frequency of lexical stress needs to be defined in more detail first. Specifically, it is necessary to be able to answer the following questions:

1. What are the structures that are relevant to lexical stress in Arabic?
2. Among these relevant structures, to what extent to are they varied in frequency?
3. What are the structures that are more frequent and less frequent?

Because previous studies did not provide answers to these questions, a corpus analysis was conducted to estimate the frequency distribution of relevant structures for lexical stress in Arabic. As indicated by the algorithm for lexical stress in Arabic, stress placement in Arabic is an interplay between syllable weight and its position, in that the location of stress in a word is determined by the syllabic structure (i.e., weight) of the syllables and their positions in the word. Additionally, the stress pattern of the word should also be considered as one of the relevant structures as sometimes the learners would treat the input as a whole without further processing it.

Previous studies on statistical learning have also identified that not only the occurrence, but also the co-occurrence of the relative structures could inform language production. This, not only do syllabic structure as well as the position of the stressed syllable matter, but so does the probability of two relevant structures co-occurring (that is, the probability of a specific relevant structure for stress given that it is preceded by a particular relevant structure). Together with the relevant structures identified from the stress algorithm, this study posits that the following structures are relevant for stress perception and production, and are therefore manipulated and controlled for: (1) the stress pattern of the word, (2) the stress position in the word, (3) the syllabic structure for the stressed syllable and (4) the conditional probability of the stressed syllable.

To elaborate, stress pattern here is construed as the abstraction of the syllabic structure of the word with information on of position of the stressed syllable. For example, for the word نتيجة "result" [na.ˈtiː.dʒa], by abstracting away the phonemes of the word into representations "C" and "V" for consonant and vowel respectively, the stress pattern for the word then becomes [CV.ˈCVV.CV] where [VV] stands for a long vowel for clarity purpose. This stress pattern can be interpreted as having a word that begins with a [CV] syllable, is followed by a stressed [CVV] syllable and ends with a [CV] syllable.

From this stress pattern, other relevant structures for stress can be determined accordingly. As dictated by the algorithm for stress in Arabic, there are three possible positions for stress, which are final, penultimate or antepenultimate position. For the stress pattern [CV.ˈCVV.CV] , the stress falls on the second from the last syllable, that is, the penultimate position. With regard to the structure of the stressed syllable, the structure [CVV] is extracted. Lastly, the conditional probability of the stressed syllable for the example can be understood as the likelihood of having a stressed [CVV] given that the preceding syllable is an unstressed [CV].

After understanding what constitutes the relevant structures for lexical stress in Arabic, it is essential to establish a link between these structures, frequency and more importantly usage-based language acquisition. As discussed Section 2.3.1, the usage-based learning mechanism is characterized as learning the regularities of language constructions, i.e., anything that goes together, by being repeatedly exposed to input that contains similar or identical structure. The more frequent a construction is, the more frequently the learners are exposed to it and the more better that construction is entrenched, resulting in a higher processing advantage for this construction. For the present study, I posit that the acquisition of stress also follows this learning mechanism, as learning the stress system in a language fundamentally is to learn how syllables with varying syllabic structure, position and its stress status go together. As a result, the more frequent these stress-related constructions are, the more the learners are exposed to them, and arguably the more these constructions are readily available for them to use, resulting in better performance in stress perception and production.

Just as L1 learning children would abstract utterances that they hear numerous times such as “take it to me”, “give it to mommy”, “bring it to daddy” to a more abstract construction “V something to someone” with a semantic prosody of a demand for delivering objects, it is posited in this study that the L2 learners will follow this learning pattern and elevate concrete linguistic input to an abstract representation of it. That is, during the course of learning Arabic, the learners are exposed to words such as كتاب [kɪ.'taːb] “a book”, or جبال [jɪ.'baːl] “mountains”, or مثال [mɪ.'θaːl] “an example” many times and notice that these words all seem to share the same prosodic structure [C_1i.'C_2aːC_3] with stress falling on the second part of the word, and specifically on a syllable with a long [a:] that begins and ends with a consonant. Accordingly, whenever the learners encounter other Arabic words that have the [C_1ɪ.C_2aːC_3] syllabic structure, they would know that the stress falls on the second syllable, and the confidence of this decision increases as they are exposed to more words with this pattern. Meanwhile, the learners would also encounter other words that share a similar construction only with differing vowel context, such as بلاد [bɪ.'laːd] "countries", سلام [sa.'la:m] "peace" or جمال [ja.'maːl] "beauty". By being exposed to these words repeatedly, the repetition would allow the L2 learners to realize that the position of the vowel does not seem to influence the way stress is assigned in the language, but vowel length does. As a result, the learners further modify and abstract the pre-existing prosodic construction [C_1i.'C_2aːC_3] to one that is more abstract such as [C_1V.'C_2VVC_3] – exactly what is referred to as the stress pattern in the earlier section.

As noted before, in order to achieve higher confidence in the learned construction, frequent exposure to the construction in question is required. If a construction is not frequent enough, the opportunity for exposure is reduced, which in turn would reduce learners' confidence in producing the construction and increase the learners' susceptibility to be influenced by competing preferences learned from other features of constructions. In other words, not only do the learners use the learned abstract presentation of entire word to help them determine which syllable to stress, but they would also use other features, such as the position of the syllable, the syllabic structure, or the surrounding syllables to help them decide the stressed syllable. For example, from the repeated exposure to bisyllabic words in Arabic, the learner could notice that stress frequently falls on the first syllable. Therefore, due to this prominent feature of bisyllabic word, when given a word with the syllabic structure of [C_1ɪ.C_2aːC_3], it is possible for the learners to incorrectly stress the first syllable *['C_1ɪ.C_2aːC_3] rather than the second. In the same vein, the learners could also be biased either positively or negatively by other cues that are available in the construction. The more frequent these cues, or "relevant structures", are in the language input, the more likely it is that the learners will be led by these cues when determining the stressed syllable. Hence, not only does the frequency of the stress pattern, the frequency of aforementioned relevant structures, i.e., position, syllabic structure and conditional probability should all be considered together to establish the link between these relevant structures of stress and their frequency, and how their frequency could influence learning.

The organization for the reminder of the section for stimulus construction is the following: Section 3.2.1 introduces the source of the dataset upon which the analysis was conducted, followed by the presentation of results of the analysis in Section 3.2.2. A comparison is made between the frequency distribution of stress obtained from the dataset and other sources in Section 3.2.3., followed by the discussion of the organization of the experimental stimuli in Section 3.2.4. Lastly, the approach taken to construct nonsense words used for the stimuli is presented in Section 3.2.5.

3.2.1 Estimating Frequency Distribution of Stress Patterns in Arabic

With the relevant structures for stress identified (i.e., stress pattern, the syllabic structure, the position and the conditional probability of the stressed syllable) the next issue is to estimate the frequency distribution of these relevant structures and to examine whether differences in frequency exist among these structures. To capture the frequency distribution of the above-mentioned relevant structures for stress, I considered three sources from which the frequency information was accessed and cross-validated. The main source for estimating frequency distribution is a frequency dictionary for Arabic, which is cross-validated by using an Arabic-English dictionary and the textbooks used by the learners' home institution.

As the main source of estimation, the Arabic frequency dictionary compiled by Tim Buckwalter and Dilworth Parkinson (2011) was selected for its relevance and the ease of accessibility at the time of designing the present study. This dictionary presents the most frequent 5000 words based on a 30-million-word corpus of spoken and written Arabic, providing an extensive overview of possible types of stress patterns in Arabic. More importantly, the frequency information of these 5000 entries is provided as well, which is greatly beneficial approximating to the real-life usage of these words. In particular, the raw frequency of these words is provided, which reveals important information regarding how often these words or these stress patterns of these words appear in real-life usage – the very information this analysis seek to accomplish.

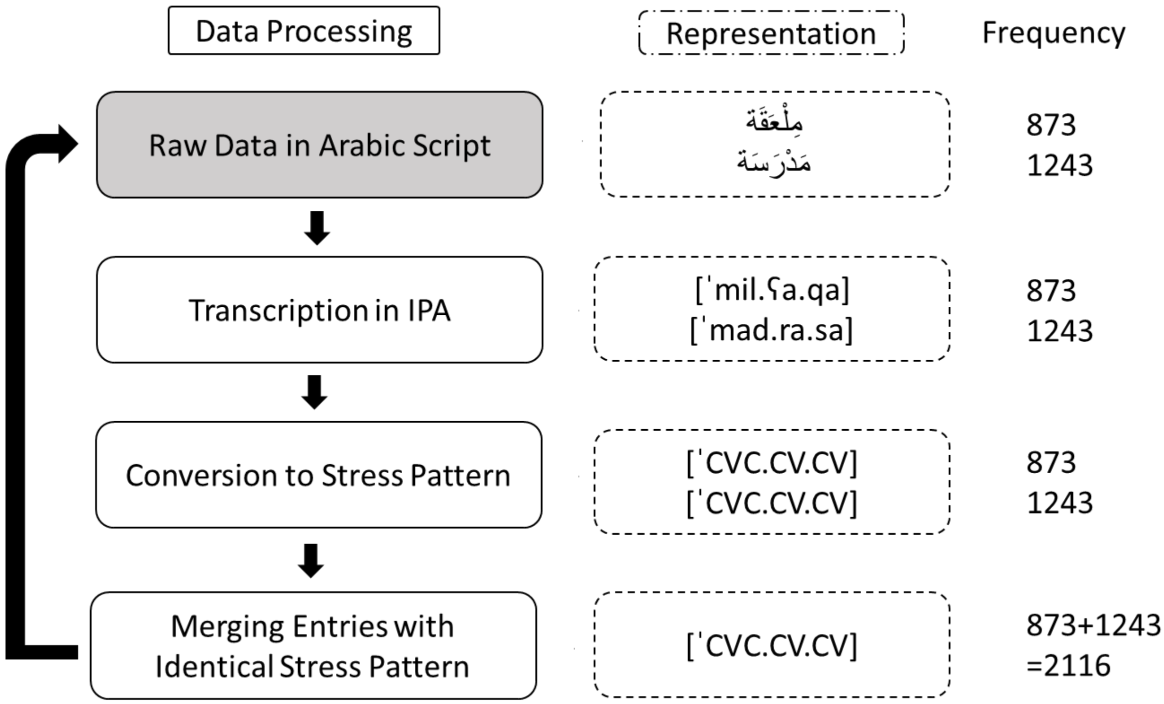
For validating the main source, the Arabic-English dictionary compiled by Hans Wehr (1993) was selected as the second data source at the time of designing the study. This iconic dictionary is an important reference for contemporary Arabic. Originally compiled in the 1950's, the dictionary has been edited and amended with new entries to better reflect the language used today. The Hans Wehr dictionary is also a popular choice among learners of Arabic to use as a learner's dictionary around the world. More importantly, the dictionary is more electronically accessible than other Arabic-English dictionaries at the time of analysis, which makes information extraction more manageable. Since the frequency dictionary only has 5000 words, it is not clear how representative these words are for the frequency distribution of stress pattern in Arabic as a whole. This dictionary is used to validate the extent to which the frequency distributions yielded from the frequency dictionary could be representative of the entire language.

Lastly, the textbooks that the learners used at their home institutions were selected to validate the third source of estimation. The language used in the textbooks is an important indicator for the accuracy of estimation, since the language, especially the vocabulary, included in the textbooks could be said to be the minimal input that is available to the learners in the curriculum. By analyzing the vocabulary included in the textbook, it becomes possible to assesses the frequency as well as the sequencing of the vocabulary in the textbook, which in turn would allow us to estimate the accumulation of stress patterns over the course of the curriculum.

3.2.2 Data Extraction

To extract the frequency distribution of stress pattern from the frequency dictionary, every single entry in the dictionary undergoes the same procedure presented in Figure 3.8. The first step is to transcribe the raw data which is presented in Arabic orthography in the dictionary into machine readable script. Here, the Arabic script was converted to the International Phonetic Alphabet (abbreviated as IPA in the figure). With the full vocalization provided by the dictionary, the pronunciation of the word along with the stressed syllable for the entry can be easily identified using the stress algorithm for MSA and fully transcribed accordingly. The raw frequency of the entry is recorded as well. The next step is to convert the IPA symbols into Cs(consonants) and Vs (Vowels) with stress information. The last step is to combine entries that have the same stress pattern in order to calculate the frequency of that stress pattern.

Figure 3.8 Data Processing Procedure for Frequency Dictionary



To illustrate, consider the entry مِلْعَقة "a spoon", which has a raw frequency of 873 meaning that the word has occurred in the corpus 873 times. The vocalic markings of the entry allow full restoration of its pronunciation, which is further transcribed as [ˈmɪl.ʕa.qa] in IPA. Accordingly, the IPA transcription is converted into representation for stress pattern by replacing [m], [l], [ʕ] and [q] with "C", and [ɪ] and [a]s with "V", resulting in the stress pattern [ˈCVC.CV.CV]. The last step is to check whether there is any processed entry which also has the same stress pattern; if so, then the frequency for the two entries will be added up. For instance, if there exists another entry مَدْرَسَة [ˈmad.ra.sa] "a school" which also happens to have the stress pattern [ˈCVC.CV.CV], then the raw frequency for the two entries مِلْعَقة and مَدْرَسَة ,which are 873 and 1243 respectively, will be merged, resulting in the raw frequency of 2116 for stress pattern [ˈCVC.CV.CV]. The same process follows until all entries are processed, which ultimately results in a set of unique stress patterns and the corresponding frequency counts. This set of stress pattern and the corresponding frequency is what the estimation seeks to achieve – a frequency distribution of stress patterns in Arabic.

As for analyzing the second data source, i.e., the textbooks used in the learners' home institution, a somewhat different approach was taken due to the different modality of the data source. Unlike a dictionary which is a list of entries and the corresponding definition, textbooks normally have more variety in their organization. A typical language textbook consists of lessons that can be further divided into main texts for the reading, instruction on grammar, drills and vocabulary. Ideally, to assess the frequency distribution of the stress patterns included in the textbook, all instances of words included in any part of the lesson of the textbook should all be counted and analyzed. However, this approach would have been viable if all the Arabic words in the textbook are adequately vocalized, which will allow extraction of the full pronunciation of a word solely from the orthography of the words. Nonetheless, full vocalization throughout the entire textbook was not available for the majority of Arabic textbooks.

For example, the *Al-Kitaab* textbooks (Brustad et al. 2011) – a popular series of Arabic curriculum consisting of three textbooks– only have fully vocalized orthography of a word in its first appearance in the vocabulary list in the beginning of the lesson. After the first occurrence, the script for the word is deprived of any vowel information. The assumption behind this practice is that after the word has been introduced with full vowel information, the learners will know the vocalic context of the word the next time they encounter it. Hence, from the author's perspective there is no need to fully "spell out" the word again in the later sections of the book. Additionally, it is assumed that words without full vocalization serve as an opportunity for practicing and activating the vocabulary.

Regardless of the question of pedagogical merits of such vocalization practices, it does pose a challenge for data extraction in the present study. It was rather challenging to disambiguate and transcribe Arabic text automatically without vowels. Without proper vowel and germination markings, the syllable structure and its full pronunciation can only be retrieved manually, which requires considerable time, given the number of words included in each textbook across over the regular 3-year curriculum. Due to the limited resources at hand at the time of designing the study, analyzing all occurrences of words in the textbooks was not a viable approach.

Instead of analyzing all instances of Arabic words in the textbook, the present study only focuses on the new vocabulary that appears in each lesson. Following the typical organization of language textbooks, *Al-Kitaab* textbooks include a vocabulary list in the beginning of every lesson. These vocabulary lists contain almost all, if not all, new words that the student will encounter in the lesson, which constitute the main body of lexicon for the learners in the target language. Therefore, by analyzing these vocabulary lists, it becomes possible to identify the lexicon that the learners are required to acquire under the curriculum, as well as the order in which the lexicon is acquired.

To analyze these vocabulary lists, the same data processing procedure was followed, as presented in Figure 3.8. However, before proceeding to the analysis, there is one remaining issue with the vocabulary lists that needs to be resolved, which is the lack of the frequency information. Unlike the frequency dictionary, the vocabulary lists from the textbook do not contain any information regarding how frequently the words occur in real-life usage or at the very least in the language classroom. Therefore, to assess the frequency of the words in the vocabulary lists, a frequency extrapolation was conducted between the vocabulary lists and the frequency dictionary. That is, for each word in the vocabulary lists, I consulted the frequency dictionary to see whether the word exists in the dictionary; if it does, then the frequency of the word will be regarded as having the same raw frequency as the entry found in the frequency dictionary. After the extrapolation, nearly all words collected from the vocabulary lists have a frequency count, which allows for data processing as presented in Figure 3.8.

The rationale of this approach is that, since learners do not have any knowledge of Arabic when they start learning the language, they generally follow the drills and activities in the textbook lesson by lesson. An important section of each lesson is the vocabulary list that the learners spend a significant portion their time on during the course of studying the material. Additionally, the vocabulary list lays out almost all, if not all, the new words that the learners are required to learn, which serves as the main input that the learners will receive in the lesson. For the learners, these vocabulary lists represent the minimal language that they have to acquire in each lesson or to a greater extent, in the curriculum. As for the analysis here, these lists represent the range of input that the learners will be exposed to in the curriculum. The benefits of analyzing these lists are two-fold. First, the analysis will provide insight into the range of the types of stress pattern that the learners will be exposed to in the curriculum. Additionally, since the curriculum generally follows the sequence of the lessons, it means that the learners will consistently encounter some of these words first and some latter. Analyzing the sequence of these words will capture the types of stress pattern that have been accumulated by the learners over the course of the curriculum. However, this analysis only indicates the types of stress pattern that the learners encounter but not how frequently they encounter these patterns. To capture the frequency of the words on the lists, I used the frequency dictionary as the reference for these words. I extrapolated the frequency from the dictionary to the words, and then calculated the frequency of the words that had the same stress pattern. By considering the sequencing of these words, i.e., the lesson where the words were found, in the analysis, it becomes possible to obtain knowledge about the trajectory of the type and the frequency of the stress pattern that the learners will encounter at any stage of the curriculum. An overview of the trajectory of frequency distribution of stress pattern will be introduced in section 4.2.

The underlying assumption is that the input, i.e., the frequency distribution of the stress patterns, in the classroom is more or less similar to the input that one would receive in real-life. This assumption means that the frequency distribution of stress pattern found in the textbook should be more or less similar to one found in the frequency dictionary. Although this assumption might not be entirely applicable for beginning learners, as they progress in the curriculum, the language used in class will become more and more similar to real-life usage. In other words, the frequency of stress patterns that the beginning learners are exposed to will become proportionally more and more similar to the frequency distribution of stress pattern identified from the frequency dictionary. This is because in the beginning of the learning process, the learners have a limited ability to produce and understand full sentences, which makes communication between instructor and learners more functional, mostly relying on keywords and simplified speech in the target language rather than full sentences. As the learners gain more knowledge about the language, the interaction between instructor and learners becomes more natural. The instructor would use complete sentences and a greater variety of vocabulary. In a more advanced classroom, i.e., the third or the fourth year in the curriculum, the instructor would interact with the learners in a way that is more similar to the interaction between native speakers. This is a common scenario in instructed second language acquisition, from which we can assume that although the input – the stress patterns – encountered by the learners might be different from real-life usage in the beginning, it will gradually become more and more similar to the input that one would receive in the real-life situation. To validate this assumption, the full analysis is presented in Section 3.2.4.

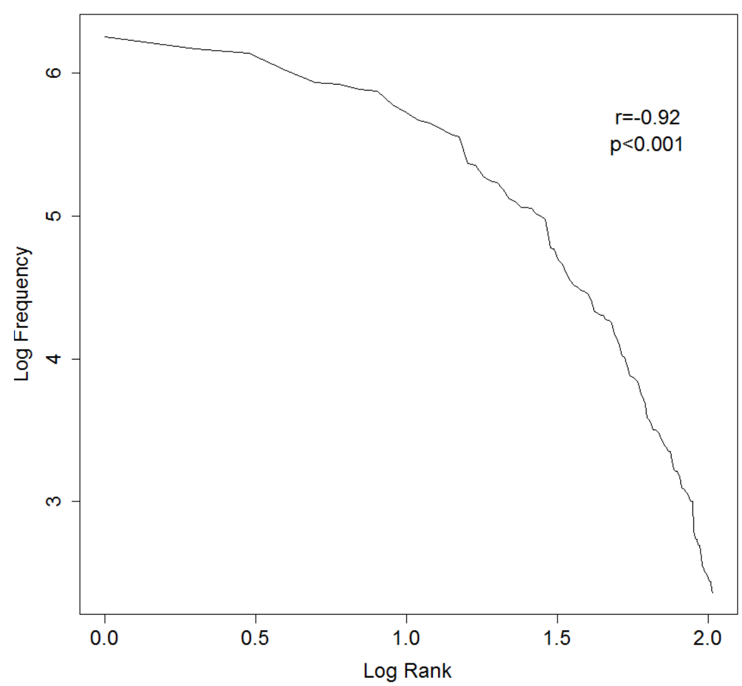
Lastly, for the third referred source, i.e., the Hans-Wehr Arabic-English Dictionary, the same procedure as in Figure 3.8 is followed. However, since the entries in the dictionary are pre-transcribed, the transcription is directly converted into stress patterns with minor adaptations from the transcription schema used in the dictionary to IPA.

3.2.3 Frequency Distribution of Stress Pattern In Arabic

This section presents the results of the analyses for the frequency distribution of relevant structures from the frequency dictionary, which are identified in the previous section as (1) the stress pattern of the word, (2) the stress position in the word, (3) the syllabic structure for the stressed syllable and (4) the conditional probability of the stressed syllable. The results of the analyses allow us to respond to the questions posed in RQ3 regarding whether there exists disparity in frequency between the relevant structures, the extent of the disparity and the subjects of disparity. It is worth noting that although all 5000 entries in the dictionary were analyzed, only entries that have bisyllabic, trisyllabic and quadrisyllabic stress pattern are presented in the section. The motivation for excluding entries that are monosyllabic or pentasyllabic is to be consistent with stimulus design that will be further discussed in Section 3.2.5. Also, among the 5000 entries found in the frequency dictionary, only ten are pentasyllabic. Having excluded monosyllabic and pentasyllabic entries, the following sections present the results obtained from a total 4283 entries in the frequency dictionary.

The first relevant structure is the stress pattern of a word, which is an abstract construction capturing the relation between consonant, vowel and the position of stress. A total of 104 distinctive stress patterns were identified from the frequency dictionary. To give an overview of the correlation between all stress patterns and their frequency, a plot using a log coordinate is presented in Figure 3.9, with the frequency rank of a stress pattern x axis and its frequency on the y axis. As shown in Figure 3.9, there is a significant negative correlation (Pearson's r = -0.92, $p < 0.001$) between the rank and the frequency. This negative correlation suggests that frequency of the stress pattern is not evenly distributed across all stress patterns.

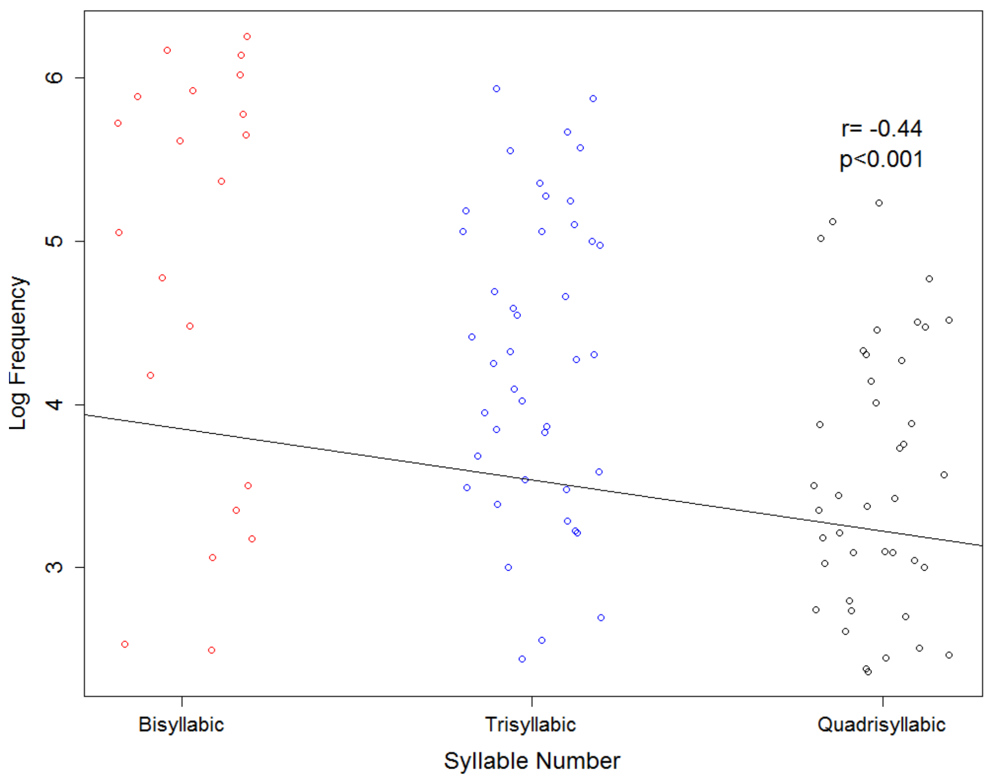
Figure 3.9 Correlation between Log Frequency and Rank of All Stress Patterns



Additionally, a comparison was made between the frequency distributions of bisyllabic, trisyllabic and quadrisyllabic stress patterns. As demonstrated in the scatterplot in Figure 3.10 with syllable number on the x axis, log frequency on the y axis and default jittering for presentational purposes, the comparison indicates a significant negative correlation

(Pearson's r = -.42, $p < 0.001$) between the number of syllables in a stress pattern and its log frequency, suggesting that stress patterns with fewer syllables, i.e., bisyllabic stress patterns, tend to be more frequent than stress patterns with more syllables, i.e., trisyllabic or quadrisyllabic.. This finding seems to conform to cross-linguistic preference for shorter words over longer ones.

Figure 3.10 Correlation between Log Frequency and Syllable Number of All Stress Patterns



The 104 stress patterns identified from the dictionary can be further divided into three groups consisting of 21 bisyllabic stress patterns, 42 trisyllabic stress patterns, and 41 quadrisyllabic stress patterns. Figures 3.11-13 present the frequency distributions of the stress patterns for these groups, with the type of stress pattern on the x axis and the log frequency rounded up to one decimal point on the y axis. The stress patterns in these figures are listed in decreasing order based on frequency count, which allows better observation for the most and the least frequent stress patterns. For example, for the bisyllabic stress patterns presented in Figure 3.11, the most frequent stress pattern is [ˈCVC.CV] with the log frequency of 6.2.

Figure 3.11 Frequency Distribution of Bisyllabic Stress Patterns

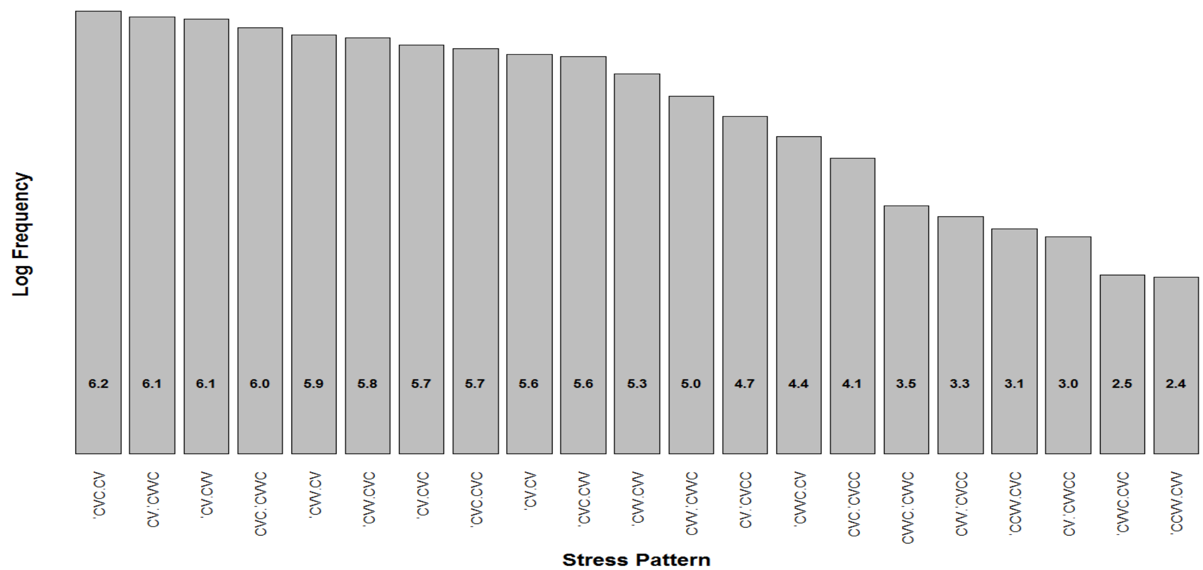


Figure 3.12 Frequency Distribution of Trisyllabic Stress Patterns

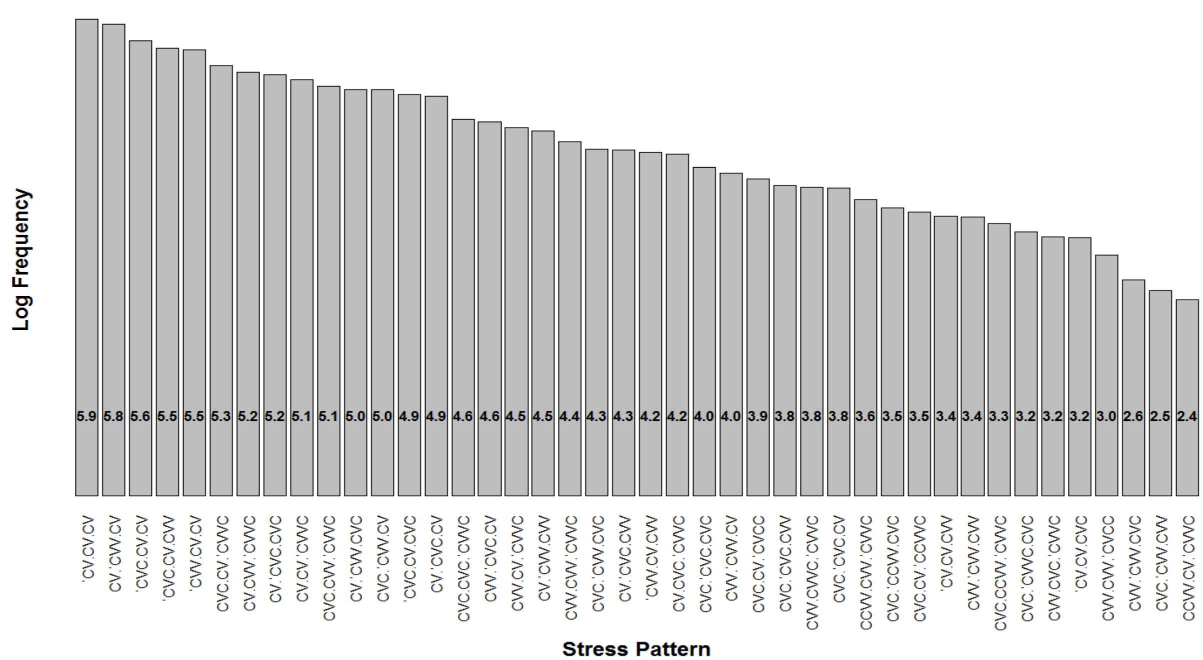
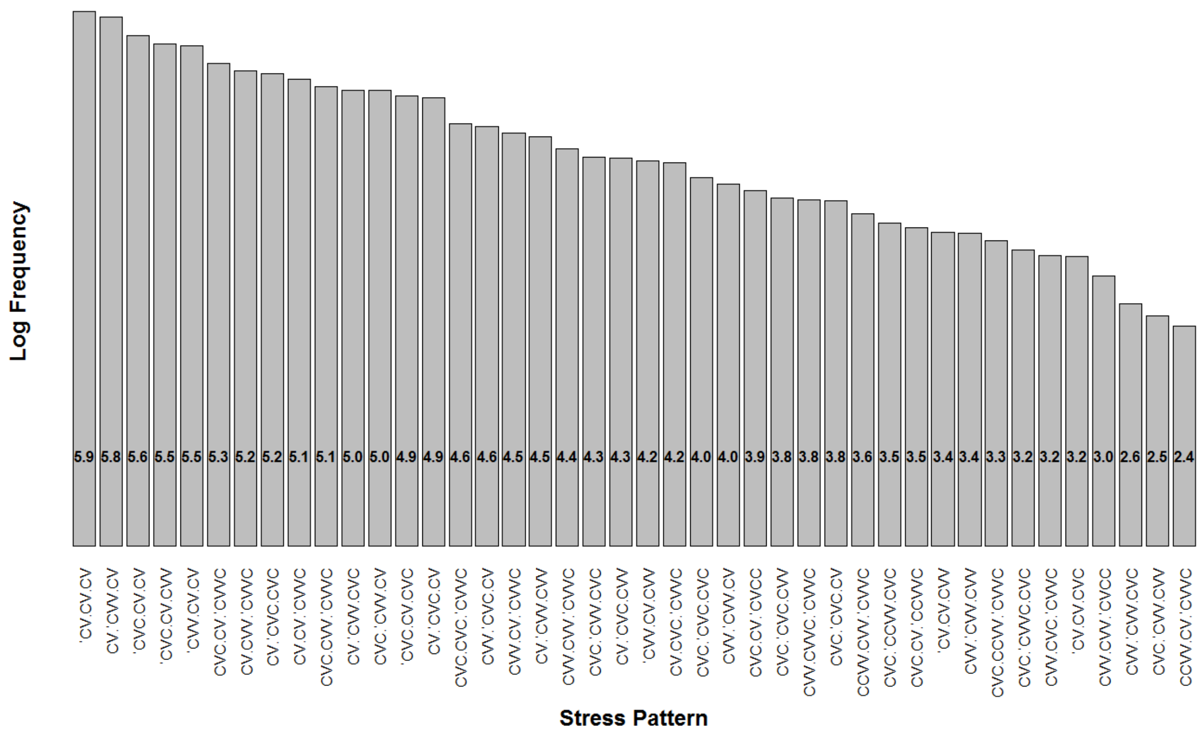


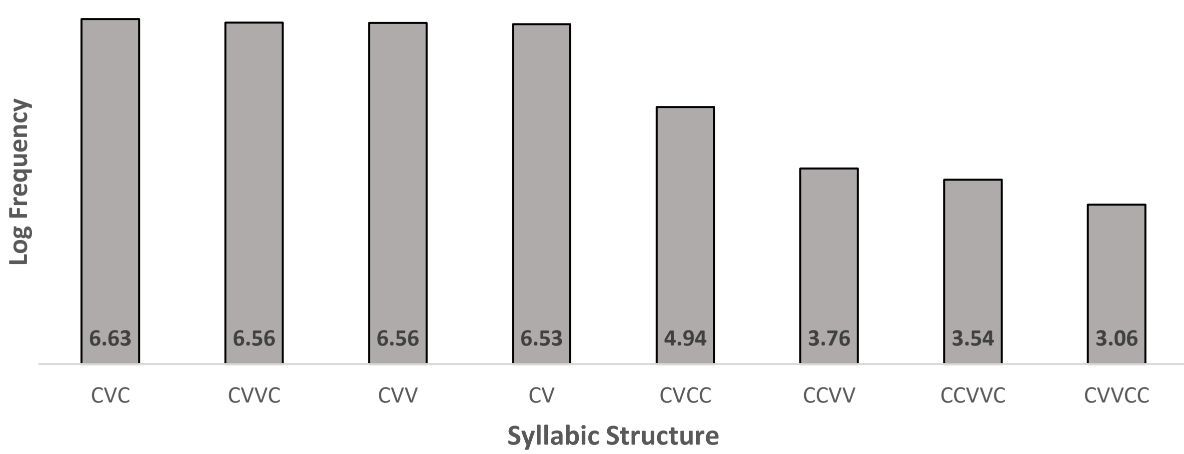
Figure 3.13 Frequency Distribution of Quadrisyllabic Stress Patterns



In Figures 3.11-13, it can be observed that for stress patterns of a specific number of syllables, some stress patterns are relatively frequent and some are so infrequent that their log frequencies could be less than a third of than the most frequent one. Given the mathematic nature of log frequency, even a 0.1 difference in log frequency could suggest a significant difference in raw frequency. The variability identified from the frequency distributions shows that the frequency of stress patterns needs to be considered as a variable, which is to be either manipulated or controlled for.

The second relevant structure examined is the structure of the stressed syllable for a given stress pattern. By analyzing the frequency distribution of the segmental structure, i.e., the consonant-vowel combination, of the stressed syllable, it becomes possible to examine whether there exists a preferred syllabic structure for stress and to identify the type of the syllabic structure that is generally preferred. Following the same procedure as extracting stress patterns, instead of extracting the syllabic structure for the entire word, only the structure of the stressed syllable is extracted along with its raw frequency. By merging the frequency count for words that have the same syllabic structure for the stressed syllable, the frequency distribution of structure of the stressed syllabic emerges as presented in Figure 3.14. Note that Figure 3.14 presents the results in descending order for all stress patterns without further grouping, with the structure for the stressed syllabic on the x axis, and the log frequency for the responding syllabic structure on the y axis.

Figure 3.14 Frequency Distribution of Syllabic Structure of Stress (All Stress Patterns)

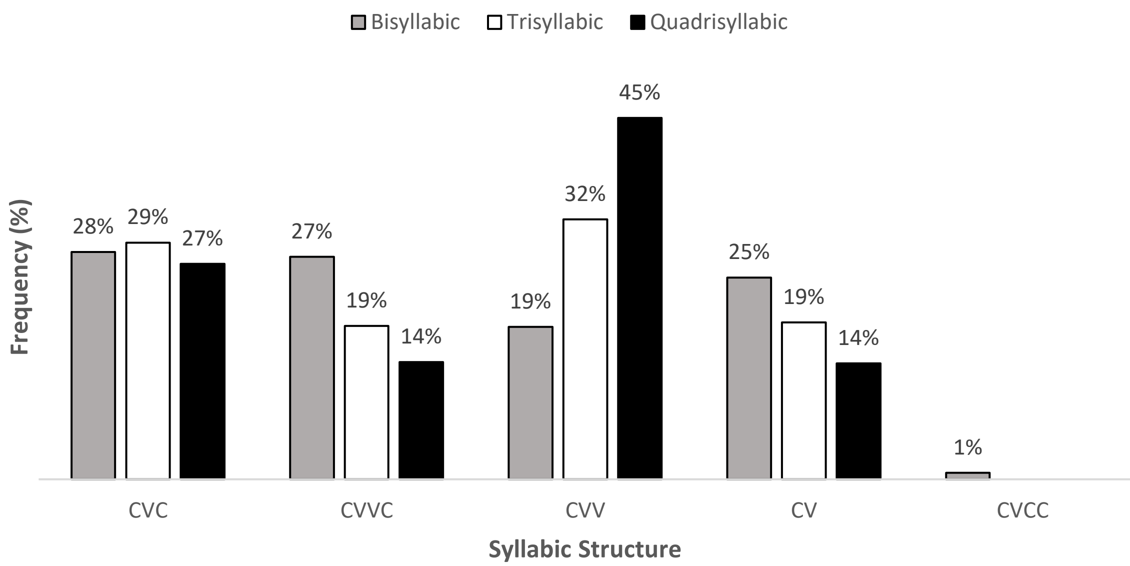


As shown in Figure 3.14, the frequency of the syllabic structure for the stressed syllable is not evenly distributed. That is, there are some syllabic structures that are preferred over some others, with relatively higher log frequency value. For example, the most frequent syllabic structure for stress is for a syllable to have a [CVC] combination, immediately followed by syllables that are either [CVVC] or [CVV], whereas syllabic structures such as [CVCC] or [CVVCC] are significantly less frequent. Note that the frequency distribution presented in Figure

3.14 has not taken the number of syllables of the stress pattern in consideration, which makes it easy to distinguish the difference in frequency between the more frequent stress patterns.

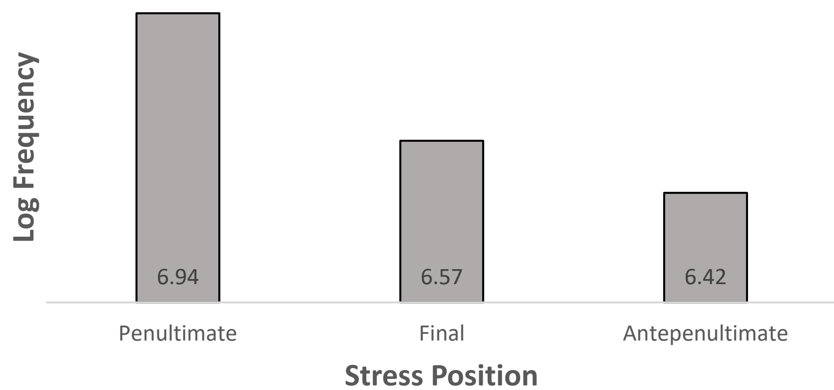
By grouping the results based on the syllable count of the stress pattern, a clear distinction between more frequent stress patterns and less frequent ones emerges, as presented in Figure 3.15, with the structure of the stressed syllable on the x axis and the frequency (presented as percentage of occurrence over all occurrences) on the y axis. Syllabic structures that account for less than 1 percent of occurrences such as [CVVCC] are excluded for the sake of conciseness of presentation. Stress patterns in words with different with numbers of syllable are presented in different colors, with stress patterns of bisyllabic words in gray, trisyllabic words in white and quadrisyllabic words in black, which makes the preferred syllabic structures more distinguishable, For example, it now becomes clear that syllables of the structure [CVC], [CVVC] and [CV] are most frequently stressed syllable in bisyllabic words, which respectively account for 28%, 27% and 25 % of all occurrences of bisyllabic stress patterns. These frequent stress patterns are followed by [CVV] with 19% occurrences, then by [CVCC] with only 1% of occurrence, which is significantly less frequent than the other bisyllabic stress patterns. Moreover, [CVCC] seems to only receive stress when the word is bisyllabic, as [CVCC] is not found to be stressed for trisyllabic and quadrisyllabic words. Similarly, for quadrisyllabic words, there seems to be a strong preference towards a [CVV] for stress, followed by [CVC] then by [CVVC] and [CV]. The results conform to the observation made in previous section, that the frequency or the preference for the structure of the stressed syllable is not evenly distributed. That is, for the stressed syllable of a word with a given syllable count, some structures are more preferred over some others. While this preference is subject to change for words with different syllable count. Therefore, the frequency of the structure of the stressed syllable needs to be taken into consideration to be either controlled or manipulated for stimulus design.

Figure 3.15 Frequency Distribution of Syllabic Structure of Stress Grouped by Syllable Number



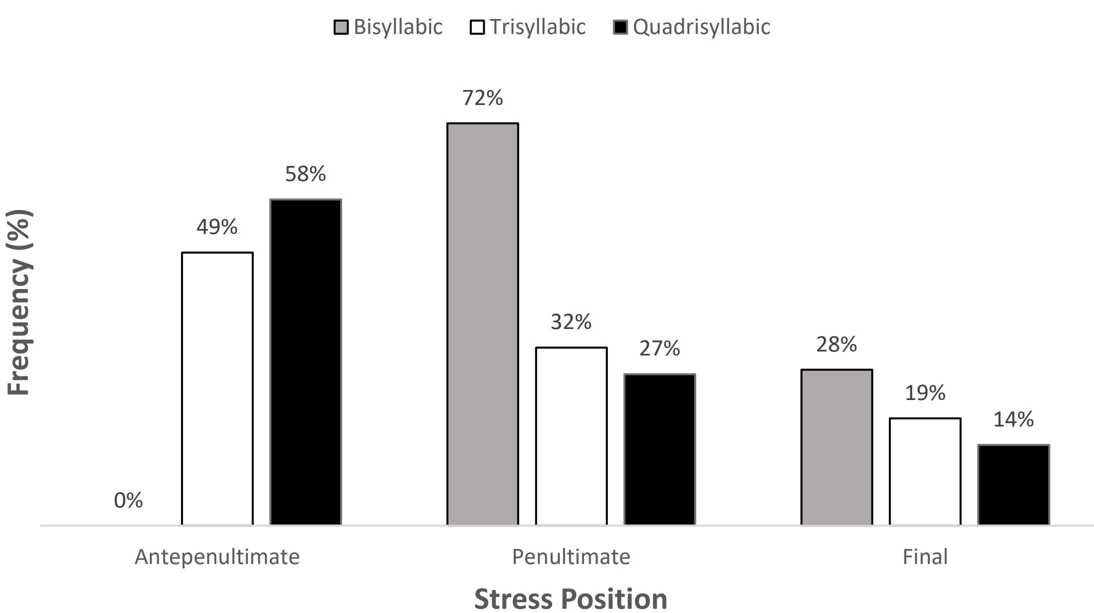
The third relevant structure for stress is the stress position; that is, the position of the stressed syllable in a given word. Based on the stress algorithm in Arabic, there are three possible positions for stress, which are either on the antepenultimate (the third syllable from the last syllable) or the penultimate (the second syllable from the last syllable) or final position (the last syllable) of any given word. Obviously, antepenultimate position is not possible for bisyllabic words, whereas the word-initial position is not possible for stress for quadrisyllabic stress patterns, as indicated by the algorithm. The procedure of obtaining the frequency distribution of stress position generally follows the procedure laid out in Figure 3.8. However, instead of extracting the entire stress pattern, only the stress position of the stress pattern is extracted along with its raw frequency and the frequency count is merged when the stressed syllables have the same stress position in the word. The result of this analysis is presented in descending order in Figure 3.16 without further grouping. The results show that overall in Arabic, the penultimate position in a word is more frequently stressed, followed by the final position then the antepenultimate. However, this difference does not seem to be as large as found in our previous analyses and the effect of number of syllables is not accurately reflected in Figure 3.16, which calls for further grouping of the results.

Figure 3.16 Frequency Distribution of Stress Position (All Stress Patterns)



By grouping the results by the number of syllables, Figure 3.17 reveals a more distinct frequency distribution that allows the examination of the preferred stress position for stress pattern of respective number of syllable. Figure 3.17 presents the results for the stress pattern of bisyllabic words in gray, the trisyllabic ones in white and quadrisyllabic ones in black, with the type of stress position on the x axis and the frequency (presented as the percentage of occurrence divided by all occurrences) on the y axis. The results show that having stress in the penultimate position accounts for 72% of occurrences of all bisyllabic stress patterns as opposed to having stress in the final position, which only accounts for 28% of all occurrences. In contrast, trisyllabic words seem to prefer stress on the antepenultimate syllable, which accounts for 49% of all occurrences, over the penultimate syllable (32%) and then the final syllable. This preference is also found in the stress patterns of quadrisyllabic words, as the antepenultimate position is again found to be the preferred position for stress (58%) followed by the penultimate position (27%) with the final position being the least favorable position (14%) for stress. The results again indicate the uneven frequency distribution of relevant structure, in that some positions in a word seem to be more frequent stressed than the others. Such preference could vary among stress patterns of different number of syllable; however, per the dataset obtained in the study, trisyllabic and quadrisyllabic words have the same preference for the position of stress, while bisyllabic words seem to have a slightly different preference for stress position. Again, this variability in frequency distribution of stress position calls for the need to consider the frequency of the position of stress to be a variable in the experiment.

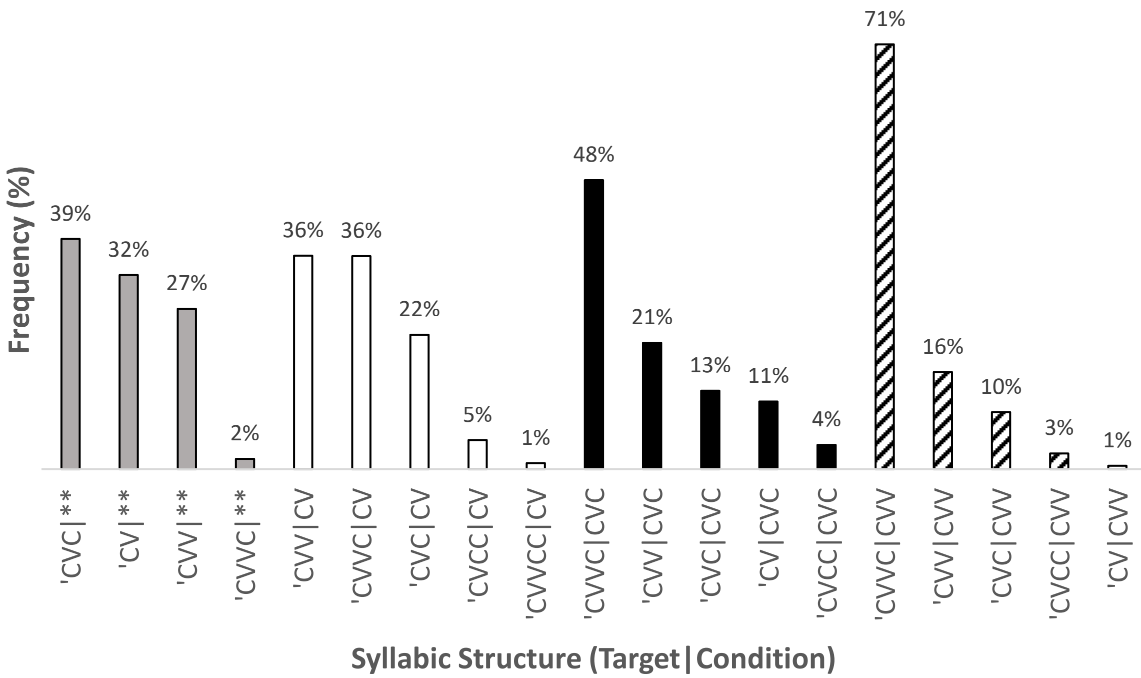
Figure 3.17 Frequency Distribution of Stress Position Grouped by Syllable Number



Lastly, as indicated in the literature, in addition to the occurrence of the relevant structures for stress, the co-occurrence of the relevant structures needs to be considered as well. The co-occurrence, or the conditional probability of the relevant structures is construed as the frequency or the probability of a stressed syllable given a preceding unstressed syllable as the condition. In other words, the conditional probability accounts for the odds of having a stressed syllable of a given syllabic structure X, granted that the preceding syllable is a unstressed syllable of a given syllabic structure Y, written as X|Y. In this case, Y is the condition for the stressed syllable X. Figure 3.18 presents the conditional probability distribution for all stress patterns grouped by the condition with the syllabic structure of the stressed syllable on the x axis and the probability on the y axis. The two asterisks "**" signify the beginning of a word; for example, "X|**" means the conditional probability of having a stressed X at the beginning of a word.

Overall, the results show that some syllabic structures appear to be more frequently stressed given a specific condition than some other syllabic structures. For example, for the stress patterns that have the condition of "**" (i.e., the beginning of a word), there is a preference for a stressed [CVC] with 39% of probability of occurrences, followed by a stressed [CV], [CVV], and lastly [CVVC] with lower probability. That is, it is more likely for a word to have a stressed [CVC] at the beginning of a word, while it is very unlikely to for a word to have a stressed [CVVC] at the beginning of a word. Similarly, given that the condition syllable is a unstressed [CVC] as presented by the black bars, a stressed [CVVC] has the highest conditional probability of 48%, followed by stressed [CVV], [CVC], and [CV] with stressed [CVCC] being the least possible to follow an unstressed [CVC]. The results reaffirm our previous findings of uneven distribution of frequency for relevant structures for stress, which needs to be taken into consideration when constructing the stimuli for the experiment.

Figure 3.18 Conditional Probability of Stress for All Stress Patterns



In sum, this section presents the frequency distributions of the relevant structures for stress in Arabic, from which the questions proposed in the earlier sections could be answered, which are re-stated below again.

1. What are the structures that are relevant to lexical stress in Arabic?
2. Among these relevant structures, to what extent to are they varied in frequency?
3. What are the structures that are more frequent and less frequent?

First, the present study posits that the frequency of (1) the stress pattern, (2) the syllabic structure, (3) the position and (4) the conditional probability of the stressed syllable in a given word might inform learners' performance. In response to the second question, the analyses conducted for the frequency distributions of these relevant structures showed that these distributions are generally heterogenous, in that the frequency for a given type of relevant structure is not unimodally distributed. Instead, the results show that there is a preference, be it strong or weak, for some specific structures to be stressed, while others are less frequently stressed, which confirms that there is difference in frequency among these structures. Additionally, difference is generally found in each frequency distribution, ranging from 0.1 to more than 3 in log frequency or 1% to more than 60% of all occurrence. Given that a slight difference in log frequency could suggest considerably distinct difference in raw frequency, hence, it could be said that within the same frequency distribution, some structures are considerably varied while some are not as distinctively different. Lastly, analyses clearly indicate the type of relevant structures as well as their relative frequency, which makes identifying both frequent and infrequent structures rather straightforward.

The variability in frequency found in these structures provides the opportunity to examine the effect of frequency by either manipulating and controlling for the frequency of relevant structures of interest. However, before proceeding to stimulus design and variable manipulation, it is essential to note that these frequency distributions are obtained solely from the frequency dictionary. To ensure that the frequency distributions obtained from the frequency dictionary is representative of the target language, and more importantly, reflective of the target language used in the language classroom, further validation is required and discussed in more detail in the next section.

3.2.4 Correlation between the Sources

As an integral part of the stimulus design, it is crucial to ensure that the frequency data obtained from the frequency dictionary is representative of actual language use, and is adequately reflective of the input that the learners received in the language classrooms. In the previous section, several frequency distributions of relevant structures were extracted from the frequency dictionary. To further investigate the validity of these frequency distributions, particularly the frequency distribution of stress patterns, these frequency distributions are compared against the frequency distributions obtained from the other referred sources, that is, the Hans Wehr English-Arabic dictionary and the textbooks of the *Al-Kitaab* curriculum. Precisely, a correlation will be made between the frequency dictionary and both the Hans Wehr Arabic-English dictionary and the frequency dictionary and the textbook.

Before proceeding to the first comparison, it must be noted that the frequency distribution obtained from the frequency distribution and the Hans Wehr dictionary represents different type of frequency distributions. As mentioned in the previous sections, each entry in the frequency dictionary comes with the raw frequency of the entry. By merging the frequency count of entries that have the same stress patterns, the frequency distribution obtained is a distribution of token frequency, i.e., the same word can be repeatedly counted as long as it occurs in a different part of the corpus. However, as in most bilingual dictionary, frequency information is rarely included in the entries, and the Hans Wehr dictionary is no exception. Therefore, by merging the entries that have the same stress pattern, the frequency distribution obtained is a distribution of type frequency, which means that the same word is only counted once even though it might have appeared in other parts of the dictionary.

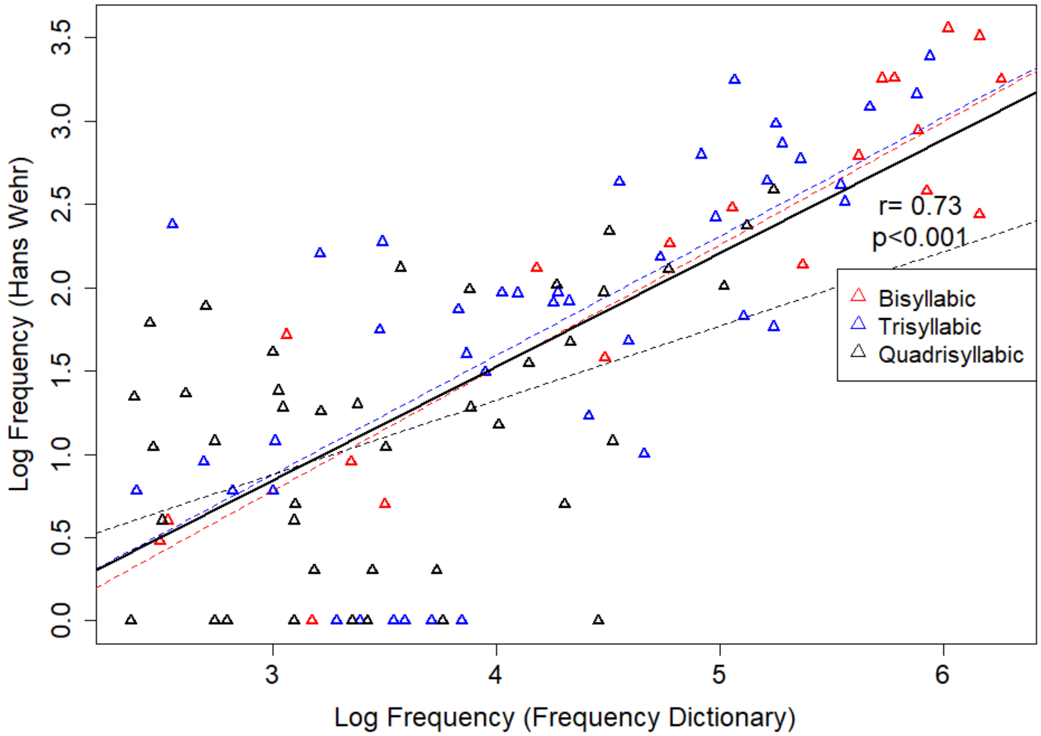
The motivation for comparing the frequency dictionary and the Hans Wehr dictionary is to examine the relation between the type and token frequency of the stress patterns, as different relation between the two types of frequency could have different interpretations of how the frequency distribution is constructed. Precisely, this comparison investigates whether the frequency distribution of the stress patterns is merely informed by the frequency of some specific words, or the distribution is reflective of the general trend for stress patterns in the lexicon of the target language.

3.2.4.1 Frequency Dictionary and Hans Wehr Dictionary

The first step in the comparison is to obtain the frequency distribution of the stress patterns from the Hans Wehr dictionary. As briefly discussed in Section 3.2.3, the data extraction follows the procedure in Figure 3.8, where each entry in the dictionary undergoes stages of transcription to IPA (which is provided in the dictionary), conversion into stress patterns, and lastly merging frequency count (which is 1 for each entry as no raw frequency information is available) with entries that have the same stress pattern. The same procedure is repeated for all the entries in the dictionary and the frequency distribution of stress pattern is obtained after the process.

The result of the correlation between the frequency distribution obtained from the frequency dictionary and the Hans Wehr dictionary is presented in a log coordinate in Figure 3.19. Each data point presented in the figure represents a stress pattern, with its log frequency obtained from the frequency dictionary on the x axis and its corresponding log frequency found in the Hans Wehr dictionary on the y axis. Stress patterns in words with different numbers of syllables are plotted in difference colors, with bisyllabic stress patterns in red, trisyllabic stress patterns in blue and quadrisyllabic stress patterns in black. The dashed trendlines follow this coloring schema, whereas the trendline for all stress patterns regardless of number of syllable is presented in black solid line.

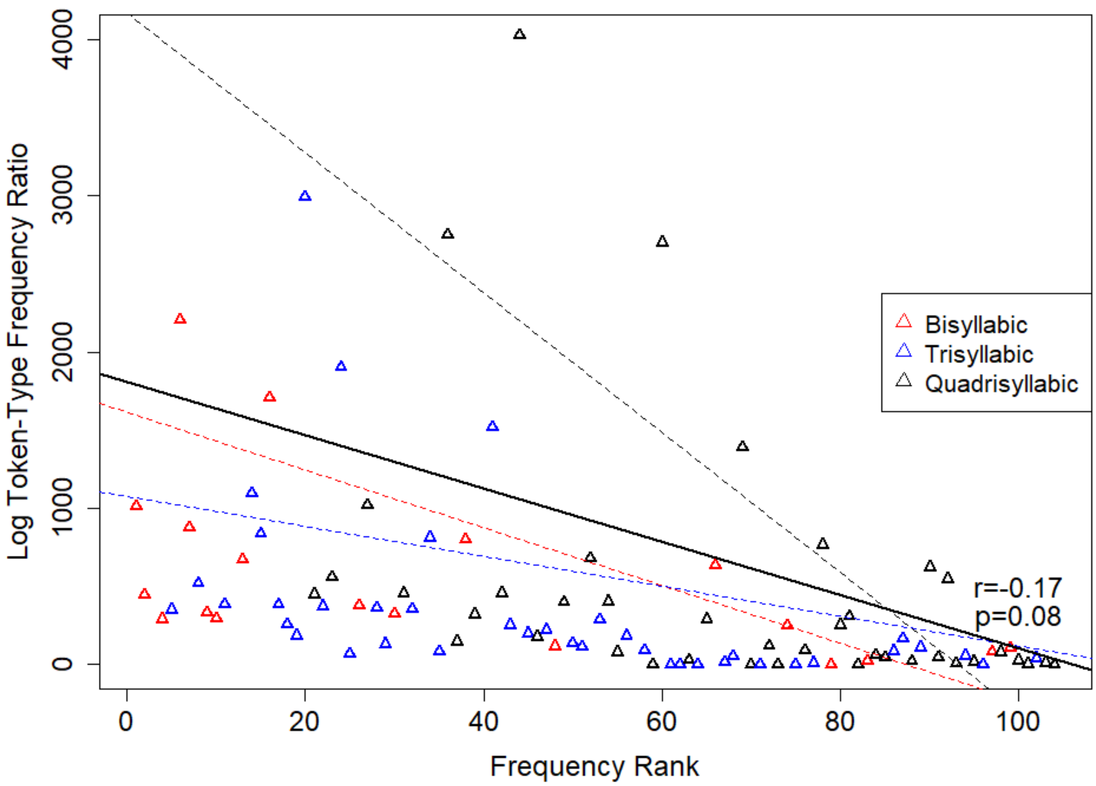
Figure 3.19 Correlation between Frequency Dictionary and Hans Wehr Dictionary



The results reveal a general positive correlation between the frequency distribution of stress patterns obtained from the frequency dictionary and the Hans Wehr dictionary. Precisely, a significant positive correlation was found for all stress patterns (Pearson's r = .73, n= 104, p<0.001), bisyllabic (Pearson's r = .69, n= 41, p<0.001), trisyllabic (Pearson's r = .73, n= 104, p<0.001) and quadrisyllabic (Pearson's r = .45, n= 42, p<0.001) stress patterns. This positive correlation between token and type frequency suggests that when a stress pattern occurs frequently in the corpus, this stress pattern is also found to be shared by many lexical items in the dictionary; conversely, if a stress pattern is identified as less frequent from the frequency dictionary, the stress pattern is less likely to be shared by many lexical items in the Hans Wehr dictionary.

Another interesting finding from the comparison between the two sources is the correlation between the token-to-type ratio and the frequency rank of the stress patterns. Figure 3.20 presents this analysis, where the x axis presents the frequency rank (the higher the rank, the more frequent the stress pattern is) of the stress patterns and y axis shows the token-to-type-frequency ratio (calculated by dividing the raw frequency of a given stress pattern with the type frequency of the stress pattern) corresponding to respective stress patterns. The results show that although there is an apparent negative correlation between the token-type frequency ratio of a stress pattern and its frequency rank, this negative correlation is generally weak and is not statistically significant for most stress patterns. Precisely, the correlation was not found to be significant for trisyllabic stress patterns (Pearson's r = -0.19, n= 41, p=0.21), quadrisyllabic stress patterns (Pearson's r = -0.24, n= 42, p=0.12), and across all stress patterns (Pearson's r = -0.17, n= 104, p=0.08); only bisyllabic stress patterns were found to have a significant negative correlation (Pearson's r =-0.43, n= 21, p<0.05). In other words, the results suggest that stress patterns that are more frequent, i.e. higher in frequency rank, do not necessarily have a higher token-type frequency ratio, which indicates that a frequent stress pattern involves many moderately frequent lexical items that share this stress pattern, instead of by few highly-frequent lexical items. This outcome suggests that when a stress pattern is common in the lexicon of a language, there is a tendency for words with this stress pattern to also be frequently used in real life. However, this finding does not seem to hold for bisyllabic stress patterns, a strong negative correlation was found, which would suggest that some frequent stress patterns are indeed dominated by a few highly frequent lexical items.

Figure 3.20 Correlation between Token-Type Frequency Ratio and Frequency Rank



The comparison thus far has shown convergence between the frequency distributions obtained from the frequency dictionary and the Hans Wehr Arabic-English dictionary, as frequent stress patterns in one dictionary are often found to be frequent in the other. This positive correlation between the two sources supports the validity of using the frequency dictionary as the main reference for the frequency information for stress patterns in Arabic.

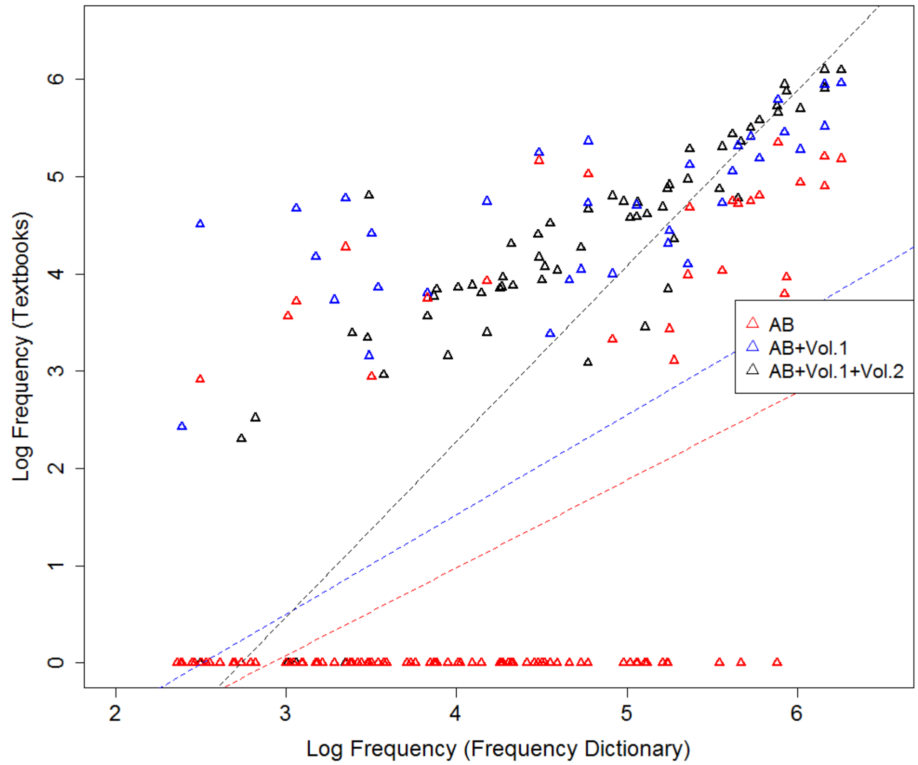
3.2.4.2 Frequency Dictionary and Textbooks

The second comparison is between the frequency dictionary and the textbooks used by the learners. The purpose of this comparison is two-fold. The first purpose is to validate the frequency information obtained from the frequency dictionary with another source of language input that is crucial to the learners ─ the textbooks. This comparison examines the degree of similarity in terms of the frequency distribution of stress patterns between the two sources. Secondly, this comparison aims to capture the progression of language input, particularly the extent to which textbooks contain frequent words throughout the curriculum, since it is not uncommon for textbooks to be criticized for not including enough frequent words, i.e. words characterized as "actually helpful" for the students in their daily lives.

As a demonstration, the three textbooks of the *Al-Kitaab* curriculum are chosen for this comparison based on their popularity for the English-speaking learners of Arabic in the U.S. The three textbooks are (1) *Alif Baa'*, (2) *Al-Kitaab* Part One, and (3) *Al-Kitaab* Part Two, presented in the order in which the textbooks are taught in the curriculum. Normally, *Alif Baa'* and parts of *Al-Kitaab* Part One are completed in the first-year class, the rest of *Al-Kitaab* Part One and parts of *Al-Kitaab* Part Two in the second-year class, and the rest of Al-Kitaab Part Two in the third-year class. As discussed in Section 3.2.2, instead of analyzing the content of the three books, the analysis focuses on the vocabulary lists that are provided in the beginning of each lesson. The frequency of the vocabulary is extrapolated from the frequency dictionary, the frequency distribution of stress patterns is extracted from the textbooks following the steps presented in Figure 3.8.

The comparison involves comparing the frequency distribution from the dictionary against the frequency distributions that are accumulated over the course of the curriculum. Precisely, the frequency dictionary is compared against the frequency distribution obtained from (1) *Alif Baa'* (AB in red) only (2) *Alif Baa'*+ *Al-Kitaab* Part One (Pt1in blue) and (3) *Alif Baa'*+ *Al-Kitaab* Part One + *Al-Kitaab* part two (Pt2 in black). In Figure 3.21, each data point presents a stress pattern with its log frequency obtained from the frequency dictionary on x axis and the corresponding log frequency obtained from the textbook on the y axis.

Figure 3.21 Correlation between Frequency Dictionary and Al-Kitaab Textbooks



The results of the comparison present a general trend of positive correlation between the frequency distribution found in the dictionary and the textbooks. A significant positive correlation was found between the frequency dictionary and *Alif Baa'* (Pearson's r =0.52, n= 104, p<0.001), Alif Baa' plus Al-Kitaab Part One (Pearson's r =0.50, n= 104, p<0.001), and all textbooks (Pearson's r =0.84, n= 104, p<0.001). In other words, when a stress pattern is frequent in the frequency dictionary, the stress pattern is also likely to be frequent in the textbooks. Despite the already-strong correlation with the *Alif Baa'* (r=0.50), this trend of positive correlation becomes even more prominent for learners who are more advanced into the curriculum, as an even stronger positive correlation (r= 0.84) was found when the dictionary is compared against all textbooks.

Another interesting finding is that some stress patterns do not seem to be available to the learners in the earlier stage of the curriculum. As shown in Figure 3.21, a cluster of data points in red are found at the bottom of the plot on the x axis, corresponding to a y value of zero. The data points near the x axis mean that the stress patterns do not have a corresponding frequency value in the textbook, as the stress pattern is not available in the source. However, this trend seems to disappear as the curriculum progresses, since not many blue or black data points are found near the x axis. In other words, the stress patterns that were not available in the earlier stage of the curriculum, i.e. in *Alif Baa'* become available to the learners.

In sum, the frequency distributions of stress patterns obtained from the textbooks closely align with that of the frequency dictionary. As discussed in the previous section, the textbooks are not only one of the main sources from which the learners receive language input but they generally anchor the range of language input that the learners receive in class. Having a higher degree of similarity manifested as high correlation means that the frequency distribution extracted from the frequency dictionary to a great extent reflect the frequency distribution of stress patterns used in the language classroom, which in turn support the use of frequency dictionary as the reference for assessing frequency information.

To conclude, the two sources, i.e. the Hans Wehr Arabic-English dictionary and *Al-Kitaab* textbooks, examined in this section support the validity of the frequency distribution extracted from the frequency dictionary and therefore supports using the dictionary as the main reference for assessing frequency information for Arabic. In the two comparisons conducted for data validation, a strong positive correlation was found between the frequency dictionary and the two sources, suggesting that the frequency distribution obtained from the frequency dictionary is reflective of the frequency distribution found in the lexicon of the language, and the language used in the language classroom. Thus, although it cannot be said that the data obtained from the frequency dictionary is completely reflective of the actual language use, the frequency information extracted from the data is, to a reasonable degree, adequately representative of the language encountered by the learners, at least for the purpose of stimulus design for the experiment.

3.2.5 Stimulus Organization

As concluded in the last section that the frequency of the relevant structures for stress in Arabic is variable, this section discusses the way in which the variability in frequency can be controlled and manipulated, as well as other considerations pertaining to stimulus design.

3.2.5.1 Real V.S. Nonsense Words

The first consideration that needs to be discussed is the general composition of the stimuli, that is, whether to include only nonsense words or to include real words as well, as the stimuli in question will be used across all tasks of the experiment. On one hand, the motivation for using only nonsense words as stimuli is the benefit of eliminating any familiarity effect that could potentially make the participants respond either faster or slower. On the other hand, having stimuli consisting of only nonsense words could deprive the sense of authenticity from the participants and unnecessarily induce uneasiness as they will not know any of the stimuli since they are all made-up words, which could in turn influence their performance.

To strike a better balance between the two approaches, the present study adopted a stimuli schema that consists of half nonsense words and half real words at the time of the development. The motivation for including real words as much as half of the stimuli is to mask the unnaturalness of the nonsense words and to promote the authenticity of experiment as a whole; that is, one that tests the participants' ability to perceive and produce "real" Arabic words. However, these real words will only serve as fillers and will not be further analyzed due the potential bias from the familiarity effect mentioned earlier.

3.2.5.2 Variable Manipulation

The second consideration for the stimulus design concerns the inclusion and exclusion of variables that were previously identified and are intrinsic to the stimuli. As concluded from the analyses in Section 3.2.3, the four relevant structures for stress are variable in frequency, which led to the decision to either control for or manipulate them. These variables are the frequency of (1) stress pattern, (2) position, (3) syllabic structure and (4) the conditional probability of the stressed syllable. To simplify these variables, each individual variable is treated as binary, as either "relatively frequent" (henceforth "frequent") or "relatively infrequent" (henceforth "infrequent"). That is, for any given stress pattern, the stress pattern could either be frequent or infrequent, depending on the other stress patterns that it is compared against. For example, a stress pattern [CV.ˈCVV.CV] could be frequent as a stress pattern, but the stress position of the stress pattern is relative infrequent; conversely, a stress pattern such as [ˈCV.CV.CVC] could be infrequent as a stress pattern, but have stress falling on a frequent position. By converting the absolute frequency count into binary values, variable manipulation could be more easily achieved.

The other variable that is intrinsic to the stimuli is the number of syllables of each stimulus. The analyses conducted for the frequency dictionary show that the majority of stress patterns are for either bisyllabic, trisyllabic and quadrisyllabic words. Among these stress patterns, the bisyllabic stress patterns are found to have the least variety, as only 21 types of stress patterns are identified as opposed to 41 and 42 types for the trisyllabic and quadrisyllabic stress patterns respectively. Besides, mathematically, the bisyllabic stress patterns have the highest likelihood of being correct if participants were to guess randomly. That is, for bisyllabic stress patterns, there is a 50% chance for that participants would correctly respond to the stimuli by random guessing as opposed to approximately 33% for the trisyllabic stress patterns and 25% for the quadrisyllabic stress patterns. Besides, bisyllabic stress patterns are generally more frequent and have higher variance in log frequency ($s=1.26$) than trisyllabic (s=0.92) and quadrisyllabic (s=0.80) stress patterns, Due to the relatively higher frequency, the bisyllabic stress patterns might be more familiar to the participants, which allow faster retrieval for perception and production than the trisyllabic and quadrisyllabic stress patterns. Together with its higher expected value for correctness, it was decided to exclude bisyllabic stress patterns from the stimuli and to focus on trisyllabic and quadrisyllabic stress patterns only.

Thus far, a total of five binary variables have been identified, which would theoretically require 2^5=32 tokens in a block to exhaustively manipulate these variables. Considering that the real-nonsense word design adopted in the study, it would require 2^6= 64 token in each block of the stimuli which, given the need for additional blocks of stimuli, would make the task overly long. Therefore, to keep stimulus set a manageable size with an adequate number of blocks, the number of manipulated variables has to be limited. Since the number of syllables and the real-nonsense contrast are not subject to adjustment, the frequency-informed variables have to be selectively controlled for. As a result, the frequency of the stress pattern and the position of the stressed syllable were manipulated while the syllabic structure for the stressed syllable as well as its conditional probability was held comparable in the stimulus design . With the frequency of the stress pattern and stress position manipulated, in addition to the number of syllable and the real-nonsense word contrast, the size of each block is effectively reduced to $2^4 = 16$ tokens.

The schema of variable manipulation as well as the selected stress patterns that correspond to the respective variable setting are presented in Table 3.1. The schema consists of two parts, one for trisyllabic stress patterns and the other for quadrisyllabic one. Within each part, each stress pattern is contrastive, be it the frequency (presented as "+" in the table) of the stress pattern or the frequency of the position of the stressed syllable while other frequency-informed variables such as the syllabic structure of the syllable as well as its conditional probability are controlled for. For example, the trisyllabic stress pattern [ˈCV.CV.CV], is a relatively frequent stress pattern (log frequency = 5.9) with stress falling in a relatively frequent position, i.e. on the antepenultimate position, compared with all other trisyllabic stress patterns identified from the frequency dictionary. In contrast, although [ˈCV.CV.CVC] has the stress in the frequent position for trisyllabic words, i.e. antepenultimate, the stress pattern itself is not a frequent one (log frequency = 3.2 and there denoted with a "-" sign) compared to other trisyllabic patterns. As for the stress pattern [CV.ˈCVV.CV], it has an overall frequent stress pattern (log frequency of 5.8), but the position of the stressed syllable is infrequent, i.e. on the penultimate syllable. Lastly, the stress pattern [CVC.ˈCVC.CVC] is relatively infrequent stress pattern (log frequency = 4.0) with stress falling on a relatively infrequent position, i.e. on the penultimate position. The same configuration applies to the quadrisyllabic stress patterns that were selected for constructing the stimuli. The detailed procedure of constructing nonsense words from these stress patterns will be discussed in Section 3.2.5. More importantly, the real words used in the stimuli also follow the eight types of stress patterns laid out in Table 3.1. These real words are randomly selected from the frequency dictionary and real words with the same root are generally avoided

Table 3.1 Schema of Variable Manipulation

	Stress Pattern	Frequency Stress Pattern	Frequency Stress Position	Frequency Syllabic Structure	Conditional Probability
Trisyllabic	ˈCV.CV.CV	**+ (5.9)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
	ˈCV.CV.CVC	**- (3.2)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
	CV.ˈCVV.CV	**+ (5.8)**	**- (5.9)**	**+ (6.1)**	**+ (3.3)**
	CVC.ˈCVC.CVC	**- (4.0)**	**- (5.9)**	**+ (6.1)**	**+ (2.5)**
Quadrisyllabic	CVC.ˈCV.CV.CV	**+ (5.0)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
	CVC.ˈCV.CV.CVC	**- (2.6)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
	CVC.CV.ˈCVV.CV	**+ (4.7)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
	CV.CV.ˈCVV.CVC	**- (3.8)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**

Plus sign (**+**) = Frequent, Minus Sign (**-**) = Infrequent, (): Log Frequency

3.2.5.3 Stimulus Blocking, Randomization and Presentation

The last consideration for stimulus design concerns the blocking of the stimuli and the randomization process that the stimuli undergo in the experiment. Firstly, as laid out in the previous section, each block of stimuli consists of 16 tokens, with each token contrasting in the frequency of its stress pattern (frequent v.s. infrequent), stress position (frequent v.s. infrequent), number of syllable (trisyllabic v.s. quadrisyllabic), naturalness (real v.s. nonsense). To keep the duration of the experiment to under one hour, it was decided to include 8 blocks of stimuli in the experiment, amounting to a total of 128 tokens (8 blocks times 16 token per block). As a post-experimental note, the duration of completion for the entire experiment averages to approximately 50 minutes, which aligns closely with the estimated duration of completed. The root pattern of nonsense words in each block varies from block to block, allowing the experiment to examine the participants' performance in stress perception and production under various consonant conditions. The approaches to selecting the root patterns for the nonsense words will be discussed in Section 3.2.5.

To avoid the priming effect that might stem from the co-occurrence of stimuli that are constructed from the same root pattern or from repeated appearance of stimuli that have the same stress pattern, measures of pre-blocking and randomization must be taken. For the pre-blocking, it was ensured that each block contains (1) the same number of real and nonsense words, and that (2) every word, be it real or nonsense, must have a different root within the block, and that (3) both real and nonsense words must contain all stress patterns. The blocking was controlled automatically by the computer. The randomization process was controlled both locally and globally. Locally, the tokens for each block are randomized; globally, the order in which each block is presented in each task is randomized as well. Similarly, the randomization is not conducted manually; instead, it is conducted via the internal setting in the PsychoPy program.

The same set of stimuli was used for the perception and the production tasks. The motivation for using the same set of stimuli is the attempt to establish a link between participants' performance in the perception and the production tasks. The interval between the stress production task and the stress identification task, where a short introduction of lexical stress is presented in English to help to minimize the influence of the earlier tasks on the later one. Together with randomization of the stimuli as well as lexical decision task serving as a filler in the stress identification task, the participants are faced with many diversions from the abovementioned measures and are less likely to perceive the stimuli to be the same. As a post-experiment note, when asked about their general feeling towards the stimuli in the short exchange with the participants after the experiment, none of the participants seemed to have noticed that the same set of the stimuli was presented in the tasks.

In sum, this section addresses several considerations that are central to stimulus design, such as the inclusion of real words in the stimuli, the manipulation of frequency and stimuli randomization and presentation. The stimulus design included both real and nonsense words with the former being treated as fillers. With respect to the array of frequency-related variables pertaining to the stimuli, only the frequency of stress pattern and the position are manipulated while the others are controlled for. Based on the selected variables, eight contrastive stress patterns were identified as the template for constructing nonsense words. Measures of pre-blocking and randomization were employed by computer program to control for any priming effect. The remaining step of stimulus design is the construction of nonsense words, which will be discussed in the following section.

3.2.6 Stimulus Construction

3.2.6.1 Main Tasks

A total 8 stress patterns were used as the template to construct nonsense words. The purpose of this section is to discuss the principles and considerations that are essential to create nonsense words for the experiment or in Arabic in general. These considerations are laid out as the following questions.

1. What constitutes a nonsense word in Arabic?
2. What are the ways to construct nonsense words?
3. What type of nonsense words should be included in the experiment?

To determine what constitute a nonsense word in Arabic, it is essential to understand first what constitutes a word. Arguably, the fundamental constituents for most Arabic words is the root and the consonantal-vocalic pattern (henceforth a "template"). A root consists of a sequence of consonants (henceforth the term "root consonants" is used interchangeably). The number of consonants in a root could vary but a tri-literal root, one that consists of three consonants, is the most frequent type of root. On the other hand, the consonantal-vocalic pattern is a template that regulates the vocalic context of the word as well as the respective position of for the root consonants. In Arabic, a word is derived by combining a root with a template. For example, the tri-literal root [ktb], which has a prototypical meaning of "write/writer", can be combined with the template [ˈC_1a:.C_2ɪC_3], which has the meaning of "the one who …" or "…er" (Alhawary, 2011). To combine the root with the template, each consonant of the root should be placed into the template by the position assigned for the consonants of the root in order. That is, the first consonant of the root [k] should be placed in the position for the first consonant C_1, the second consonant of the root [t] in the position for the second consonant C_2, then third consonant in the root [k] in the position for the third consonant C_3. After the insertion, the root and the vocalic pattern derive the word [ˈka:.tɪb], which means "the one who writes" or "writer". In the same vein, with the root [qrʔ] "to read" with the same vocalic pattern [ˈC_1a:.C_2ɪC_3], the derived word [qa:.rɪʔ] has the meaning of "the one who reads" or "reader".

However, not all consonants can be freely combined to become a root in Arabic. There are some phonotactic constraints blocking certain combinations of consonants. One well documented constraint is the homorganic consonant, which restricts the co-occurrence of consonants with similar place of articulation within the root (Greenberg 1950). This constraint is also sensitive to the position of the root. For example, the first and the second consonant of the root cannot be the same consonant; as a result, roots such as [mmd] are never attested in Arabic. Conversely, having the same consonant as the second and the third consonant of the root is allowed, which makes a root such as [mdd] possible in Arabic.

This homorganic constraint helps us to understand what constitutes a real Arabic, in that real Arabic words are ones derived from attested roots and attested vocalic patterns. Conversely, Arabic nonsense words are necessarily derived from either unattested roots or unattested template. However, unattested templates do not consistently derive nonsense words, as some templates could derive nonsense words with some specific roots but not with the others. For instance, the template [ˈC_1ɪː.C_2aC_3] could derive a nonsense word when combined with the root [ktb] as [ˈkɪ.tab]; nonetheless, when combined with the root [qym], a real word [ˈqɪ.yam] "values" is derived. Due to this inconsistency in deriving nonsense words, the constructing nonsense words from manipulating the template did not seem to be an effective approach. Thus, the present study pursues the other approach to constructing nonsense words and uses unattested roots.

Unlike implementing an unattested template, an unattested root could always derive a word that is not attested in the language, i.e. a nonsense word. The aforementioned homorganic constraint is instructive in locating unattested roots. The homorganic constraint implies that the less frequently two consonants co-occur in the root, the less likely for the two consonants to constitute an attested root and accordingly a real word. Therefore, the task becomes to identify consonants are less likely to co-occur. To this end, the roots included in Lane's Arabic-English dictionary (Lane 2003) are analyzed. The reason for using the Lane dictionary rather than the Hans Wehr dictionary is due to its accessibility, since the root information in the Lane's dictionary is readily available electronically, which could simplify the analysis and accelerate the processing time. Additionally, since most of the roots is tri-literal and the size and the type of the tri-literal roots are rather stable over time, root information in this slightly dated dictionary shall suffice the purpose of identifying triliteral roots.

Appendix 3.7 gives the type of co-occurrence of the two consonants as well as its conditional probability; this is based on the 1573 triliteral roots identified from the dictionary. To illustrate, Table 3.2 is the partial table from the Appendix 3.7, presenting the various types of co-occurrence of root consonants. The type co-occurrence is presented in a conditional probability statement, in that co-occurrence type "$R_2|R_1$" should be construed as the probability of having the root consonant R_2 given that the preceding root consonant is R_1 in a root. For example, the first type of co-occurrence "r|b" is understood as having a [r] in the root given that the preceding root consonant is [b], i.e. the sequence of [br] in the root. This type of co-occurrence is relatively frequent in the 1573 roots examined, and the likelihood of having this sequence is 8.9%. On the other hand, the co-occurrence types "dʒ|b" and "ʃ|b" are comparatively less frequent, as only one instance of [bdʒ] sequence or [bʃ] sequence is found among the 1573 roots included in the dictionary. For the purpose of nonsense word construction, infrequent root consonant sequences as such are essential, as a word that is derived from a root with an infrequent root consonant sequence such as [bʃ] is more likely be a nonsense word than one that that is derived from a root with a frequent root consonant sequence such as [br]. In other words, nonsense words could be derived from roots that contain infrequent consonant root sequence, and to derive such roots, we could either replace a root consonant in the root with one that results in the lowest conditional probability. Alternatively, we could replace the entire root with one that consists of infrequent co-occurrences of consonant roots as well.

Table 3.2 Type of Co-Occurrence of Root Consonants and Its Conditional Probability (Partial)

Type of Co-Occurrence	Frequency Count	Conditional Probability
r|b	26	8.9%
l|b	14	4.8%
b|b	12	4.1%
dʒ|b	1	0.3%
ʃ|b	1	0.3%

Although the two approaches can both derive nonsense words with confidence, the underlying difference between the two approaches lies in the degree of naturalness. By replacing only one root consonant, the derived nonsense word would bear more resemblance to the original word, which in turn is more natural. On the other hand, if the entire root is swapped for one consisting of sequences of root consonants that are less frequently attested to co-occur, the derived nonsense word could be less natural due to the composition of the root; nonetheless, this approach avoids biases introduced by the familiarity effect, as the participants might have different reaction due to the resemblance born by the nonsense words with only one root consonant replaced. Since both approaches have their advantages and shortcomings and to control for the biases of only adopting one of the two approaches, it was decided to adopt both approaches to construct the nonsense words. That is, the stimuli include two types of nonsense words, half of which will be derived from replacing one root consonant in an existing word with one that infrequently co-occur with the other root consonants in the word, especially with the preceding root consonant. The other half of the nonsense words will be constructed from a controlled templated with unattested roots that are constructed from sequences of root consonants with infrequent co-occurrence.

The first type of nonsense words could be easily derived from the real words included in the stimuli, as these real words are selected based on the stress patterns identified for the nonsense words. Table 3.3 presents a block of nonsense words as well as the corresponding real words. As shown in Table 3.3, each row represents a minimal pair of stimuli that contrast in the its naturalness. For each example, the first row shows the nonsense word نَغَرَ [ˈna.ɣa.ra] that conforms to the [ˈCV.CV.CV] stress pattern. This nonsense word is derived from the real word نَشَرَ [ˈna.ʃa.ra] "to spread" by replacing the second root consonant [ʃ] with [ɣ] since the root consonant sequence [nɣ] has a relatively low probability as indicated by results of the previous analyses of the Lane dictionary. The same process is followed for words that are either trisyllabic and quadrisyllabic.

Table 3.3 Sample Block of Nonsense Words Derived from Real Word

	NONSENSE				REAL	
	AR script	**Stress Pattern**	**Pronunciation**	**Template**	**AR script**	**Pronunciation**
Trisyllabic	نَغَرَ	ˈCV.CV.CV	ˈna.ɣa.ra	Minimal Pair	نَشَرَ	ˈna.ʃa.ra
	طَثَعَتْ	ˈCV.CV.CVC	ˈtˤa.θa.ʕat	Minimal Pair	طَلَعَتْ	ˈtˤa.la.ʕat
	خِلازَ	CV.ˈCVV.CV	xi.ˈla:.za	Minimal Pair	خِلالَ	xi.ˈla:.la
	مُسْتَخْزَم	CVC.ˈCVC.CVC	mus.ˈtax.zam	Minimal Pair	مُسْتَخْدَم	mus.ˈtax.dam
Quadrisyllabic	مُتَظانِر	CV.CV.ˈCVV.CVC	mu.ta.ˈðˤa:.nir	Minimal Pair	مُتَظاهِر	mu.ta.ˈðˤa:.hir
	اِكْتَشَزَ	CVC.ˈCV.CV.CV	ʔik.ˈta.ʃa.za	Minimal Pair	اِكْتَشَفَ	ʔik.ˈta.ʃa.fa
	اِنْهَلَعَت	CVC.ˈCV.CV.CVC	ʔin.ˈha.la.ʕat	Minimal Pair	اِنْدَلَعَت	ʔin.ˈda.fa.ʕat
	اِسْتِقاخَة	CVC.CV.ˈCVV.CV	ʔis.ti.ˈqa:.xa	Minimal Pair	اِسْتِقالَة	ʔis.ti.ˈqa:.la

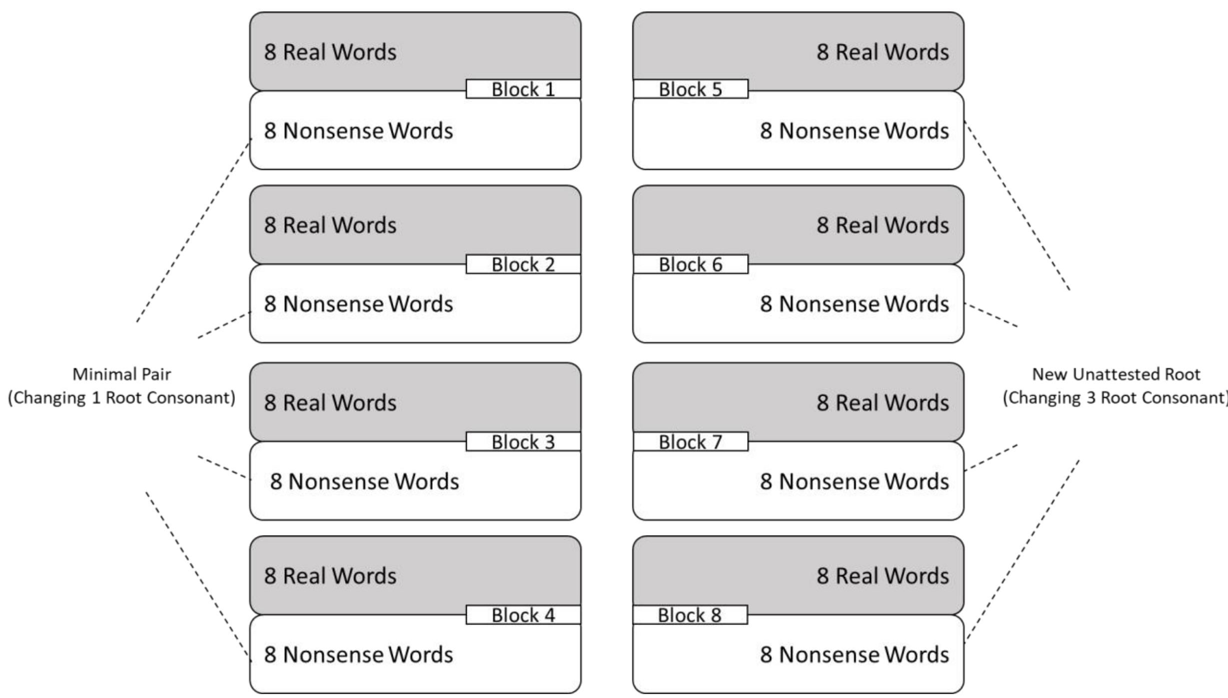
Nonsense words of the second type are derived from an unattested root with a predetermined template for each stress pattern. First, the template for each is selected from the most frequent template for the stress pattern in the frequency dictionary. The unattested root for the nonsense words is identified by referring to the conditional probability table for the root consonants presented in Appendix 3.7. For example, to identify the nonsense root [fqz], the procedure starts by randomly selecting a consonant, which is [f] in the case of [fqz]. Accordingly, the conditional probability table is referred to determine which root consonant is the least likely to occur, given that [f] is the preceding root consonant, which is [q]. The same process was followed to see which consonant is least likely to co-occur with [q], which gives [z]; hence, the nonsense root [fqz] is derived. The unattested root [fqz] is then used to combine with the template to form stimulus for respective stress pattern in the block. The same procedure was followed to identifyd the other three unattested roots that are used for constructing the remaining nonsense words. Table 3.4 presents a sample block of nonsense words constructed using the unattested root [fqz] along with real words with corresponding stress pattern but not necessarily with the same template. For example, for the stress pattern [ˈCV.CV.CV], the nonsense word فَقَزَ [ˈfa.qa.za] is constructed by inserting each consonant in the root [fqz] into the template [ˈC1a.C2a.C3a]. The corresponding real word وَثِقَ [ˈwa.θi.qa] "to trust" shares the same stress pattern [ˈCV.CV.CV], but differs in the template, which is [ˈC1a.C2i.C3a].

Table 3.4 Sample Block of Nonsense Words Derived from Unattested Root [Fqz]

	NONSENSE				REAL	
	AR script	**Stress Pattern**	**Pronunciation**	**Template**	**AR script**	**Pronunciation**
Trisyllabic	فَقَزَ	ˈCV.CV.CV	ˈfa.qa.za	ˈC_1a.C_2a.C_3a	وَثِقَ	ˈwa.θi.qa
	فَقَزَت	ˈCV.CV.CVC	ˈfa.qa.zat	ˈC_1a.C_2a.C_3at	جَعَلَت	ˈdʒa.ʕa.lat
	فِقَازَ	CV.ˈCVV.CV	fi.ˈqa:.za	C_1i.ˈC_2a:.C_3a	نَتِيجَة	na.ˈti:.dʒa
	مُسْتَفقَزْ	CVC.ˈCVC.CVC	mus.ˈtaf.qaz	mus.ˈtaC_1.C_2aC_3	مُسْتَوْرَد	mus.ˈtaw.rad
Quadrisyllabic	مُتَفَاقِز	CV.CV.ˈCVV.CVC	mu.ta.ˈfa:.qiz	mu.ta.ˈC_1a:.C_2iC_3	مُتَتالٍ	mu.ta.ˈta:.lin
	اِفتَقَزَ	CVC.ˈCV.CV.CV	ʔif.ˈta.qa.za	ʔiC_1.ˈta.C_2a.C_3a	اِعْتَمَدَ	ʔiʕ.ˈta.ma.da
	اِفتَقَزَتْ	CVC.ˈCV.CV.CVC	ʔif.ˈta.qa.zat	ʔiC_1.ˈta.C_2a.C_3at	اِرْتَكَزَت	ʔir.ˈta.ka.zat
	اِفتِقَازَ	CVC.CV.ˈCVV.CV	ʔif.ti.ˈqa:.za	ʔiC_1.ti.ˈC_2a:.C_3a	اِسْتَجابَ	ʔis.ti.ˈdʒa:.ba

Lastly, all constructed nonsenses words undergo further validation as to whether these nonsense words are indeed unattested. Specifically, nonsense words were checked against two dictionaries (Hans Wehr Arabic-English dictionary and Mawrid dictionary) and none was found, which further confirms their status as nonsense words. As mentioned in previous section, the experiment contains 8 blocks of stimuli totaling 128 tokens, half of which consists of real words with the other half being nonsense words. Because there are two types of nonsense words, this means that the 4 blocks of stimuli contained nonsense words derived using real words with corresponding stress patterns, while the other 4 blocks of stimuli contained nonsense words derived from unattested roots that are constructed from scratch with predetermined templates that correspond to the eight stress patterns identified for the experiment. Figure 3.22 reflects this configuration of stimuli and the complete list of stimuli is presented in Appendix 3.8.

Figure 3.22 Stimuli Configuration



3.2.6.2 Warm-Up Activity

Compared to the stimulus design for the main tasks, the stimulus development for the warm-up activity, i.e. the speed reading drill, is relatively straightforward. As a reminder, in the speed reading drill, the participants are presented with a list of real Arabic words and are required to read aloud the words as quickly as possible. The main purpose of this activity is to acquaint the participants with the testing atmosphere as well as to help the participants enter the mode of using the target language. Therefore, the task should be mechanical and manageable and yet somehow related the modality of the next task, which is production.

To achieve this goal, words that correspond to the 104 stress patterns identified from the frequency dictionary were selected from that dictionary. The 104 words could not all be included in the experiment, as it would significantly increase the duration of the activity. In order to keep the activity within a manageable duration, preferably within 3 minutes, the number of words included in the list were reduced to 48 words. The motivation for including 48 words is because the estimated time for completing each word is approximately 5 seconds for the beginners, which requires approximately 200 seconds or 3 minutes. Besides, the number of words that could be accommodated in each row on the testing devices is approximately 8 words per row for a total of six rows without compromising readability. As a post-experiment note, the duration of completion for the warm-up activity fell around the 3 minutes target even for the participants who are in the beginning level of Arabic class.

The 48 words on the wordlist for the warm-up activity are mainly trisyllabic and quadrisyllabic words as the stimuli used in main task are trisyllabic and quadrisyllabic as well. Each individual word represents a stress pattern identified in the frequency dictionary with varying roots. These words were randomly selected by the computer to avoid biased introduced by manual selection. The full list of words used for the practice phase as well as the testing phase is given in Appendix 3.2.

3.3. Participants

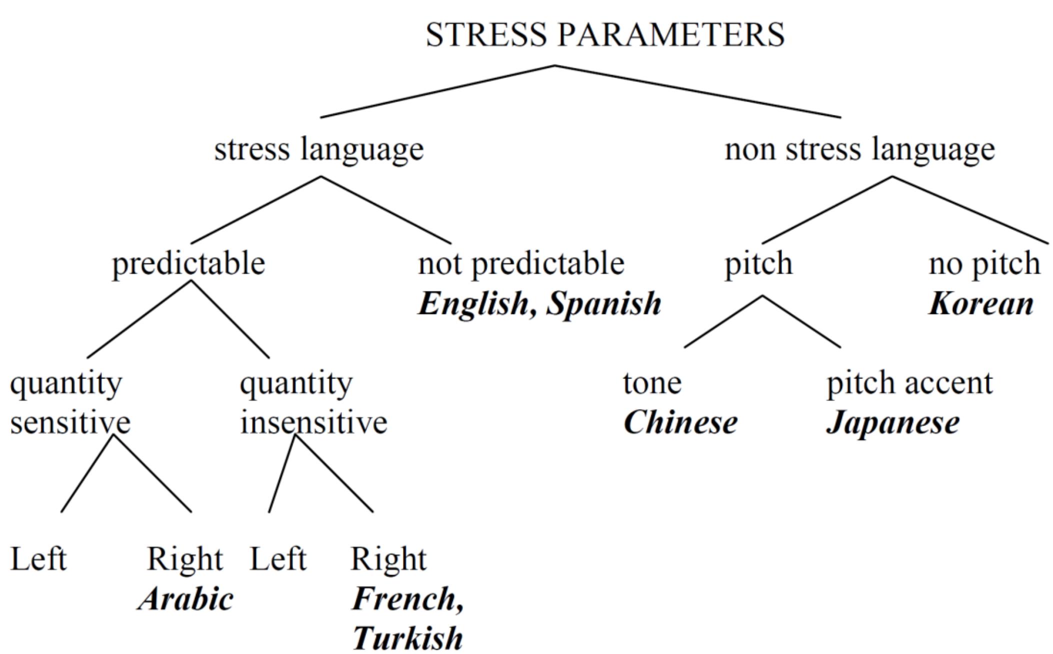
To examine the effect of learner's native language on stress perception and production (RQ1), two groups of learners of Arabic were recruited to participate in the experiment. One group consisted of native speakers of English who were studying Arabic in the U.S. while the other group consisted of native speakers of Mandarin Chinese (henceforth Chinese) who were studying Arabic in Taiwan. Additionally, to investigate the role of proficiency in stress perception and production (RQ 2), the experiment recruited learners of three levels of proficiency to capture the learners' ability in stress perception and production in different stages of language acquisition. Precisely, based on their level of proficiency, i.e. the level of the class in which they were enrolled, the recruited learners were further divided into three subgroups: the first-year group, the second-year group and the third-year group. However, due to the limited number of participants in the third-year group, the data collected was not large enough to proceed with any statistical analyses and therefore are not reported. Lastly, as the control for the experiment, a group of native speakers of Arabic was recruited to participate in the experiment. Incentive was given out for all participants who completed the experiment. Detailed information regarding the two experimental groups and the control group is presented in Section 3.2.3.

3.2.1 Recruitment Criteria

3.2.1.1 Native Language

The motivation for recruiting learners who were native speakers of English or Mandarin Chinese was primarily to allow better examination of the effect of native language on speakers' performance on stress since English and Chinese are typologically distinctive from each other and from the target language Arabic in terms of the representation of stress. This typological difference is demonstrated by the Stress Typology Model (Altmann 2006) in Figure 3.23, which is a binary tree that classifies languages based on the stress parameters and prosodic features of the languages. For the languages of interest in this study (i.e., Arabic, English and Chinese), the model demonstrates clear distinctions between the three languages as each is located in a different end of the binary tree.

Figure 3.23 Stress Typology Model



As discussed in Chapter 2, Arabic is a stress language but lexical stress in Arabic is not phonemic; that is, the position of stress does not change the meaning or the part of speech of a word. Additionally, stress in Arabic is quantity-sensitive, which means that the position of the stressed syllable is completely predictable by following the algorithms for respective varieties of Arabic. However, this is not the case for English. Despite being a stress language, lexical stress in English is rather different from Arabic, as lexical stress in English is phonemic and is not quantity-sensitive. This quantity-insensitive property means that the stressed syllable of a word cannot be determined based on the relative quantity/weight of the syllables and is therefore not completely predicable as in Arabic. It worth noting that the stress position of some specific words in English could be somehow predictable based on the part of speech, such as <u>*re*</u>cord (n.) v.s re<u>*cord*</u> (v.), but not on the quantity/weight of the syllables in the word.

Although the position of the stressed syllable in an English word cannot be confidently determined by solely comparing the weight of its syllables, studies have identified some tendencies for the position of stress. Firstly, it was reported that English words tend not to have stress on the last syllable of a multisyllabic word, with only 11.6% of the words having stress on the final syllable (Clopper 2002). Secondly, when the final syllable is superheavy (CVCC, CVVC), only 15% of t English words have stress on this superheavy syllable (Domahs et al. 2014), which supports the claim that English is a quantity-insensitive language or otherwise superheavy syllables tend to attract stress in other quantity-sensitive language. Lastly, studies also indicated that initial stress, i.e., stress falling on the first syllable of the word, seems to be the default stress pattern in English, in that 57-69% of polysyllabic words have initial stress (Cutler & Carter 1987) and that 92% of lexically invariant bisyllabic nouns have initial stress (Francis & Kucera 1983).

On the other hand, the prosody system of Mandarin Chinese falls on the other end of the spectrum. As demonstrated in the first bifurcation of the typology model in Figure 3.1, Chinese is generally recognized as a non-stress, tonal language (see Duanmu 2002 for an alternative view on whether Chinese has stress). Being a tonal language means that Chinese utilizes phonemic tone as the prosodic feature on the word level instead of stress and a tone is assigned to each word. Since words are monosyllabic in Chinese, this in turn means that every single syllable is assigned with one of the four phonemic tones. In terms of the frequency distribution of tones, a brief calculation using a wordlist available online indicates that the tones are not evenly distributed across the lexicon. Precisely, the falling tone (i.e. Fourth Tone) has higher frequency than the rest of the tones, accounting for approximately 33%, followed by the high-level tone (i.e. First Tone) and the rising tone (Second Tone) accounting respectively for 25% of occurrence. The dipping tone (i.e. Third Tone) is the least frequent tone, accounting for only 17% of the usage.

3.2.1.2 Proficiency

The second consideration for the participants was their proficiency in Arabic. Although there are various ways (e.g. via a standardized test, or based on the grades that the participants received from class, etc.) to assess the proficiency of the participants in the target language, the present used the level of the Arabic class that the participants were taking at the time of recruitment as the measure for the proficiency level of the participants. Precisely, the participants fell into two proficiency levels: the first-year and the second-year. As the name suggests, the first-year group consists of participants who were enrolled in Arabic classes for first-year learners of Arabic while the second-year group with participants in second-year Arabic class.

The motivation for limiting the recruitment to learners in their first-year and second-year classes. First, as one of the less commonly taught languages, Arabic programs have fewer body of students enrolled in the third and fourth-year Arabic class. Secondly, students enrolled in higher-level class is more probable to have travelled or studied in an Arabic-speaking country (the reason why not enough third-year students participated in the experiment) – a main criterion of recruitment of this study that will be discussed in the following section, which again reduces the body of learners that meet the recruitment criteria. Therefore, to avoid the risk of recruiting third and fourth year students who might not eventually constitute a statistically sound sample size, the recruitment focused on the first-year and the second-year learners of Arabic.

Admittedly, this approach of assessing participants' level of proficiency does not completely reflect the individual variability in proficiency between participants enrolled in the same level of Arabic class. Therefore, it is essential to complement this recruitment criterion with another layer of assessment of participants' proficiency, which is the motivation for the inclusion of the speed reading drill (see Section 3.1.1) in the experiment design. This approach allowed several benefits to the experiment as well. Compared to implementing a separate assessment task to evaluate proficiency, using the level of enrollment as the level of proficiency simplified the recruitment process and significantly reduced the duration of the entire experiment. By making the experiment more compact in duration, it relieved the potential fatigue from participants, and might allow better concentration on the tasks in the experiment.

3.2.1.3 Amount of Input

One important consideration with regards to recruitment was the participants' previous exposure to the target language at the time of recruitment. To better examine the effect of frequency on learners' performance on stress perception and production, it is crucial to control for the amount of input received by the participants of a specific level of proficiency. The participants with the same proficiency level were expected to have received more or less similar amount of input of the target language during the course of the learning Arabic.

Thus, two types of learners of Arabic were excluded from participating in the experiment. The first type of learners is heritage speakers of Arabic, that is, learners who learned Arabic at home from their parents in a non-Arabic-speaking context. Although it is not uncommon to have heritage speakers of Arabic studying Arabic at an university where the language of communication is not Arabic, due to their previous experience with Arabic, heritage speakers are likely to have received more language input than their non-heritage peers and are therefore excluded from the recruitment. The second type of learners who were excluded from recruitment were learners who have lived or studied in an Arabic-speaking country for more than one semester/three months. These learners also would be likely to have received more input from the native-speaking learning environment than their counterparts who only study at the foreign-language classroom setting at their home institution.

In other words, the present study focuses on the largest body of learners at a foreign-language classroom setting to maximize the possibility of reaching statistically sound sample size. These learners had not had any exposure to Arabic before taking Arabic classes at their home institution, nor had they lived or studied in an Arabic-speaking country for an extended period of time.

3.2.1.4 Variety of Arabic Taught

Another essential recruitment criterion was for all participants in the experimental group to study the same variety of Arabic at their home institutions. This criterion is particularly relevant for the English-speaking learners since as the Arabic programs in the United States have more varying preferences for the variety of Arabic that is used as the main language of instruction in the classroom than the Chinese-speaking learners. These programs could be roughly divided into three categories: MSA-only programs, MSA-dialect programs and dialect-MSA programs. For the MSA-only programs, as the name suggests, MSA is the only language of instruction and the content of the class is based on MSA only. The MSA-dialect programs, MSA is used as the main language of instruction and the content of instruction is mainly MSA as well with only marginal focus on dialect. Lastly, the dialect-MSA programs use dialect of choice (e.g. mainly Egyptian or Levantine) as the language of instruction and the content of instruction focuses on both the dialect and the MSA at the same time.

In contrast, in Taiwan, because there is only one Arabic program hence there is no varying preference in terms of the variety of Arabic that is taught in the classrooms. The Arabic program in Taiwan falls under the MSA-only category in that the textbook and the language used in class are all MSA-based. Since the Arabic program in Taiwan is a MSA-only, the English-speaking learners of Arabic were recruited from a MSA-only program.

A tier-one private research university on the east coast of the United States was selected as the data collection site. This data collection site hosts a MSA-only Arabic program, which offers Arabic as an undergraduate major and graduate degrees. The Chinese-speaking participants were recruited from a tier-one public research university in Taiwan. The Taiwanese university hosts an MSA-only Arabic program, which also offers Arabic as an undergraduate major as well. The control group, the participants were either college students at the University of Michigan or ESL students at a language institute in the Ann Arbor area.

To summarize this section, learners of Arabic who were native speakers of English or Mandarin Chinese were recruited. These learners were enrolled in either the first-year or the second-year of Arabic class offered at their home institutions at the time of recruitment. The participants were not heritage speakers of Arabic, and had not lived or studied Arabic in an Arabic-speaking country for more than one semester or three months.

3.2.2 Recruitment Procedure

The same process was followed to recruit participants in both data collection sites. First, the department chairs were contacted to gain permission to recruit learners of Arabic. The instructors for the Arabic classes at the data collection sites were then contacted to acquire permission for a class visit and for promoting recruitment in class. After receiving permission from the instructor, a brief introduction to the experiment was presented at the beginning of the class visit and recruitment flyers were handed out to the students in class as well as pasted to bulletin boards on the campus. Neither the introduction nor the flyer mentioned the core design of the experiment, i.e. lexical stress in Arabic; instead, a general description of the tasks such as "you will read and listen to some Arabic words" was presented Students were told that participation was completely voluntary, but that incentives would be offered for completed participation .. The recruitment flyer was prepared in both English and Chinese, as presented in Appendix 3.9. As a note, the English version of recruitment flyer was used for the recruitment of English-speaking participants as well as the participants in the control group.

3.2.3 Recruitment Results

Based on the recruitment criteria set for the experiment, a total of 76 participants were recruited to participate in the experiment. The complete breakdown of the participants is presented in Table 3.5. A total of 67 participants completed the experiment in the experimental group (henceforth the non-native groups) whereas 9 participants in the control completed the task. Within the experimental group, 33 English-speaking learners of Arabic and 34 Chinese-speaking learners of Arabic participated in the experiment. The English-speaking group had 20 participants from the first-year and 13 participants from the second-year whereas the Chinese-speaking group had 21 first-year participants and 13 second-year participants. The mean age for the English-speaking group was 20.9 years-old for participants in the first-year group, 22.7 years-old in the second-year group whereas the mean age of the first-year and second-year Chinese-speaking participants was 19.1 years-old and 19.7 years-old, respectively. The mean age for the control group was 25.9 years-old. While the mean age of the two non-native groups was comparable, English-speaking participants had a greater standard deviation for age than the Chinese-speaking participants. The participants in the control group had greater standard deviation in age than both the non-native groups.

Table 3.5 Composition of Participants

Groups	**Length of Exposure**	**Credit Hours Enrolled in**	**Gender M/F**	**Age Mean**	**SD**
English L1					
Group 1 (n=20)	Year 1	5	8/12	20.9	4.2
Group 2 (n=13)	Year 2	5	6/7	22.7	3.5
Mandarin Chinese L1					
Group 1 (n=21)	Year 1	7	6/15	19.1	0.9
Group 2 (n=13)	Year 2	5	4/9	19.7	0.8
Arabic L1: Control (n=9)	--	--	7/2	25.9	7.1

M/F = Total Males/Total Females

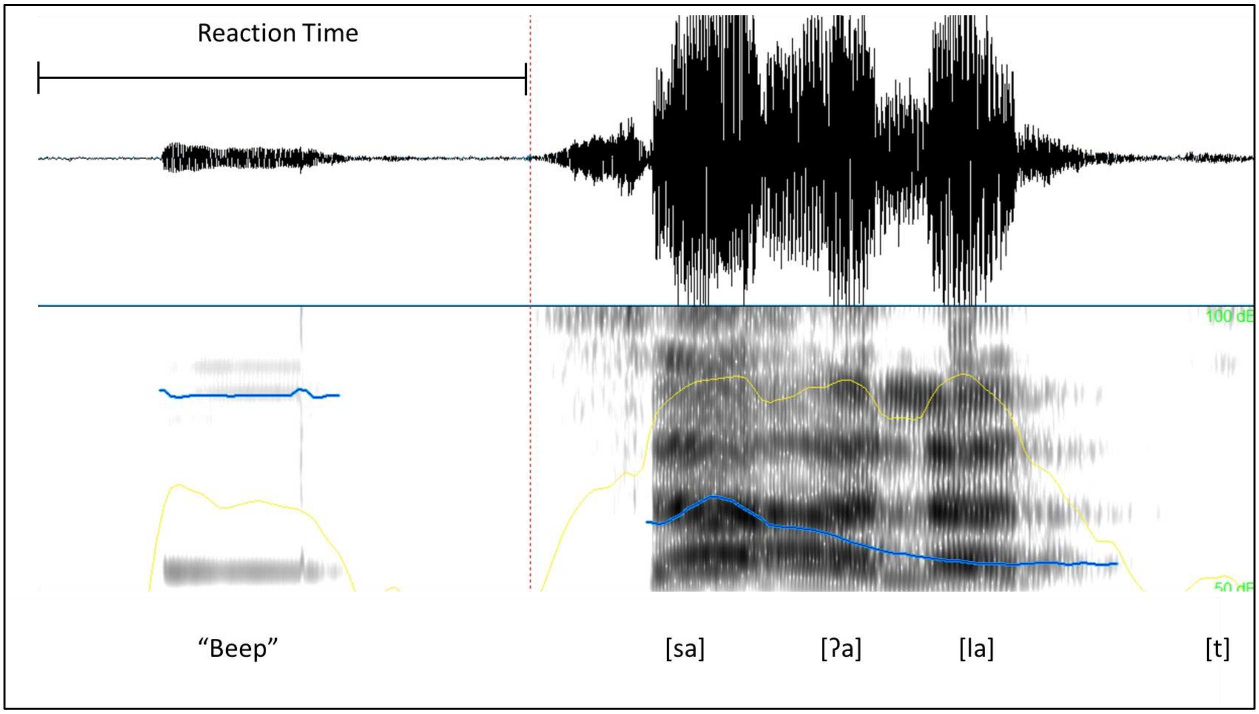
3.4. Data Analysis

As discussed in earlier sections in this chapter, for most tasks in the experiment, the reaction time and correctness of participants' responses in the experiment were automatically maintained and analyzed by the PsychoPy program used for testing, except for the stress production task, which required manual analysis using external programs. This section discusses the process of data analysis for the responses obtained in the stress production task.

Recall that, in the stress production task, participants are presented with a visual stimulus, which is either a real Arabic word or a nonsense word. The same stimulus is prompted twice and each time the participants see the stimuli, they are required to read out the word, resulting in two repetitions for the same stimuli, which are referred to as trial 1 and trial 2. The purpose of the analysis was to determine the reaction time and the correctness of the production in the two repetitions. The analysis of the two measures was conducted using the software program PRAAT, which is widely implemented for phonetic analysis.

To determine the reaction time of each production, each response was analyzed using the following procedure. Figure 3.24 presents a sample analysis of a production. First, each repetition was prompted with a beep to help anchor the participant' production in the program, which is shown on the left part of the figure. The onset of the production is identified by the red line; the location was determined by inspection of the spectrogram and the intensity level (yellow curve). The reaction time is the difference in ms between the timestamp of the onset of production and the beginning of the recording with a PRAAT script that iterates over all annotated data. Incomplete productions are not included in the analysis.

Figure 3.24 Sample Analysis for Stress Production Responses



The second analysis was to determine the correctness of the response, that is, to determine whether the participants place stress on the correct syllable. There were two stages for this analysis. The values for the acoustic correlates for each syllable of the production were extracted and then the syllable that was most likely to be stressed was identified based on the values of these correlates. Three correlates are most relevant for stress, higher F0, duration and intensity for stressed than unstressed syllable. To extract the values of the three correlates, an analysis that is similar to determining the reaction time of the production was conducted. The only difference is that the analysis focuses on the onset and closure of the vowel in each syllable of the production instead of only on the onset of the entire production. That is, the analysis identified the timestamp of the beginning and the end of all vowels in the production as stress is primarily carried by the vowels. A PRAAT script was run to iterate over all annotated recording to extract the three correlates for each syllable of each production.

In the second stage of analysis, these values of the correlates for each individual production were submitted to a scoring schema to evaluate the stressed syllable of the production. The scoring schema operated as follows: for each correlate, compare the strength of the correlate across all syllables and assign a score to each syllable based on the value of the correlate; that is, the higher the value of the correlate, the higher the score given. The score ranges from 1 to N where N is equal to the number of syllables in the word. This process is repeated for all three correlates and each syllable at this point will have three scores based on the relative strength of the three correlates. These three scores are added again to become the final score for the syllable. The syllable that receives the highest final score is determined to be the stressed syllable of the production. For example, consider a hypothetical production for the trisyllabic word مِلْعَقة "a spoon" [ˈmɪl.ʕa.qa] with the value of the correlates for respective syllable presented in Table 3.6. The highest score that can be assigned is 3 since the word has three syllables. Accordingly, the first correlate is evaluated across all syllables and it is found that the first syllable has the highest F0, followed by the second syllable then by the last. As a result, the first syllable receives the highest score 3 and the second syllable receives 2 and the last 1. After evaluating the remaining correlates, i.e., duration and intensity, the total score received by each syllable is calculated, and it is shown that the first receives the highest score among the three syllables, and therefore is determined to be the stressed syllable of that token. Productions with missing values for a correlate are not analyzed. In the event that two or more syllables have the same final score, the F0 score is used as a tie-breaker. That is, the syllable with higher F0 is chosen as the stressed syllable.

Table 3.6 Correlates for A Hypothetical Production

	CORRELATE VALUE			**SCORE**		
	1st Syllable	2nd Syllable	3rd Syllable	1st Syllable	2nd Syllable	3rd Syllable
F0 (Hz)	200	185	150	3	2	1
Duration (ms)	90	60	100	2	1	3
Intensity (dB)	70	60	50	3	2	1
			Total	**8**	**5**	**5**

It is worth noting that previous studies in stress perception and production (e.g., Chen et al., 2001 and Lai 2008) have reported that L1 English speakers in many situations do not utilize F0 to contrast stressed and unstressed syllables in their production of English words. To examine whether such preference holds in English speakers' production of Arabic stress, their production of stressed and unstressed syllable in the stress production task were analyzed and the results are presented in Table 3.7. The value in each cell is the average correlate for the stressed syllables deducted by that of the unstressed syllables. In other words, if the value is positive, it means that the values of the correlates are higher in stressed syllables than the unstressed ones. The results confirmed that the stress correlates are generally higher for the stressed syllables than the unstressed ones as the difference was generally positive for all groups of participants, except for the English speakers. Specifically, the F0 value for the stressed syllables were generally lower than the unstressed ones in the production of the English-speaking groups, suggesting that English speakers do not seem to utilize F0 contrast their production of stress.

Table 3.7 General Difference in Correlates between Stressed and Unstressed Syllables

	TRISYLLABIC			QUADRISYLLABIC		
	Δ F0	**Δ Duration**	**Δ Intensity**	**Δ F0**	**Δ Duration**	**Δ Intensity**
AR	0.80	0.28	0.87	0.77	1.03	1.35
EN 1st Year	-0.10	0.27	0.54	-0.34	0.71	0.47
EN 2nd Year	-0.18	0.32	0.64	-0.35	0.76	0.72
TW 1st Year	0.67	0.55	0.68	0.67	0.95	0.71
TW 2nd Year	1.03	0.64	0.89	1.07	0.88	0.96

This lack of utilization of F0 for stress seems to justify a special accommodation for the English speakers by lowering the weight of the F0 in the scoring scheme since they used duration and intensity to signify stress more than they do for F0. The proposed accommodation was to multiply the score obtained for each correlate by 3, except for the score obtained from F0 only for the English-speaking groups.

However, the results yielded with such accommodation were not significantly different from those without such accommodation. To be methodically sound and to avoid any potential bias from this special accommodation, the score schema remains as originally proposed; that is, without any accommodation (lowering the weight of F0) for the lack of utilization of F0 by English speakers.

3.5 Summary

In this chapter, the essential components and considerations regarding the many facets of the experiment were discussed. The discussion mainly focused on four aspects of the experiment: the tasks, the stimuli, the participants and the analyses.

To sum up, the present design includes a warm-up activity, three main tasks to assess participants' fluency and accuracy in stress perception and production, and a questionnaire. The warm-up activity was a speed reading drill which required the participants to quickly read aloud a list of words. The three main tasks were the stress production task followed by the stress identification tasks held concurrently with the lexical decision task. In the stress production task, the participants were presented with visual stimuli and were required to quickly read aloud each stimulus twice. The stress identification task required the participants to determine the position of the stressed syllable of the stimuli that were presented to the participants aurally. For each stimulus in the stress identification task, the participants were also prompted in a lexical decision task to determine whether the stimulus was a real Arabic word. Lastly, to inform across-participants variability, a questionnaire was compiled that probed the participants' perception of instruction on stress received from the instructor and textbook, their self-awareness and notice of stress, their explicit knowledge of stress via a short quiz, and their social engagement with the target language.

The crux of the stimulus design was to incorporate stimuli with contrastive frequency features. To identify these frequency features, the Buckwalter-Parkinson frequency dictionary was used to extract frequency distributions of relevant structures (i.e., stress pattern, stress position, syllabic structure and conditional probability) for lexical stress in Arabic. The data obtained from the frequency dictionary was further validated across other sources, including the Hans Wehr Arabic-English dictionary and the *Al-Kitaab* textbooks series. Accordingly, these frequency distributions were converted into binary frequency feature and were further manipulated or controlled to create a list of stress patterns for stimulus design . Since it was decided to include both real and nonsense words in the stimuli, real words that match the stress patterns on the list were selected from the frequency dictionary. On the other hand, nonsense words were constructed using two approaches, which all had to do with root consonants that have low co-occurrence probability. The first approach used the real words in the stimuli and replaced one root consonant from the real words with one that infrequently co-occurs with other root consonants in the word. The other approach created a complete new nonsense word by creating a new unattested root with root consonants that infrequently co-occur and combined it with a pre-determined template.

The experiment tests two groups of learners of Arabic: native speakers of English and Mandarin Chinese who are not heritage speakers of Arabic and who have not lived in Arabic-speaking countries for more than 3 months or one semester. A group of native speakers of Arabic is also tested as the control group. The participants from the experiment had different levels of proficiency in the target language; they were enrolled in either first-year, or second-year Arabic classes at the time of recruitment.

Lastly, since the responses were automatically analyzed in terms of reaction time and correctness of stress placement in nearly all tasks in the experiment, the discussion of data analysis focused on the procedure and processes involved for analyzing the responses obtained from the stress production task that cannot be analyzed automatically. To this end, the present study conducted an acoustic analysis of each produced syllable to determine stress location. A scoring schema was implemented that evaluated and assigned scores based on the relatively strength of the correlates across all syllables in a production, and chose the syllable that had the highest score to be the stressed syllable of the production.

In sum, the experiment incorporates many variables in its design, which includes the following:

1. Internal and proficiency factors based on the participants: native language (L1) and their level of proficiency
2. Input factors based on the stimuli: number of syllable (length), relative frequency, stress patterns and nonsense patterns of the stimuli.
3. External factors based on the participants: explicit knowledge of lexical stress in Arabic, extracurricular engagement with the language, awareness, instruction and textbook

Through this experiment, the present study examined these variables and learners' performance in stress perception and production. A more detailed discussion of the results obtained from the experiment is presented in Chapter 4.

Chapter 4

Results

In this chapter, I report the results of the three main tasks in the experiment. The tasks are discussed below in the following order (not in the order in which they were administered): the stress identification task, then the lexical decision task followed by the stress production task. In addition, the results of the speed reading task, which served as the warm-up task for the experiment and as another dimension of participants' overall proficiency, are reported as well. I examine how the various factors identified in the study influenced the fluency and accuracy of the learners in these tasks to answer the research questions proposed in Chapter 1. As discussed in Chapter 3 in Section 3.5, these factors can be divided into three main categories below:

4. Internal factors of the participants: native language (L1) and proficiency level (proficiency henceforth)
5. Input factors of the stimuli: number of syllables (length), relative frequency, stress patterns and nonsense patterns of the stimuli.
6. External factors based on the participants' background: explicit knowledge of lexical stress in Arabic, extracurricular engagement with the language, awareness, instruction and textbook

In the following sections, I first report my observations of classroom instruction received by the two groups of participants in their respective Arabic programs at the two data collection sites. Next, I present the results from the speed reading task, then the stress identification task, followed by the results from the lexical decision task and conclude with the production task results. For each task, the experimental results were analyzed by the factors mentioned above, mainly by the internal and input factors. The interaction between external factors and participants' performance was also analyzed and the results are presented at the end of this Chapter. It must be noted that the results presented in this Chapter were participants' responses to stimuli that were nonsense words only, to avoid any bias that might be introduced by participants' previous knowledge of the stimuli.

4.1 Observation of Classroom Instruction

For data collection, I visited a private research university on the East Coast of the United States to recruit English-speaking learners of Arabic. Additionally, I visited a public research university in Taiwan to recruit participants who are Chinese-speaking learners of Arabic. Both institutions offer Arabic as an undergraduate major. There were many similarities shared by the two institutions as well as some disparities between the two. The similarities and the disparities were previously discussed in the Participants section in Chapter 3. I was fortunate to be given the opportunity to visit nearly all Arabic classes offered at the English-speaking site at both levels, the first-year and second-year. At for the Chinese-speaking site, although the opportunity to be physically inside the Arabic classes was not available, I was allowed to observe the class outside of the classroom without disrupting the instruction for most of the classes offered.

In this section, I report my observations from my class visits at the two Arabic programs, regarding topics that are instrumental to better understanding the input frequency that the participants received. These topics include: (1) the composition of students, (2) the contact hours with the target language, (3) the general pedagogical approach in teaching Arabic, (4) the variety of Arabic used in the classroom, and (5) the textbook used in the program.

My first observation is that the composition of the students was rather distinct, which largely had to do with the education systems in each country. At the English-speaking site, the students enrolled in the Arabic classes did not necessarily major in Arabic. Results of the post-experiment questionnaire that was completed by the participants showed that the English-speaking participants had a greater span in their specializations than the Chinese-speaking participants, with the specializations of the English-speaking participants ranging from humanities and social sciences to natural sciences. These more varied backgrounds suggested that the English-speaking participants took other classes required by their specification as they are taking the Arabic class, which resulted in a more diversified class schedule.

In contrast, the Chinese-speaking participants had more homogenous backgrounds, as the specialization of the participants at Chinese-speaking site was predominately Arabic language and literature. As shown in the post-experiment questionnaire, all Chinese-speaking participants majored in Arabic, except for one participant who majored in Economics but also double-majored in Arabic. The fact that all Chinese-speaking participants were Arabic majors is due to the collegial education system in Taiwan. In the Taiwanese education system, the students choose their major prior to entering the university, and the classes that they take are predominately determined by their major, resulting in less mobility between specializations, i.e. less transfer between majors, and less flexibility in the classes they take. Therefore, for the

Chinese-speaking participants in the experiment, the main component of their collegial curriculum was classes in Arabic language and Arabic-related topics, with less focus on other subjects. It is worth noting that both programs did not have many heritage learners of Arabic. Fewer than five heritage learners were identified (none of them was recruited as heritage speakers do not meet the recruitment criteria) at the English-speaking site, and no heritage speaker was identified at the Chinese-speaking site.

The distinct composition of the participants in the two data collection sites led to another disparity between the two groups: the contact hours with the target language. The Arabic classes offered at the English-speaking data collection site in most cases are under one unified class title for each level, such as “first year Arabic” or “second year Arabic” and so on. These classes normally are 5 credit hours and the class met every day during the week for approximately one hour per day, totaling 5 contact hours weekly for all levels of Arabic classes that are of interest for this study.

It is worth noting that there was a group of graduate students at the English-speaking site who were required to enrolled in a special section of Arabic class taught by a designated crew of faculty members. For these special sections, the class met five hours a week as the regular section of Arabic class. Additionally, the special section of Arabic class met two extra hours weekly for conversation hours, which is a social activity facilitated by their peers and sometimes by a faculty member. The participants who enrolled in these special sections were in the second of learning Arabic. Specifically, five second-year participants were from this special section of Arabic class. None of the participants in the first-year English-speaking group enrolled in the special section of Arabic classes. The motivation to include these graduate students who enrolled in the special section of Arabic class was to maximize the sample size for each participant group, as the number of participants who enrolled in the regular section, i.e. 9 participants in the second-year group, might not reach the ideal number for statistical tests. Since the participants from this special section of Arabic class spent two extra hours weekly compared to other English-speaking participants who enrolled in regular curriculum, these extra hours might potentially introduce biases in the performance of the participants which in turn might potentially influence the results of the experiment. This topic is addressed in a later section (see Section 4.6) by comparing the results with and without these participants who enrolled in these special sections.

Nevertheless, the Chinese-speaking participants had similar weekly contact hours as the English-speaking participants. The Arabic program at the Chinese-speaking site required the first-year students to enroll in two separate Arabic classes, "College Arabic" and "Arabic tutorial", totaling seven credit hours, equivalent to seven contact hours weekly. The second-year students enrolled in two similar classes, with a total of five credit hours, equivalent to 5 contact hours weekly. In addition, the second-year students were also required to enroll in "Arabic Grammar", which met two extra contact hours. Since the class was conducted in Chinese and the input in Arabic was rather limited, the two hours spent in the class were not considered to be contact hours in the present study. Although the number of contact hours at the Chinese-speaking site seemed to be generally more than the contact hours at the English-speaking group, these contact hours were not as evenly distributed at the Chinese-speaking site as in the English-speaking site. Unlike the Arabic classes offered at the English-speaking site that met every day during the week, the "College Arabic" class and the "Arabic Tutorial" classes offered at the Chinese-speaking site met twice a week only. It is then possible for the Chinese-speaking students not to have any Arabic class on some days during the week, and minimally two days in a week if the two classes all meet on the same day.

The class size differed at the two data collection sites as well. Overall, the class size at the Chinese-speaking site was larger. For the first-year classes, each section of the first-year College Arabic class had around 15 students, whereas the first-year Arabic Tutorial class had approximately 20 students in each section. For the second-year class, the class size had an average of 20 students in each section. The Arabic Grammar class had a larger class size, with approximately 35 students. At the English-speaking data collection site, class sizes averaged around 13 students or less for some sections of the class throughout from first-year to second-year classes. The smaller size might suggest that the English-speaking students had greater opportunity to interact with the instructor in class, compared to the Chinese-speaking students.

Lastly, it must be noted that the average length of a semester is generally longer at the Chinese-speaking site than at the English-speaking site. As an example, the length of the Winter semester in 2017 for the Chinese-speaking site is nineteen weeks, as opposed to fifteen weeks, including winter break, for the English-speaking site.

The pedagogical approach adopted by the instructor is another factor that could influence the amount of input received by the participants. During my visit to the Arabic classes, I found the pedagogical approach adopted by both programs to be rather similar, in that the teaching did not lean towards extremes of either teacher-centered, lecture-based instruction or a pair-work-intensive, fully communicative approach. The pedagogy of the two programs fell somewhere between the two extremes, as interaction existed in class but it often occurred between the students and the instructor and occurred rather less often between students, giving it a slightly teacher-centered mode of instruction in both programs. Nonetheless, the classroom atmosphere was enjoyable, and the teacher-student interaction was pleasant and encouraging. The students seemed motivated to come to class and were very engaged in participating in class activities.

As for the quality of input that the participants received at the two locations, it is essential to examine the quality of the instructor's productions, especially – for the purpose of this study – the accuracy of stress production. Because a good amount of language input that the participants received comes from the instructor, incorrect input or stress production might result in incorrect or non-MSA production of stress pattern. At the English-speaking site, the majority of the instructors were native speakers of Arabic, predominantly Egyptian with a few faculty members from other regions of the Middle East. Although the relatively few non-native speakers mainly worked as teaching assistants that assist the main instructor, some (i.e. ones that are in the later stages of their advanced degrees) did work as the main instructor for the class. Of the five first-year classes visited, four were taught by Egyptian instructors, and one by a Jordanian instructor, with two teaching assistants from Saudi Arabia. For the four second-year classes visited, two were taught by an Egyptian instructor, one by a Palestinian instructor and the other by a non-native instructor, with one teaching assistant from Saudi Arabia. In contrast, of the five full-time faculty members at the Chinese-speaking site, two were native speakers from Jordan, and the remaining faculty members as well as other adjunct instructors were non-native speakers.

Although the opportunity for recording the instructor's voice was not available at either data collection site, I was given permission to sit in on the classes and observe the instruction and the language produced by the instructors at the English-speaking site. Although some features that are specific to the Egyptian dialect occasionally occurred during instruction, e.g., the use of [g] in place of [dʒ], the stress patterns used during teaching conformed to the stress pattern in Modern Standard Arabic (MSA). The stress patterns in MSA were consistently produced by all instructors observed, native and non-native alike. Additionally, the productions of the students also conformed to MSA stress patterns and did not seem to be influenced by other dialects.

For the Chinese-speaking site, even though the opportunity for in-class observation was not available, the opportunity to observe the class from outside of the classroom was available. The instruction and the language produced by the instructor and the students were clear from outside of class. Similar to the English-speaking counterparts, the instructors at the Chinese-speaking site consistently produced stress patterns that conformed to MSA without notable use of the Jordanian dialect.

Finally, there is overlap between the textbooks used at both data collection sites. At the English-speaking site, the second-edition *Al-Kitaab* curriculum (Brustad et al., 2011) was adopted extensively through all two levels of Arabic classes. In addition, supplementary materials, such as news articles, reports and worksheets, were provided to the students, especially those in the second-year class. For the Chinese-speaking groups, two textbooks were mainly used, which were the second-edition *Al-Kitaab* curriculum, *Al-ʻArabiyya bayna yadayk* (Al-Fuzan et al., 2014), and supplementary materials made by the instructors, as the level and the topic of the class (recall that there were three different classes taken by the Chinese-speaking students, the College Arabic, the Arabic Tutorial and the Arabic Grammar) could influence the choice of the textbook used for the class. It is, however, safe to say that both institutions had a strong MSA orientation in that not many dialectal features were introduced in the language class.

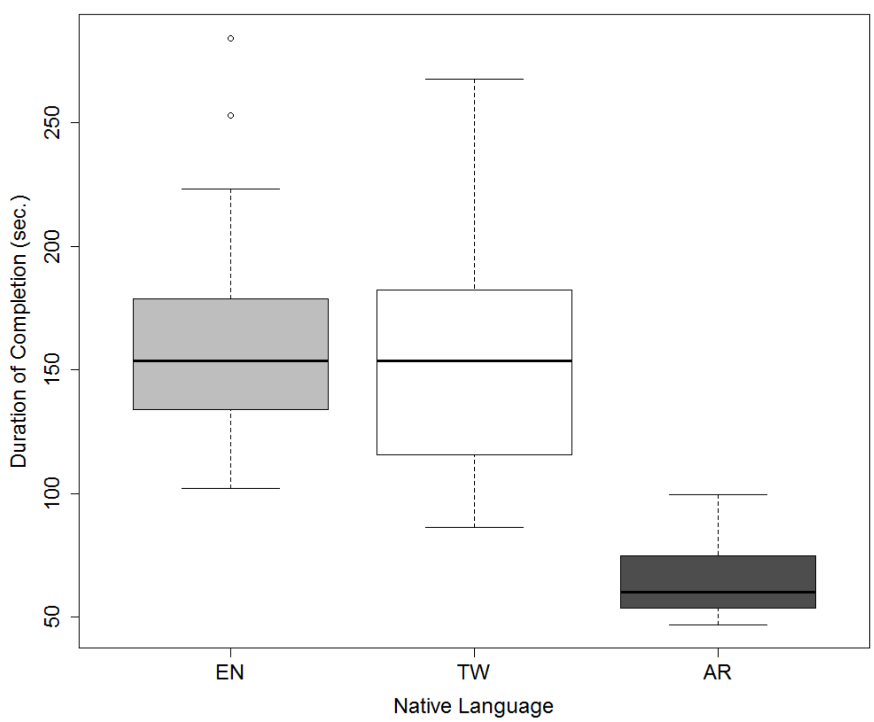
To sum up, the two data collection sites had much in common in terms of the quality of the input given to their students, as both sites used MSA as the main variety of Arabic and the inclination towards MSA was reinforced by their choice of textbooks. However, the amount of input that the students received have a greater variation. Even though the English-speaking learners of Arabic had more evenly distributed contact hours and smaller class sizes, the Chinese-speaking learners of Arabic had generally longer contact hours weekly as well as longer semesters, resulting more exposure to the target language.

4.2 Warm-Up Task: Speed Reading

The speed reading task was the first task in the experiment, serving as the warm-up activity for producing Arabic. In this task, all participants were given the same wordlist that contained 48 real Arabic words with various stress patterns in the same order, and were asked to read aloud every word on the list as fast as they could. The purpose of the speed reading task was not only to better prepare the participants for the main tasks of the experiment, as the stimuli in the main tasks were more controlled, but also to measure the participants' reading fluency, serving as another dimension of participants' overall proficiency. Fluency here refers to the duration of completion of the task; in other words, it assesses the time that the participants took to finish reading all words on the list. Specifically , the participants were prompted with a beeping sound at the onset of the task to begin reading. The duration of completion was measured from the end of the beeping to the end of their production of the last word on the wordlist.

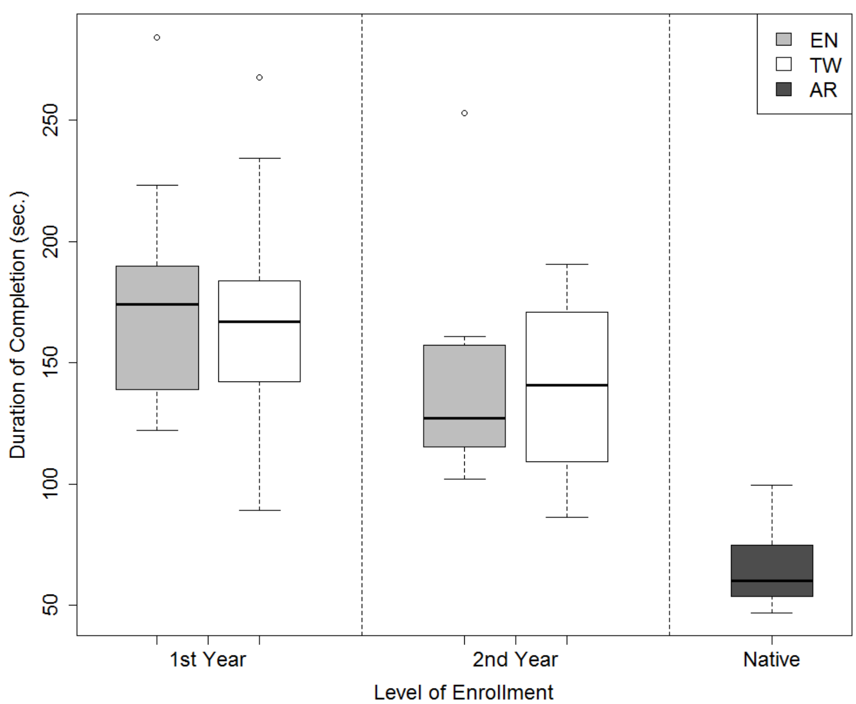
Since the speed reading task was constructed using real Arabic words, it might not be methodologically sound to examine the performance of the participants with some of the factors identified in the study, such as frequencies of stress-related cues. Therefore, the general distribution of performance by different participant groups is reported. Figure 4.1 presents duration of completion (in seconds) according the participants' native language, including native English- (EN) and Chinese- (TW) speaking testing groups as well as the native Arabic- (AR) speaking control group. Briefly, English-speaking and Chinese-speaking groups appear to have similar performance on the task whereas the Arabic group completed the task more quickly than both non-native groups. A one-way analysis of variance (ANOVA) with native language (three levels: Arabic, English, and Mandarin Chinese) as a between-group variable and duration of completion as a dependent variable was conducted and confirmed this observation. The analysis revealed that the main effect of native language was significant ($F(2, 86) = 16.04$, $p <.001$, partial $\eta^2 =0.27$). Post-hoc comparisons using the Tukey tests showed that the native group significantly outperformed both the English-speaking group ($p < 0.001$) and the Chinese-speaking group ($p < 0.001$) with significantly shorter duration of completion; however, no significant difference was found between the two non-native groups ($p = 0.36$).

Figure 4.1 Boxplot for Duration of Completion for Speed Reading According to L1



The lack of significant difference between the non-native groups, however, could be due to collapsing together participants who vary in proficiency. To further analyze performance of the non-native groups, a two-way ANOVA was conducted to further investigate the effect of proficiency on the performance of non-native participants, with native language (two levels: English, and Mandarin Chinese) as a between-group variable, proficiency level (two levels: 1^{st} year and 2^{nd} year) as a within-group variable, and duration of completion as dependent variable. Figure 4.2. demonstrates the performance of participants in the speed reading task grouped by the native language and proficiency of the participants. There was a significant effect of proficiency ($F(1, 63) = 9.51$, $p < 0.001$, partial $\eta^2 = 0.13$) on the duration of completion, while there was no significant main effect of native language nor an interaction of native language and proficiency. Post-hoc comparisons using Tukey tests indicated that the second-year participants had significantly shorter ($p < 0.001$) duration of completion than their first-year counterparts. In other words, the results suggest that the participants with higher proficiency – the second-year students compared to the first-year students – completed the task more quickly, regardless of their native language.

Figure 4.2 Boxplot for Duration of Completion for Speed Reading Drill According to L1 And Proficiency



Two (one for first-year and the other for second-year non-native participants) additional one-way ANOVA tests were conducted to compare the performance of non-native speakers of a specific level of proficiency and that of the native speakers, with native language as the between-group variable (three level; Arabic, English and Mandarin Chinese) and duration of completion as the dependent variable. Both tests found a significant main effect native language on duration of completion. Post-hoc comparisons using Tukey tests indicated that the native group took significantly less time to complete the task than either the first-year ($F(2, 47) = 27.14$, $p < 0.001$, partial $\eta^2 = 0.53$) or second-year ($F(2, 32) = 14.88$, $p < 0.001$, partial $\eta^2 = 0.48$) participants, while no significant difference was found between the English-speaking and Chinese-speaking participants, regardless of their level of proficiency.

In sum, the results indicate that Arabic-speaking participants completed the warm-up activity more quickly than the English-speaking and the Chinese-speaking groups. Additionally, the results show that participants with higher proficiency completed the task faster, in that the second-year completed the task significantly more quickly than the first-year participants.

4.3 Stress Identification Task

The stress identification task was one of the three main tasks in the experiment. In this task, the participants simultaneously heard an Arabic word and saw an Arabic script that corresponded to the word. After hearing the word, the participants were asked to identify the position of the stressed syllable in the word by pressing corresponding keys on the keypad. In this task, the reaction time (RT) and the response accuracy (RA) were measured for all participants. The reaction time is the amount of time it took the participants to respond to each stimulus, which is automatically measured by the computer program PsychoPy, used to administer the experiment. However, as discussed in Chapter 3, the program only records the duration between the onset of each trial and the time tag of the keypress. Because stimulus duration also affects response time, the reaction time used for statistical analysis is the duration recorded by the computer program (measured in milliseconds) minus the duration of the stimulus that the participants heard. As for the response accuracy, the responses of the participants were automatically determined by PsychoPy to be either correct (coded as 1 in the dataset) or incorrect (coded as 0 in the dataset) using the answer keys that were pre-programmed in the testing interface, and the proportion correct is calculated for each individual to represent the overall accuracy. For the analyses for all main tasks, response accuracy and reaction time served as the main dependent variables.

It is equally important to note that unless otherwise specified, the results reported in this Chapter are based on the analyses of participants' responses to stimuli that are nonsense words.

4.3.1 Native Language and Performance

Figure 4.3 gives response accuracy according to the participants' native language where "EN" represents English-speaking participants, "TW" Chinese-speaking participants and "AR" the participants in the control group. Both non-native groups received above-chance accuracy, while the control, surprisingly, had nearly below-chance accuracy, suggesting that the non-native participants were able to identify the stressed syllable in the stimuli whereas the native speakers had difficulty with the task. A one-way ANOVA was conducted to determine the extent of this difference, with native language as a between-group variable and correctness rate as the dependent variable. The test indicated that there was a significant effect of native language on response accuracy ($F(2, 73) = 11.3$, $p < 0.001$, partial $\eta^2 = 0.23$). Post hoc comparisons using

Tukey tests showed that while no significant difference (p = 0.94) was found between the English-speaking and Chinese-speaking groups, the non-native groups had significantly higher (p <0 .001) accuracy than the control group.

Figure 4.3. Boxplot for The Response Accuracy for the Stress Identification Task According to L1

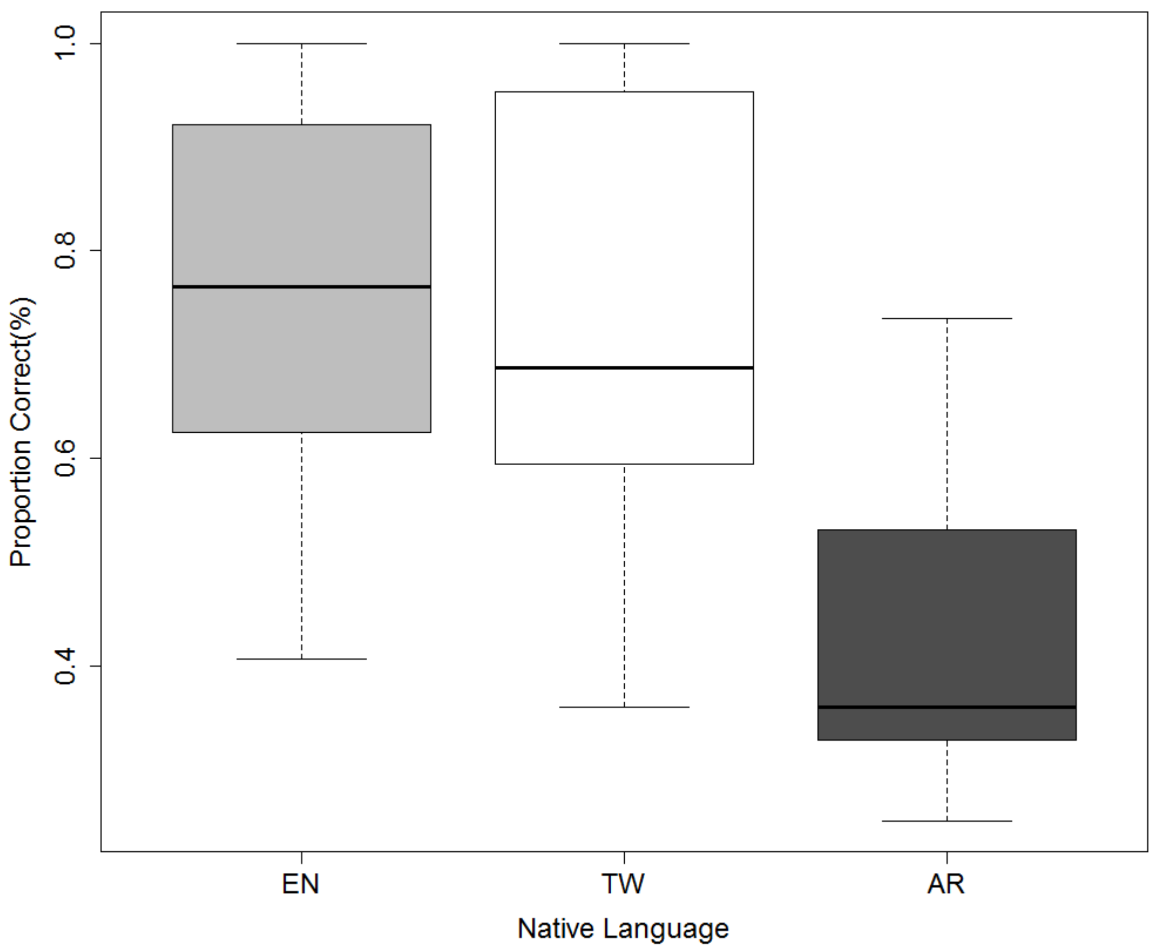
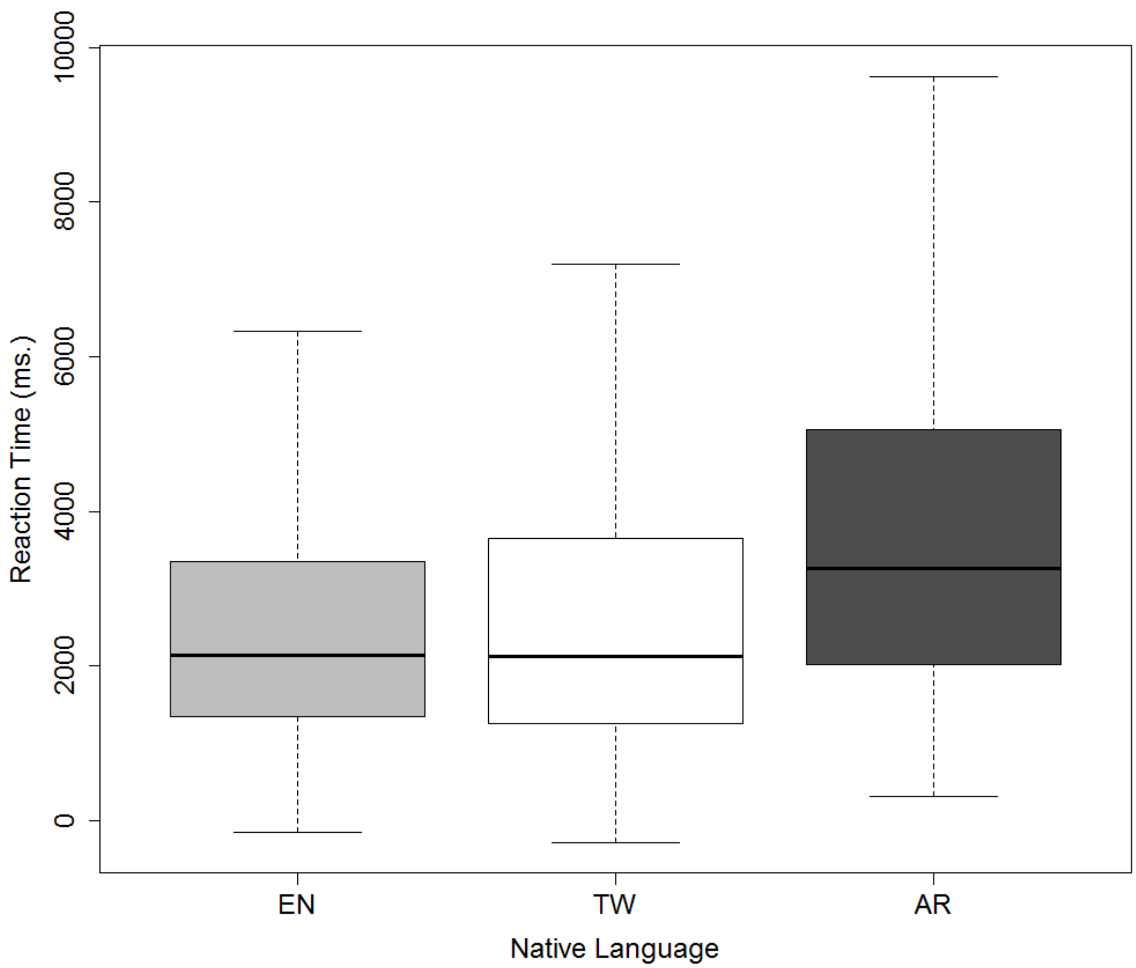


Figure 4.4 presents the reaction time results according to the native language of the participants, where the participants from the non-native groups have shorter reaction times on average. A one-way analysis of variance (ANOVA) with repeated measures was conducted to examine the extent of this difference, with native language as a between-group variable, participants as a random variable to control for individual variability, and reaction time as the

dependent variable. The ANOVA test showed a significant effect of native language ($F(2, 73) = 6.15$, $p < 0.001$, partial $\eta^2 = 0.11$) on reaction time. Follow-up post-hoc Tukey comparisons showed that although there was not a significant difference in reaction time between the non-native groups, both the English-speaking and Chinese-speaking groups had significantly shorter reaction times ($p < 0.001$) than the native group. In other words, while no difference in performance was found between the participants from non-native groups, they took less time to identify the stressed syllable in stimuli than the Arabic-speaking participants did.

Figure 4.4. Boxplot for the Reaction Time
Grouped by Native Language for the Stress Identification Task

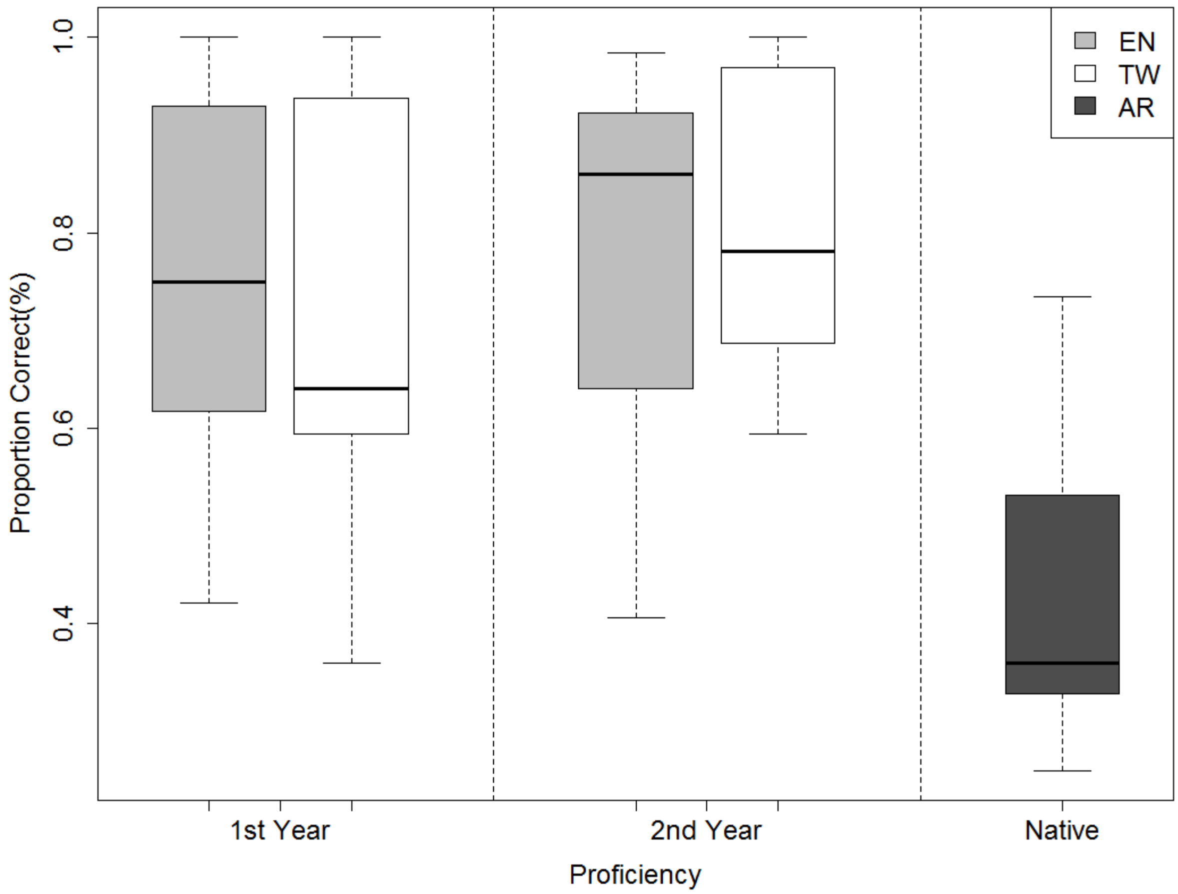


Overall, the analyses thus far found no significant difference in either reaction time or response accuracy between the English-speaking and Chinese-speaking groups. Comparing the native group and the non-native groups, the analyses found that participants who were native speakers of Arabic spent more time on the task and were less successful in determining the stressed syllable in the stimuli than their English-speaking and Chinese-speaking counterparts.

4.3.2 Proficiency and Performance

To further examine the difference in performance between the two non-native groups, I examined the performance of the participants based on their proficiency, which was based the level of class they were enrolled in. Figure 4.5 presents the response accuracy according to the native language (L1) and the proficiency of the non-native participants. A two-way ANOVA was conducted with native language (two levels: English and Mandarin Chinese) as a between-group variable, and proficiency (three levels: 1st year and 2nd year) as a within-group variable and responses accuracy as the dependent variable. The result shows that there was not a main effect of proficiency level nor an interaction between native language and proficiency, which suggests that the non-native-speaking participants had similar response accuracy regardless of their level of proficiency in the target language.

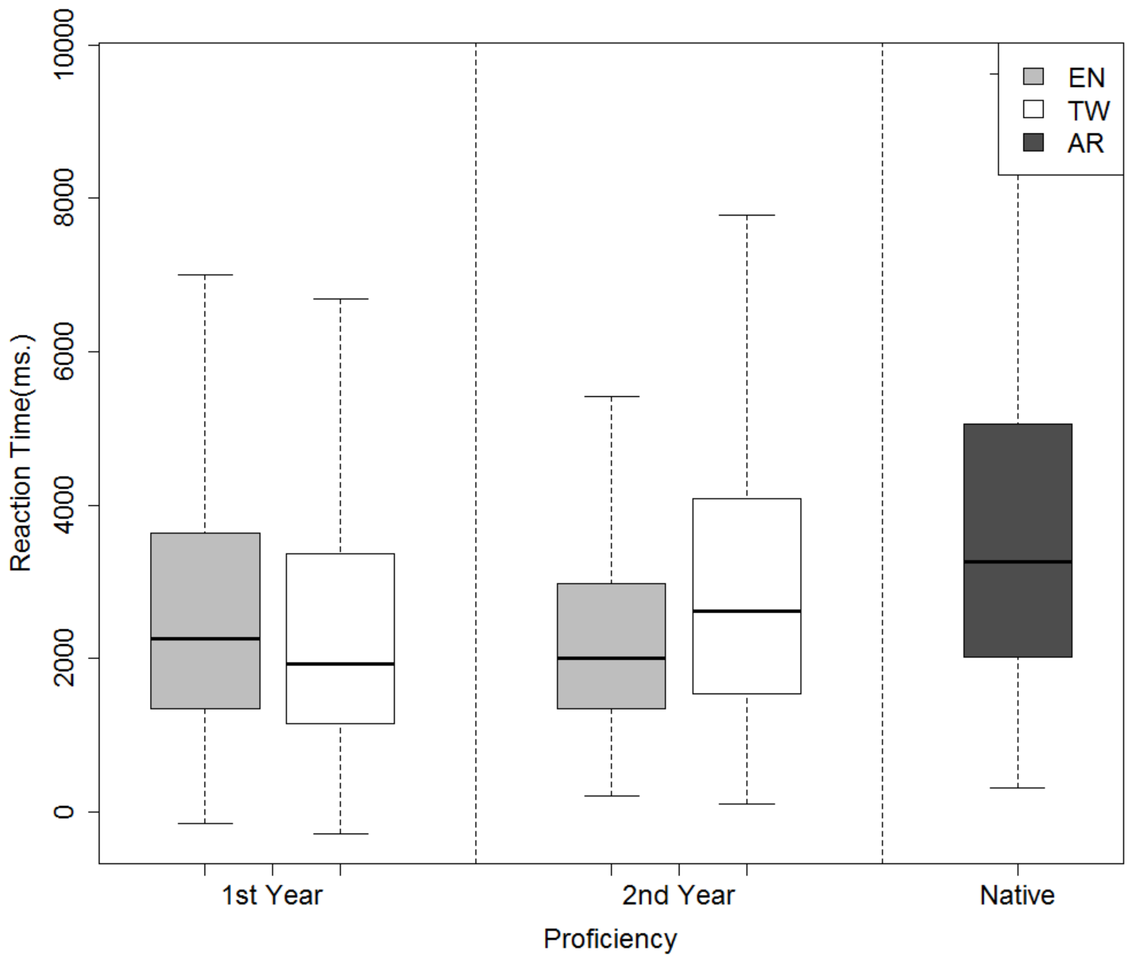
Figure 4.5. Boxplot for the Response Accuracy for the Stress Identification Task According to Participants' L1 and Proficiency



The same analsysis was conducted for reaction time as well. Figure 4.6 represents the reaction times according to the native language and the proficiency of the participants. The same two-way ANOVA was conducted, with reaction time as the dependent variable. However, as in the accuracy analysis, the ANOVA test did not reveal a significant effect of proficiency nor an interaction between native language and proficiency. In other words, the results suggest that the English-speaking and Chinese-speaking groups took a similar amount of time to respond to the stimuli, regardless of their proficiency.

Figure 4.6. Boxplot for the Reaction Time

Grouped by Participants' L1 and Proficiency for the Stress Identification Task



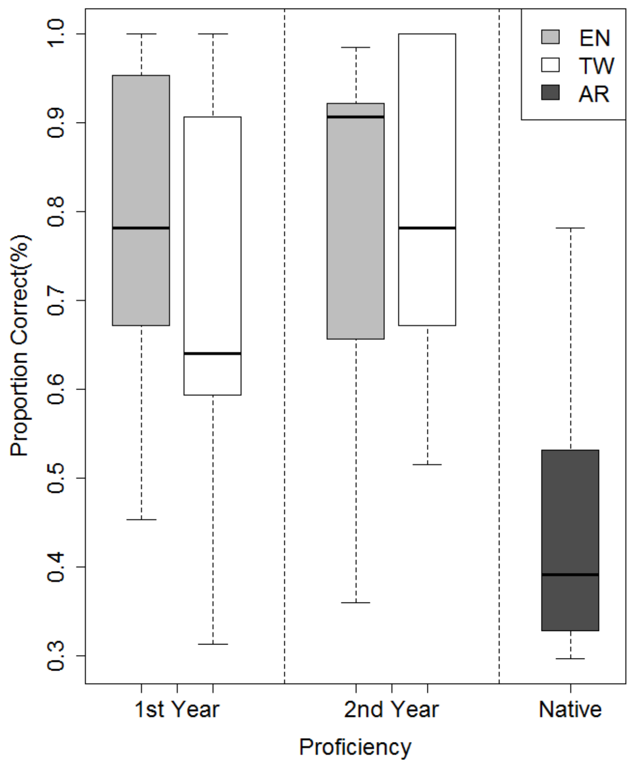
Taken together, the analyses suggest that both the English-speaking groups and the Chinese-speaking groups were similarly accurate in identifying the stressed syllable of the stimuli stress, and their proficiency - as measured in Arabic course level - did not help them do better or worse in the task.

In sum, the results of the analyses conducted in this section all indicate that there was no significant effect of proficiency on either reaction time or response accuracy. That is to say, more years of Arabic study did not help the non-native participants perform better on the stress identification task: they took a similar amount of time to determine the stressed syllable of the stimuli and they were comparably successful in identifying the stressed syllable. On the other hand, the native-speaking participants in the control group took more time to respond to the stimuli and were less successful at identifying the stressed syllable than all English-speaking groups and some Chinese-speaking groups.

The control groups' poorer performance in the task is rather unexpected as one would generally expect the control group to have ceiling performance, which is the opposite of what was observed in the task. Therefore, it is essential to examine whether such poor performance was due to the naturalness of the stimuli, as the analyses discussed before were based on participants' responses to stimuli that were nonsense words. To this end, additional analyses were conducted, now focusing on stimuli that were real Arabic words.

Figure 4.7 presents the accuracy of responses according to the native language (L1) and the proficiency of the non-native participants. As the Figure shows, the control group again had distinctively lower response accuracy compared to the non-native groups. The ANOVA tests confirmed this observation, indicating that there was a significant main effect of native language on the response accuracy ($F(2, 73) = 10.6$, $p < 0.001$, partial $\eta^2 = 0.2$) while there was no effect of proficiency nor was there an interaction between native language and proficiency. Follow-up Tukey tests showed that the participants in the control group were significantly less accurate in the task than their non-native counterparts. That is to say, the native speakers had difficulties in identifying the stressed syllable in the stimuli, regardless of the naturalness (i.e., real or nonsense) of the stimuli.

Figure 4.7. Boxplot For the Response Accuracy (Real Words) Grouped by Participants' L1 and Proficiency for the Stress Identification Task



4.3.3 Stress Pattern and Performance

To provide a more detailed examination, further statistical analyses were conducted by grouping the participants' responses based on the type of stress pattern of the stimuli. As discussed in Chapter 3, the stimuli consist of eight stress patterns. In Table 4.1, the number for the type of stress pattern, and the number of syllables that the stress pattern has, as well as the frequency for stress pattern and stress position are given. For example, stress pattern 1 is trisyllabic and has a frequent stress pattern and frequent stress position. Grouping the participants' responses based on stress pattern types was done to examine whether the participants would have different responses to stimuli with different syllabic structures.

Table 4.1. Stress Patterns Included in the Experiment

	Stress Pattern	Frequency Stress Pattern	Frequency Stress Position	Frequency Syllabic Structure	Conditional Probability
Trisyllabic	ˈCV.CV.CV	**+ (5.9)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
	ˈCV.CV.CVC	**- (3.2)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
	CV.ˈCVV.CV	**+ (5.8)**	**- (5.9)**	**+ (6.1)**	**+ (3.3)**
	CVC.ˈCVC.CVC	**- (4.0)**	**- (5.9)**	**+ (6.1)**	**+ (2.5)**
Quadrisyllabic	CVC.ˈCV.CV.CV	**+ (5.0)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
	CVC.ˈCV.CV.CVC	**- (2.6)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
	CVC.CV.ˈCVV.CV	**+ (4.7)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
	CV.CV.ˈCVV.CVC	**- (3.8)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**

Plus sign (**+**) = Frequent, Minus Sign (**-**) = Infrequent, (): Log Frequency

To examine whether the participants responded differently in terms of accuracy and reaction time to individual stress pattern, a set of ANOVA tests were conducted for each stress pattern to inspect the main effect of native language and proficiency. Specifically, a one-way ANOVA was conducted with native language (three levels: Arabic, English, Mandarin Chinese) as a between-group variable and reaction time/response accuracy as dependent variable, followed by a two-way ANOVA to inspect the effect of proficiency level on reaction times for non-native participants, with native language (two levels: Arabic, English, Mandarin Chinese) as a between-group variable, proficiency (two levels: first-year and second-year) as a within-group variable, and reaction time/response accuracy as dependent variable. If a significant effect was found in either ANOVA test, post-hoc Tukey tests were followed for multiple comparisons of either native language and proficiency. The summary for all ANOVA tests and Tukey tests for stress identification task can be found in Appendix 4.1.

To demonstrate, Table 4.2 presents a portion of the summary table for the main effects of native language and proficiency from Appendix 4.1 for stimuli that had stress pattern 2. The table shows that there was a significant effect of native language ($F(2,87) = 7.17$, $p < 0.001$) on reaction time, but the effect of proficiency and the interaction between native language and proficiency were not significant.

Table 4.2. Summary Table for Main Effect for Stress Pattern 2

	STRESS PATTERN 2	
	F_statistics	P_Value
Native Language	F(2,87) =7.17	p < 0.001***
Proficiency	F(2,78) =0.38	p = 0.69
Native Language x Proficiency	F(2,75) =1.22	p = 0.3

The analysis was followed by post hoc multiple comparisons using Tukey tests to compare the reaction times of participants of different native languages. The partial summary table for multiple comparisons is presented in Table 4.3, where multiple comparisons for groups with different native language backgrounds are given. The table gives the Estimate value for the comparison is presented as well as the significance level. The estimate value represents the mean difference in reaction time between the two groups. For example, a -2030.62 estimate value shown for the comparison between the English-speaking and the Arabic-speaking groups indicated that the average difference between the two group is -2030 (ms), suggesting that the English-speaking group had significantly shorter reaction time than the Arabic-speaking group. Conversely, a positive estimate value, such as the 545.57estimate value found for the comparison between the Chinese-speaking groups and the English-speaking group, suggests that the Chinese-speaking groups had a longer reaction time (albeit not significantly longer) than the English-speaking groups.

Table 4.3. Summary Table for Multiple Comparisons for Stress Pattern 2

Multiple Comparisons – Native Language

	STRESS PATTERN 2	
	Estimate	P_Value
EN-AR	-2030.62	p < 0.001***
TW-AR	-1485.05	p < 0.01**
TW-EN	545.57	p = 0.22

Observing from the summary table in Appendix 4.1.1 for response accuracy and 4.1.2 for reaction time, there was a significant effect of native language on both reaction time and response accuracy in nearly all stress pattern groups, while no significant effect of proficiency was found. The post-hoc Tukey tests showed that, although there was overall no significant difference in either reaction time or response accuracy between the two non-native groups, they were found to have a significantly better performance than the control group. The multiple comparisons showed that the English-speaking participants took significantly shorter time and were more accurate in the task than Arabic-speaking group. A similar advantage was found for the Chinese-speaking group as well, albeit less strongly, as the Chinese-speaking participants also had significantly shorter reaction time and higher response accuracy than their Arabic-speaking counterparts in the control group.

On the other hand, there was generally no significant effect of proficiency on task performance. The only exceptions were the interaction found between native language and proficiency in stress patterns 3 and 4, where the post-hoc tests show that Chinese-speaking participants with higher proficiency surprisingly spent significantly longer time to respond.

The findings from the stress pattern group comparison further confirmed the findings from the “Native Language and Performance” section of the stress identification task. Native language did not have significant effect on the non-native participants' reaction time or response accuracy, as both English-speaking and Chinese-speaking had comparable performance on the task. Surprisingly, the post hoc multiple comparisons found that the native speakers of Arabic did significantly worse than their English-speaking and Chinese-speaking counterparts in both reaction time and correctness rate for nearly all stress patterns. Additionally, the results again confirm that the proficiency of participants overall had little effect on either the reaction time or correctness of response for English-speaking participants.

4.3.4 Relative Frequency and Performance

To examine the role of frequency of the stress pattern on the performance of the participants, I identified five groups of stress patterns for comparison. Each group consisted of stimuli that belong to two stress patterns that were minimally contrastive in their relative frequencies. Of the five groups, three were trisyllabic and two of them were quadrisyllabic. Table 4.4 gives the information for the stress patterns in each group. The relative frequency of the stress patterns within each group was determined by comparing the frequencies of their stress pattern and stress position. For example, for the two stress patterns in stimulus group 1, [ˈCV.CV.CV] was more frequent in terms of its stress pattern as well as the stress position, compared to the other pattern in the group, [CVC.ˈCVC.CVC], which has both an infrequent stress pattern and an infrequent stress position. Stress patterns that were more frequent were marked with the plus sign "+", whereas the less frequent ones were marked with the minus sign "-" in the "Relative Frequency" column.

Table 4.4 Groups of Stress Patterns with Contrastive Relative Frequency

Number of Syllable	Stress Pattern	Frequency of Stress Pattern	Frequency of Stress Position	Relative Frequency	Vowel Length	Stimuli Group
3	ˈCV.CV.CV	**+ (5.9)**	**+ (6.2)**	+	Short	1
3	CVC.ˈCVC.CVC	**- (4.0)**	**- (5.9)**	-	Short	1
3	ˈCV.CV.CV	**+ (5.9)**	**+ (6.2)**	+	Short	2
3	ˈCV.CV.CVC	**- (3.2)**	**+ (6.2)**	-	Short	2
3	ˈCV.CV.CVC	**- (3.2)**	**+ (6.2)**	+	Short	3
3	CVC.ˈCVC.CVC	**- (4.0)**	**- (5.9)**	-	Short	3
4	CVC.ˈCV.CV.CV	**+ (5.0)**	**+ (5.3)**	+	Short	4
4	CVC.ˈCV.CV.CVC	**- (2.6)**	**+ (5.3)**	-	Short	4
4	CVC.CV.ˈCVV.CV	**+ (4.7)**	**- (5.0)**	+	Long	5
4	CV.CV.ˈCVV.CVC	**- (3.8)**	**- (5.0)**	-	Long	5

Plus sign (**+**) = Frequent, Minus Sign (**-**) = Infrequent, (): Log Frequency

To determine the effect of frequency on the participants’ responses to stimuli in Groups 1-5, a repeated-measures logistic regression was conducted with relative frequency (two levels: frequent and infrequent) as a within-group variable, participant as the random variable to control for individual variability, and the correctness of response (two levels: correct and incorrect) as the dependent variable. The repeated-measures logistic regression was consistently done for each stimulus group across all groups of participants.

The full summary table for the interaction between relative frequency and response accuracy can be found in Appendix 4.2.1. Table 4.5 is extracted from the full summary table found in Appendix 4.2.1 and displays the results that had a relatively clear pattern for the effect of relative frequency. Table 4.5 presents both descriptive and inferential statistics of the analyses. For the descriptive statistics, the mean and the standard deviation of the responses were reported. As for the inferential statistics, the Estimate of the statistical analysis, which indicates the weighted degree of difference in response accuracy between frequent stress pattern and infrequent one, and the p value for the test. A positive estimate value suggests that the average degree of correctness of response for stimuli that had a frequent stress pattern was higher than the infrequent one. In contrast, a negative value in the estimate section suggests that the average correctness of response for stimuli that had infrequent stress pattern was higher than the frequent one. Odds ratio is the effect size statistics, which reports the odds of gain in correctness given a frequent stress pattern. That is, if the odds ratio is bigger than one, it means better odds for the frequent stress pattern to have higher degree of response accuracy, as opposed to having an odds ratio smaller than 1, which would then suggest better odds for the infrequent stress pattern to have higher degree of correctness of response. Based on the predictions made in the study for frequency effect, it is expected that participants will have shorter reaction time and higher response accuracy for stimuli with frequent stress patterns than ones with infrequent stress patterns. Namely, for the frequency effect to be evident, a positive estimate value is expected for response accuracy analysis and a negative estimate value is expected for reaction time analyses.

Table 4.5 Summary of Frequency Effect

on Correctness of Response for Stress Identification Task (Partial)

		Stimulus Group 3		Stimulus Group 4	
		Mean	SD	Mean	SD
AR	Frequent	0.1	0.3	0.39	0.49
	Infrequent	0.12	0.33	0.38	0.49
EN 1st	Frequent	0.55	0.5	0.78	0.42
	Infrequent	0.58	0.49	0.71	0.45
EN 2nd	Frequent	0.52	0.5	0.83	0.38
	Infrequent	0.55	0.5	0.82	0.39
TW 1st	Frequent	0.42	0.5	0.73	0.44
	Infrequent	0.46	0.5	0.65	0.48
TW 2nd	Frequent	0.72	0.45	0.86	0.35
	Infrequent	0.62	0.49	0.77	0.42

	Group 3			Group 4		
	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio
AR	-0.96	$p < 0.01$**	0.39	-0.33	$p = 0.47$	0.72
EN 1st	-0.66	$p < 0.01$**	0.52	-1.3	$p < 0.01$**	0.28
EN 2nd	-0.72	$p = 0.12$	0.5	-1.04	$p = 0.1$	0.36
TW 1st	-0.56	$p < 0.05$*	0.58	-1.91	$p < 0.001$***	0.15
TW 2nd	0.21	$p = 0.58$	1.23	-1.16	$p = 0.17$	0.32

Although a negative estimate value seemed to be the trend for the stimuli in stimulus groups 1,3 and 4, the main effect of relative frequency was insignificant for most groups. Only some groups were found to have a significant difference in performance and these groups were rather scattered across the different groups of stimuli. A more general trend of significant preference did not seem to emerge. However, when significant effects emerge, they are in the unexpected direction. For example, participants' responses to stimulus group 3 were more accurate for stimuli with an infrequent stress pattern for the English- and Chinese-speaking first-year students.

Among these groups of participants, the main effect of relative frequency was significant for the native group and the first-year non-native groups. A similar trend was found in the group 4 stimuli as well, where all groups had a negative estimate value and significant difference was found in the first-year non-native groups. In other words, the results suggest that participants responded more accurately to the stimuli with infrequent stress pattern than ones that had frequent stress pattern. Again, holistically speaking, the main effect of relative frequency was not found to be significant in most of the stimuli when looking at the full summary table in the appendix, and a conclusive pattern did not seem to emerge.

A similar analysis was conducted for the reaction time for the stress identification task. Table 4.5 presents the results for the stimulus groups for which a relatively consistent pattern of frequency effects emerged. In the table, the results were organized according to the stimulus group. For each stimulus group, the table presents the Estimate value, i.e. the mean difference in reaction time between responses frequent and infrequent stress patterns obtained from the post hoc comparisons, the F statistics and the P value from the one-way ANOVA with repeated measures done for each group of participants. The estimate value was calculated by deducting the mean reaction time (in millisecond) to stimuli that had infrequent stress pattern from the mean of the reaction time to stimuli that had frequent stress patterns. For example, if a group of participants had an average of 2330 milliseconds for their reaction time to stimuli with frequent stress pattern and 2960 milliseconds for infrequent stress pattern, the estimate is 2335 minus 2960, which equals -625 as what is shown for the Arabic-speaking participants in their responses to the stimuli in Group 1 in Table 4.5. In other words, a negative estimate would suggest that the average reaction time for the frequent stress pattern is shorter than the average reaction time of the infrequent stress pattern. The same applies to correctness of response; if the difference is negative, it suggests that the degree of correctness of response for stimuli with frequent stress pattern is lower than that of the infrequent stress pattern.

Table 4.6 Summary of Frequency Effect

on Reaction Time for Stress Identification Task (Partial)

		Stimulus Group 1		Stimulus Group 4	
		Mean	SD	Mean	SD
AR	Frequent	3955	3609	3955	3609
	Infrequent	4579	3472	4666	3625
EN 1st	Frequent	2681	1985	2681	1985
	Infrequent	2980	2034	2898	1958
EN 2nd	Frequent	2048	1253	2048	1253
	Infrequent	2489	1689	2383	1512
TW 1st	Frequent	3573	10242	3573	10242
	Infrequent	2966	1804	2978	2240
TW 2nd	Frequent	2741	2044	2741	2044
	Infrequent	4198	3249	3489	2446

	Group 1			Group 4		
	Estimate	F Statistics	P value	Estimate	F Statistics	P value
AR	-625	F(1,134) = 1.8	p = 0.18	-228	F(1,134) = 0.25	p = 0.62
EN 1st	-297	F(1,313) = 2.63	p = 0.11	409	F(1,312) = 4.45	p < 0.05*
EN 2nd	-444	F(1,193) = 7.45	p < 0.01**	497	F(1,194) = 9.26	p < 0.01**
TW 1st	608	F(1,314) = 0.6	p = 0.44	995	F(1,314) = 33.97	p < 0.001***
TW 2nd	-1458	F(1,194) = 22.19	p < 0.001***	1291	F(1,194) = 38.51	p < 0.001***

A consistent pattern was found for the second-year non-native participants' reaction times for the stimuli in stimulus group 1, which were trisyllabic with stress on a short vowel and frequency contrast in both stress pattern and stress position. A trend of shorter reaction time for stimuli with a frequent stress pattern than stimuli with an infrequent stress pattern was found in stimulus group 1. The estimate value for all groups of participants, except for the first-year Chinese-speaking group, was negative. Although the main effect of relative frequency was not significant for the native group (p = 0.18), it was significant for the second-year ($F(1,193)$ = 7.45, p <0.01, partial η^2 =0.05) English-speaking groups. An identical pattern was found the Chinese-speaking groups, where the main effect of relative frequency was significant for the second-year (F(1,194) = 22.19, p < 0.001, partial η^2 = 0.14) group, but not for the first year group. In other words, when responding to stimuli that had a frequent stress pattern, the more experienced non-native speaking participants responded more rapidly than they did to stimuli with an infrequent stress pattern, conforming to the predictions for the frequency effect.

However, unexpectedly, a reverse trend was found for the group 4 stimuli, which were quadrisyllabic with stress on a short vowel and frequency contrast only in stress pattern. All groups of participants, except for the native group, had a positive estimate value, indicating that the reaction time was longer for stimuli with more frequent stress pattern than ones with the infrequent stress pattern. The ANOVA test found the main effect of relative frequency to be significant in nearly all groups of participants, except for the native group. Thus, participants spent more time on stimuli with a frequent stress pattern than ones with an infrequent stress pattern, which was contrary to our previous finding for group 1 stimuli, disagreeing with the prediction for the frequency effect.

As for the responses for stimuli in stimulus groups 2, 3, and 5, the trend for frequent stress pattern was not as strong. For stimulus group 2, while there was a significant effect of relative frequency on reaction time for the second-year English-speaking group, the second-year Chinese-speaking groups with significantly shorter reaction time for stimuli with the frequent stress pattern, the effect of relative frequency was not significant in the remaining groups. Similarly, for stimulus group 3, significant difference in reaction time was only found in the second-year Chinese-speaking groups with significant shorter reaction time for stimuli with frequent stress patterns; no significant effect of relative frequency was found in the remaining groups. No significant effect of relative frequency was found for any group of participants for stimulus group 5 either. Although the significant differences in reaction time found for stimulus groups 2, 3 and 5 all conformed to the prediction for frequency effect, the effect of relative frequency in these groups seemed rather scattered (especially in stimulus groups 3 and 5) and therefore did not provide strong evidence to support the role of frequency in performance.

In short, this section did find the main effect of relative frequency to be significant on both accuracy and fluency; however, the significant differences seem to be found more locally, i.e., within a given stimulus group. Unlike the effect of native language found in Section 4.3.1 (Native language and Performance) and Section 4.3.3 (Stress pattern and performance) part of this section, an across-the-board effect of frequency that applied to all groups of participants and stimuli was not found. Moreover, the preference for frequent vs. infrequent stress patterns diverged at times. Recall that the usage-based perspective on language would predict a shorter reaction time and greater accuracy of response for stimuli that had frequent stress pattern than ones with infrequent stress pattern. In the stress identification task, such preference was not consistently found. For example, most of the participants had significantly shorter reaction times for group 1 stimuli and few others in stimulus groups 2 and 3. However, the participants had a reverse preference for the group 4 stimuli, as they responded more slowly to stimuli with a frequent stress pattern than with an infrequent stress pattern. The trend for correctness of response also seemed to contradict the predictions for frequent stress pattern, as the participants seemed to respond less accurately to stimuli with a frequent stress pattern than ones with an infrequent stress pattern, although this pattern was less obvious than the trend found for reaction time. Taken together, there was evidence that conforms to the predictions made for the frequency effect and evidence against these predictions. It seemed to be case that there was more evidence in support of a frequency effect in the reaction time analyses, while the accuracy analyses offer a rather inconclusive evidence to support the prediction for a frequency effect.

4.3.5 Summary of Results

Table 4. 7 presents a summary of the results for the stress identification task according to the three main variables analyzed in this study. Participants in the native group, i.e. the Arabic-speaking group, responded to the stimuli significantly more slowly and less accurately than the participants in the English-speaking and Chinese-speaking groups. With more detailed multiple comparisons by the type of stress pattern, it was found that the main effect of native language remained strong and seemed globally applicable to most stress patterns. The multiple comparisons also showed that proficiency had little effect either on the reaction time or the correctness of participants' responses in the stress identification task. Nonetheless, the multiple

comparisons did find some traces in the Chinese-speaking group showing the advantage of higher proficiency on reaction times to specific stress patterns.

Table 4.7 Summary Table for Findings of Stress Identification Task

	REACTION TIME	RESPONSE ACCURACY
Native Language	1. Longer RT for native group than the non-native groups 2. No difference between the two non-native groups	1. Lower correctness rate for native group than the non-native groups 2. No Difference between the two non-native groups
Proficiency	No Effect	No Effect
Frequency	1. Shorter RT for stimulus with frequent stress pattern in stimulus group 1 2. Longer RT for stimulus with frequent stress pattern in stimulus group 4	1. Lower accuracy for stimuli with frequent stress pattern in stimulus group 3 and 4 2. Less strong trend than the results for reaction time

The analyses of the frequency effect showed mixed results for the stress identification task. On one hand, the results indicated that some participants had significantly shorter reaction times for stimuli that were trisyllabic with a frequent stress pattern than for stimuli with an infrequent stress pattern. On the other hand, some groups of participants had significantly longer reaction times for the stimuli in group 4 with a frequent stress pattern. Such mixed results extended to the correctness of response, as a significantly lower response accuracy was found in some groups of participants for some stimuli that were trisyllabic and some that were quadrisyllabic. In short, the significant differences found for both reaction time and correctness of response were rather scattered amongst different groups of participants, and no obvious pattern emerged from the results.

Another major finding from the stress identification task is that the native speakers performed significantly more poorly than the non-native speakers. They responded more slowly and less correctly to the stimuli for nearly all stress patterns. This finding contradicted the general expectation of native speakers, as they are normally associated with faster and more accurate production, which was not the case for their performance in the stress identification task. A detailed discussion of this finding is provided in the Chapter 5.

4.4 Lexical Decision Task

The second main task of the experiment was the lexical decision task, which was conducted together with the stress identification task. That is, immediately after participants completed a trial in the stress identification task, a screen was prompted in the testing interface, asking the participants whether they thought the word that they just heard in the trial was a real Arabic word or a made-up one. The participants were asked to press the buttons on the keypad to indicate their response. Similar to the measures used in the stress identification task, the participants' responses were measured by the reaction time of the response, i.e., the duration of time (measured in milliseconds) that it took the participants to respond, and the correctness of the response. Both measures were automatically captured by the testing program, PsychoPy. No further calculation was implemented for the reaction time and correctness of response. As in the stress identification task, the focus of this task was not on the real words in the stimuli; instead, the nonsense words in the stimuli were the focus of this task. Therefore, the data used for the following analyses were measures of the participants' responses to stimuli that were nonsense words.

4.4.1 Native Language and Performance

Figure 4.8 shows the response accuracy according to participants' native language. A clear difference can be seen between the native group and the non-native groups, as the former had relatively higher (almost ceiling) accuracy in the task, while the latter had difficulty in the task as their accuracy was below-chance (less than 50%). A one-way ANOVA was conducted to examine the extent of this difference, with native language (Arabic, English and Mandarin Chinese) as a between-group variable and response accuracy (in proportion correct) as the dependent variable. The analysis revealed that there was significant effect of native language ($F(2,73) = 26.2$, $p < 0.001$, $\eta^2 = 0.41$) on response accuracy. Post-hoc comparisons using Tukey tests showed that the Arabic-speaking group outperformed the two non-native groups with significantly higher response accuracy ($p < 0.001$). However, no significant difference was found between the English-speaking groups and the Chinese-speaking groups.

Figure 4.8. Boxplot for the Response Accuracy for the Lexical Decision Task Accordingly to Participants' L1

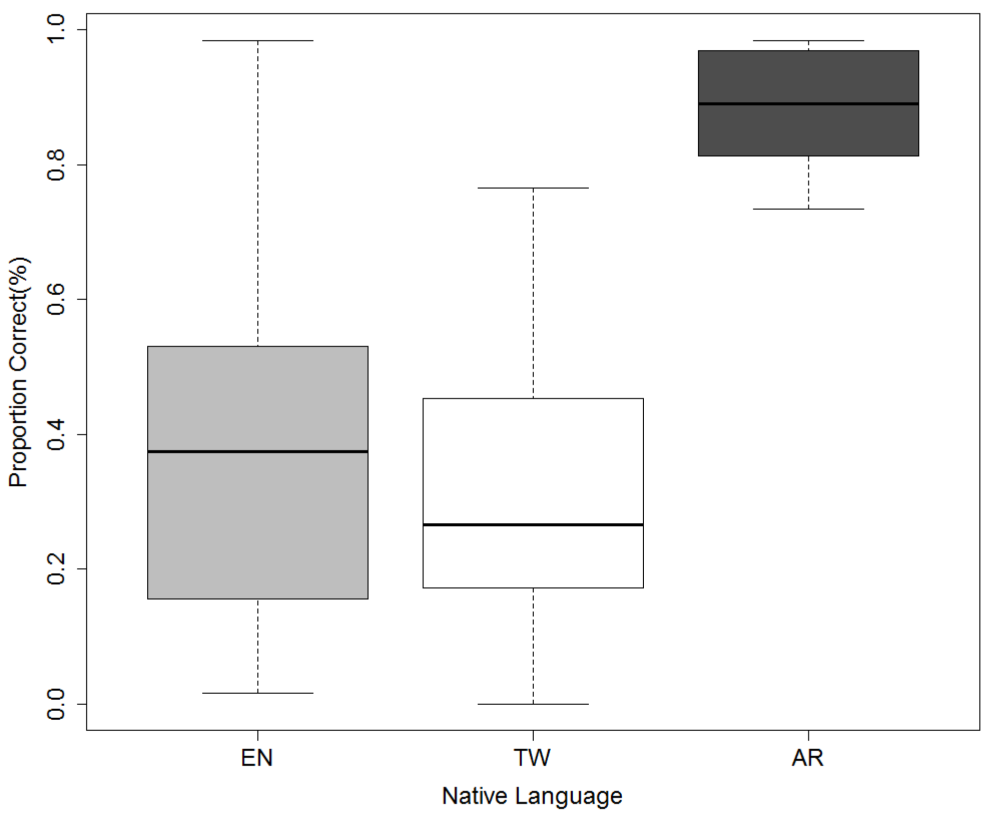
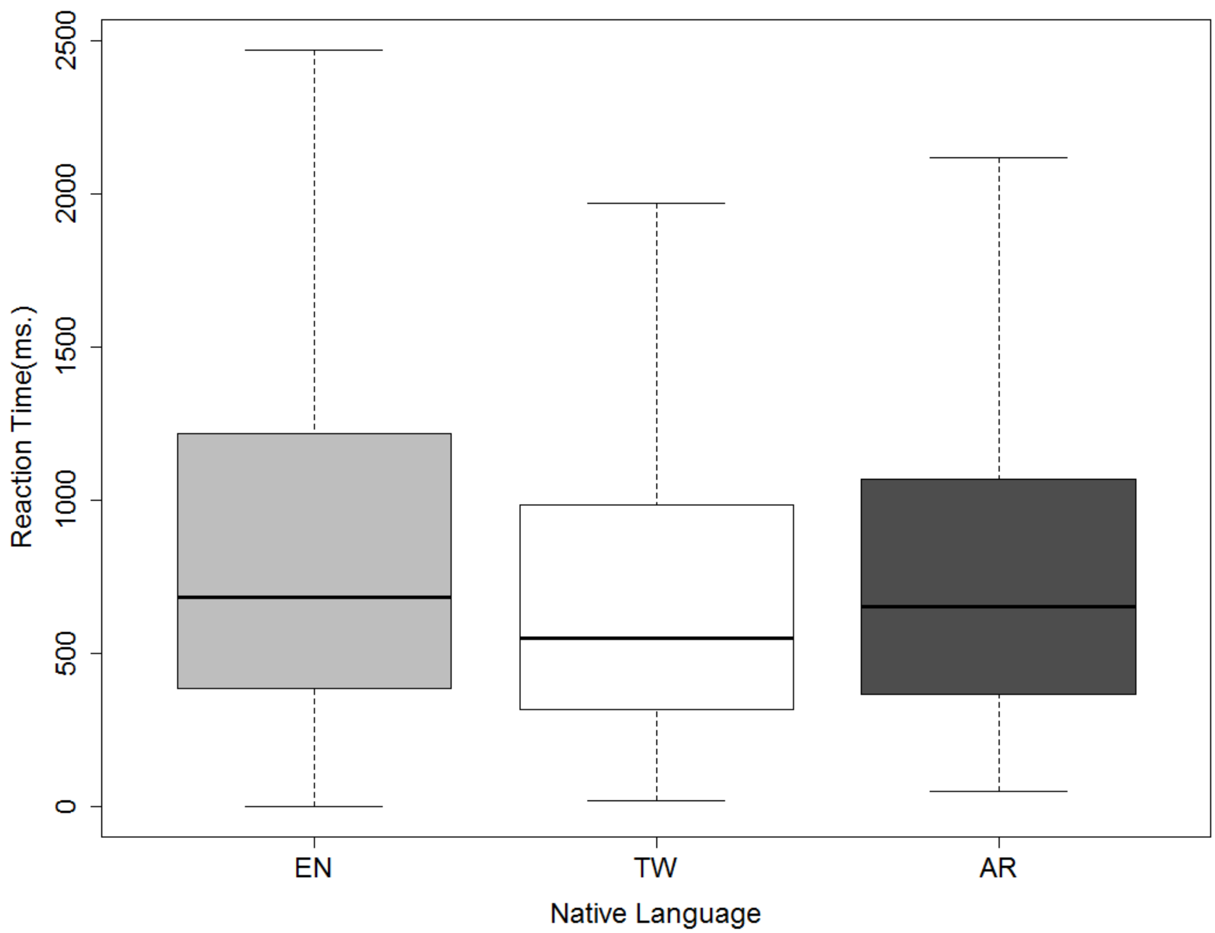


Figure 4.9. presents the average reaction times for each language group. A one-way analysis of variance (ANOVA) with reaction time as dependent variable was conducted. However, the ANOVA test did not find the main effect of native language to be significant ($F(2,73) = 0.73$, $p = 0.48$), suggesting that the participants took similar time to respond to the stimuli in the task, regardless of their native language.

Figure 4.9. Boxplot for the Reaction Time
Grouped by Native Language for the Lexical Decision Task

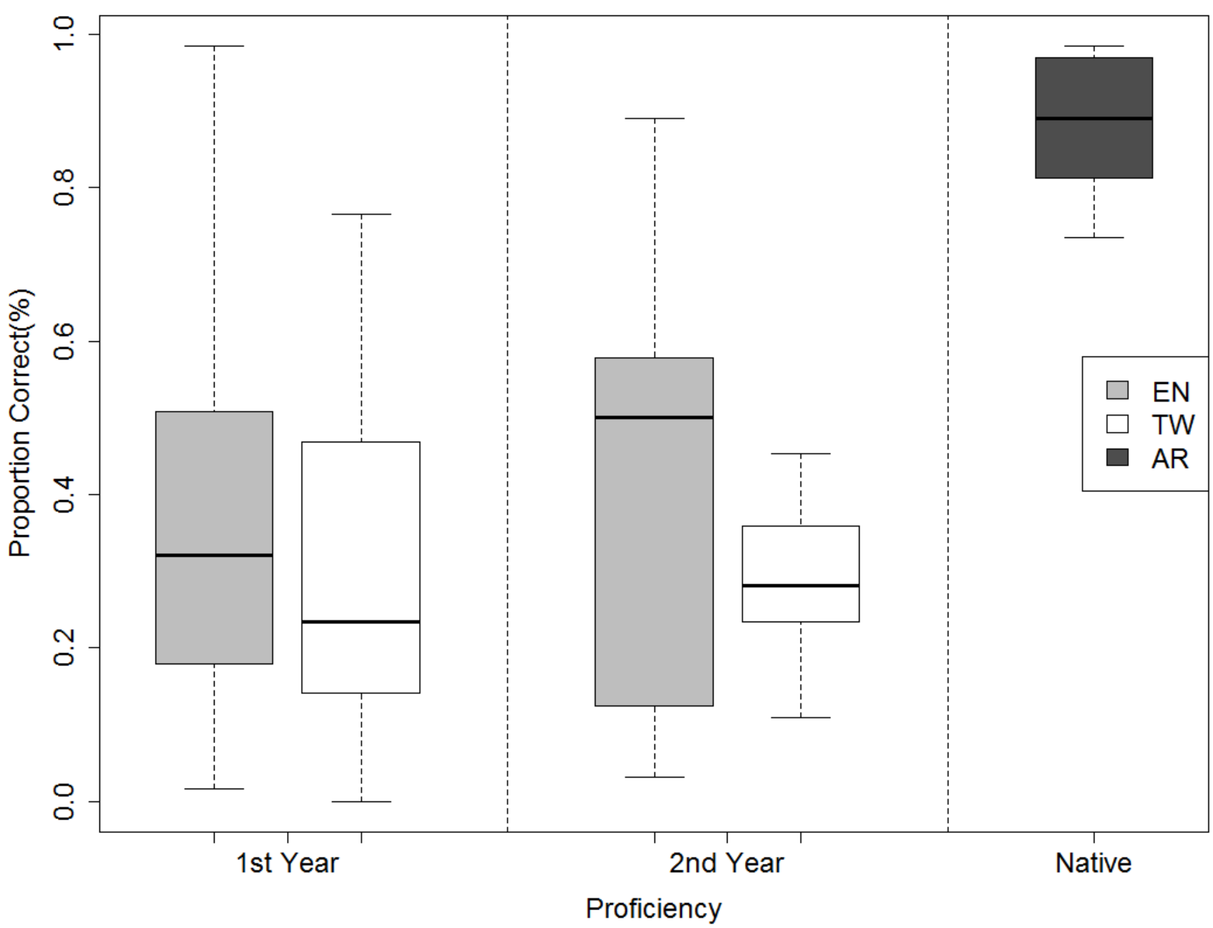


In short, the analyses in this section showed that although the Arabic-speaking participants spent similar time to respond to the stimuli as the participants in the non-native groups, their responses were significantly more accurate than their non-native counterparts.

4.4.2 Proficiency and Performance

To further investigate the effect of proficiency on the response accuracy, Figure 4.10 shows the reaction times according to participants' native language as well their level of proficiency compared with participants in the control group. Figure 4.10 shows that the average response accuracy increases along with the proficiency, as second-year participants had higher average accuracy than their first-year counterparts. However, the overall performance was low, as the average response accuracy remains only around chance even for the second-year participants.

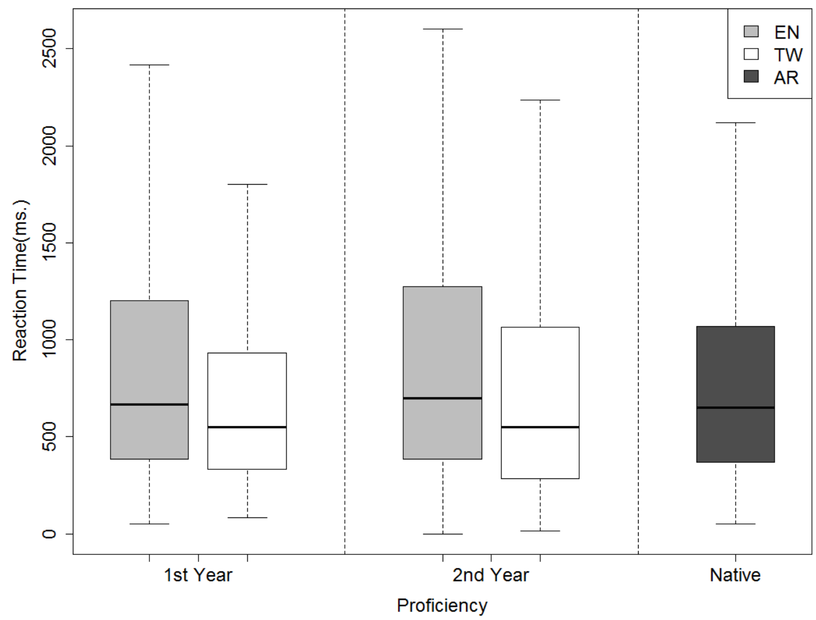
Figure 4.10. Boxplot for the Correctness Rate
Grouped by Participants' L1 and Proficiency for the Lexical Decision Task



To examine the extent of the difference in response accuracy in the non-native groups, a two-way ANOVA was conducted with native language as a between-group variable (two levels: English and Mandarin Chinese), proficiency (two levels: first-year and second-year) as a within-group variable and response accuracy as the dependent variable. The results indicate that there was no significant main effect of proficiency nor an interaction between native language and proficiency, suggesting that the non-native participants had similar response accuracy in the task, regardless of their proficiency.

Figure 4.11 shows the reaction times according to participants' native language and their proficiency. A similar two-way ANOVA was conducted to examine to effect of proficiency on reaction time, with native language as a between-group variable (two levels: English and Mandarin Chinese), proficiency (two levels: first-year and second-year) as a within-group variable and reaction time as the dependent variable. Again, there was not a significant effect of proficiency nor an interaction between native language and proficiency, suggesting again that higher proficiency did not help the non-native participants respond to the stimuli more quickly.

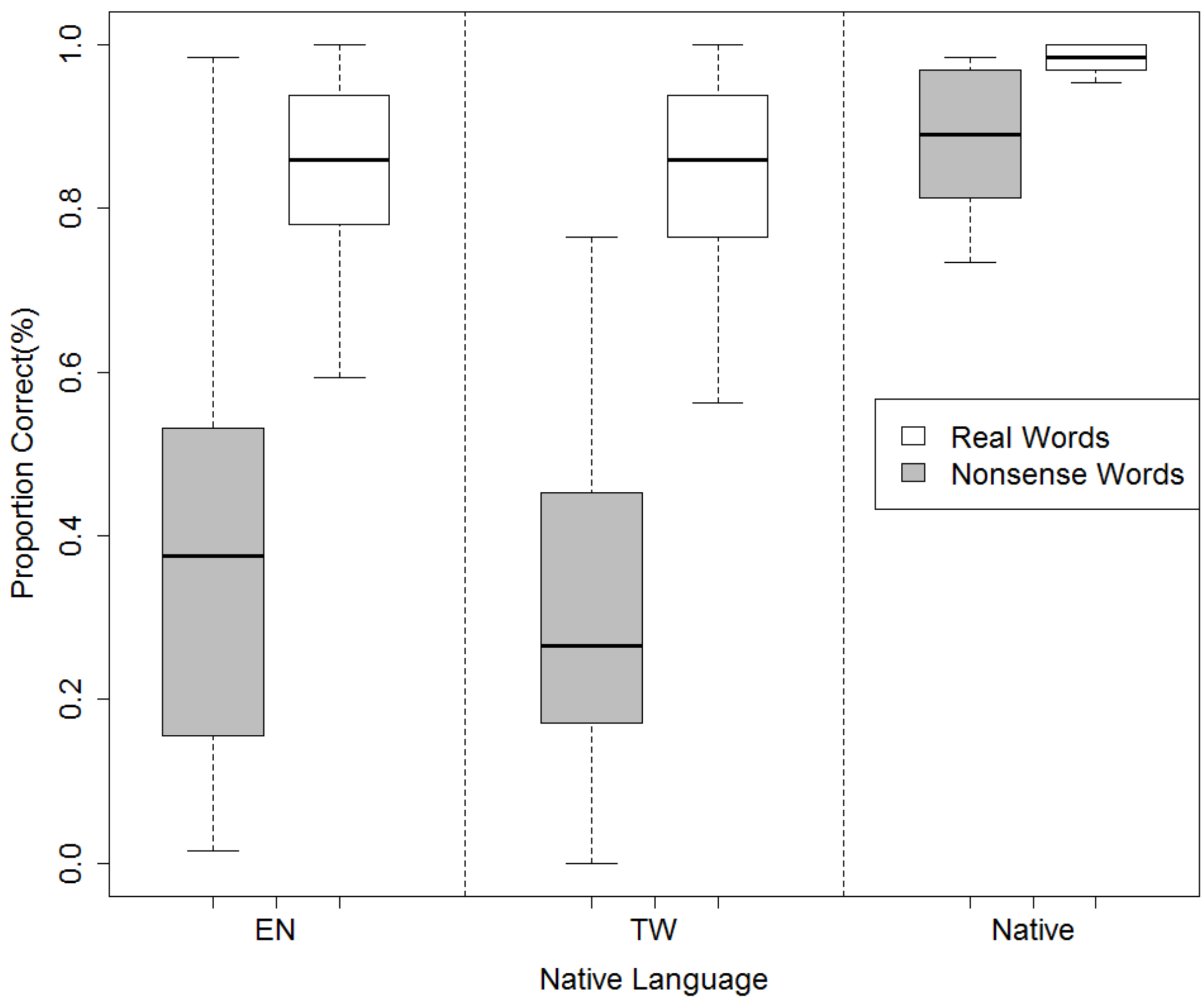
Figure 4.11. Boxplot for the Reaction Time
Grouped by Participants' L1 and Proficiency for the Lexical Decision Task



Finally, although it is not the focus of the present study, it is essential to examine participants' responses to stimuli that are real words in the task. The motivation for attending to real word is that lexical decision, unlike other tasks in the experiment, could yield an inflated accuracy rate for nonsense word if listeners simply always respond "no" (i.e., not a word). If the participants only press "yes" to all the stimuli, they will receive 100% proportion correct on stimuli that are real words and 0% on stimuli that are nonsense words. Responses as such will inevitably introduce bias as the score does not reflect their actual performance on the task, which will negatively influence subsequent analyses these data.

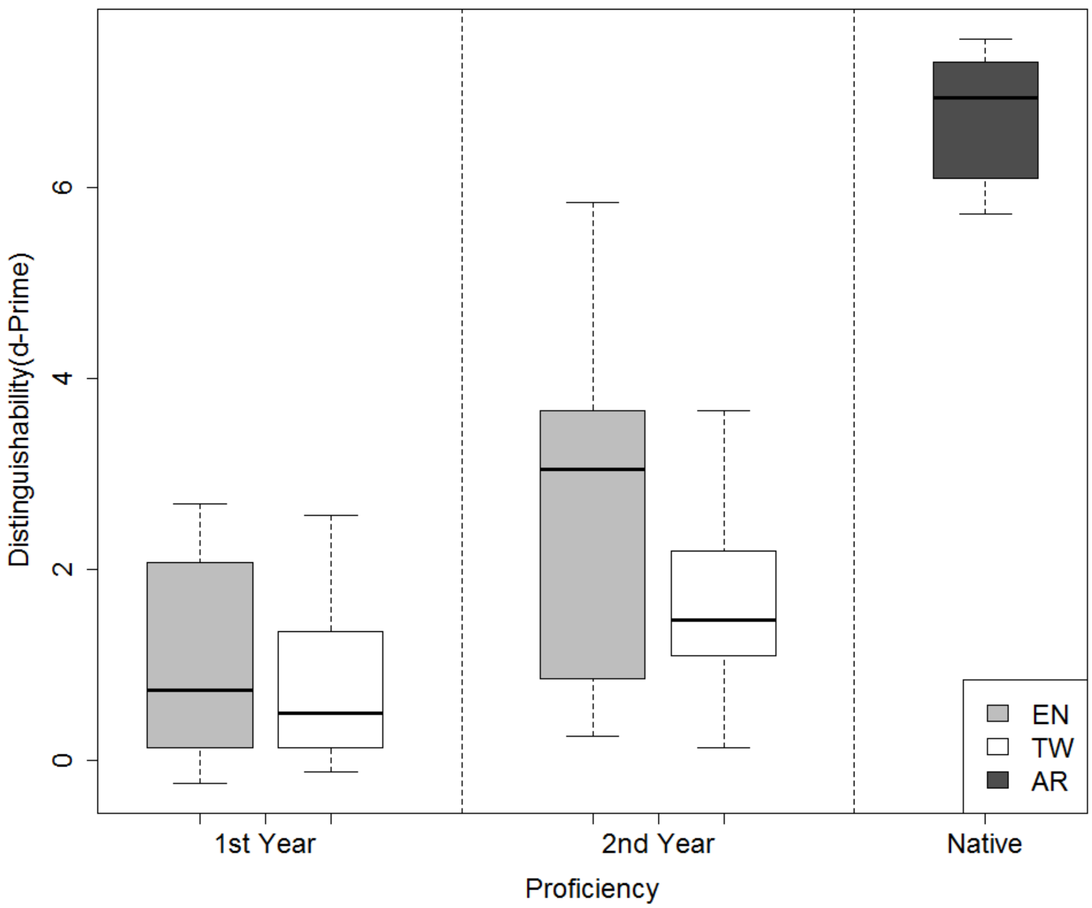
Figure 4.12 shows the response accuracy according to the native language of the participants as well as the lexical status of the stimuli (i.e. whether they are real Arabic words or not). The comparison shows that the native speakers had similar response accuracy for both real and nonsense words, suggesting that they were attending to the stimuli rather than pressing the same button with attending to the stimuli. However, the contrast in the non-native groups was alarming, as a significant difference was found between the participants' response accuracy to real words and nonsense words. As the figure suggests, the non-native participants had higher proportion correct on nonsense words than real ones, which suggests that some of the participants might have pressed "no" without actually processing the stimuli.

Figure 4.12. Boxplot for the Response Accuracy for the Lexical Decision Task According to L1 and Lexical Status of Stimuli



This biased performance calls for a more detailed analysis of participants' responses in the lexical decision task. By converting their responses into *d*-prime scores, it allows a better observation of whether the participants were indeed able to distinguish real word from nonsense words. Figure 4.13 presents the *d*-prime scores for each participant group. A one-way ANOVA was conducted with native language (three levels: Mandarin Chinese, English and Arabic) as independent variable and *d*-prime score as dependent variable. The test indicated that there was a significant effect of native language ($F(2,74) = 82.6$, $p < 0.001$, $\eta^2 = 0.69$) on the *d*-prime score. Post-hoc Tukey tests found that the participants in the control group had significantly higher ($p<0.001$) *d*-prime scores than the non-native participants. Another two-way ANOVA examining the performance of the non-native groups with native language (two levels: Mandarin Chinese and English) and proficiency (two levels: 1st year and 2nd year) as independent variables and *d*-prime score as dependent variable found a significant main effect of proficiency ($F(1,64) = 16.3$, $p < 0.001$, $\eta^2 = 0.20$) on d-prime scores while there was no interaction between native language and proficiency. Post-hoc Tukey tests indicated that the second-year participants had significantly higher *d*-prime scores than their first-year counterparts.

Figure 4.13. Boxplot for the D-prime Score for the Lexical Decision Task According to L1 and Proficiency



In other words, this analysis found that the native speakers were rather confident in distinguishing the real words from the nonsense ones in the stimuli. In contrast, although improvement was found as the proficiency of the participants increases, the non-native participants' occasionally accurate responses might be due to chance or habitually pressing the same key, but not their lexical knowledge about the stimuli. In other words, it has to be noted that the results yielded from the non-native participants in the task should be taken with a grain of salt, as the results might be biased and might not be completely reflective of their actual lexical knowledge due to their relative lower d-prime scores. However, the findings regarding the native speakers will remain credible since they demonstrated consistent performance across real and nonsense words with their high d-prime scores.

In sum, this section found no effect of proficiency for the non-native participants. No significant difference was found in either reaction or accuracy between participants of different proficiency levels. Taken together, the analyses conducted in this section showed that the participants in the native group were significantly more accurate in determining whether a stimulus is a real Arabic word or not than their non-native counterparts. However, they took similar amount of time to make such decision as their non-native counterparts. As for the non-native participants, their performance was comparable, as they have similar reaction time and response accuracy, regardless of their proficiency. Lastly, due to non-native participants' relatively lower d-prime scores (suggesting lower reliability), the following analyses focus more on the performance of the native speakers than on that of the non-native speakers.

4.4.3 Stress Pattern and Performance

To delve more into the effect of native language and proficiency, more in-depth analyses were conducted with subgroups of stimuli, which was based on the type of stress patterns of the stimuli. A similar set of ANOVAs (as discussed in Section 4.3.3) was conducted for the participants' responses to the eight types of stress pattern

The summary table for all tests conducted for response accuracy and reaction time is presented in Appendix 4.3.1 and 4.3.2 respectively. The analyses showed a consistent effect of native language on response accuracy, while the effect of proficiency as well as an interaction between native language and proficiency was not found. The post-hoc Tukey tests found that the participants in the native group were significant more accurate in the task than their non-native participants in all stress patterns. As for reaction time, the tests did not find an effect of native language, proficiency, or an interaction between native language and proficiency across all types of stress patterns.

In sum, the results confirm the findings from Section 4.4.1 (Native language and performance) and Section 4.4.2 (Proficiency and performance) for the lexical decision task. That is, although the participants in the control group did not respond to the stimuli more quickly than their non-native counterparts, they did maintain a strong advantage for determining whether a word is real or not, regardless of its stress pattern. Additionally, the non-native participants did not exhibit any difference in their performance, regardless of their native language and their proficiency levels.

4.4.4 Relative Frequency and Performance

A rather interesting pattern emerged upon examining the response accuracy in the lexical decision task to nonsense words. Following previous analyses of correctness of response, a repeated-measures logistic regression was conducted with relative frequency (two levels: frequent and infrequent) as a within-group variable, participant as the random variable to control for individual variability, and the correctness of response (two levels: correct and incorrect) as the dependent variable. The analysis was consistently done for all groups of stimuli across all groups of participants. The full summary table for analyses is presented in Appendix 4.4.1. Table 4.8 presents a portion of the full summary table where a more distinct pattern was observed.

For the stimuli in stimulus group 1, the analyses found that all groups of participants, except for the first-year Chinese-speaking group, had a negative estimate value and the effect of relative frequency was significant. A negative estimate value for correctness of response indicated that the participants responded less correctly to stimuli with a frequent stress pattern and more correctly to stimuli with an infrequent stress pattern. In other words, for the stimuli in this stimulus group, the participants were significantly more correct when responding to stimuli with an infrequent stress pattern than to stimuli with a frequent stress pattern.

Table 4.8 Summary of Frequency Effect

on Correctness of Response for Lexical Decision Task (Partial)

		Stimulus Group 1	
		Mean	SD
AR	Frequent	0.78	0.42
	Infrequent	0.92	0.28
EN 1st	Frequent	0.26	0.44
	Infrequent	0.44	0.5
EN 2nd	Frequent	0.27	0.45
	Infrequent	0.42	0.5
TW 1st	Frequent	0.31	0.46
	Infrequent	0.35	0.48
TW 2nd	Frequent	0.17	0.38
	Infrequent	0.5	0.5

	Group 1		
	Estimate	P Value	Odds Ratio
AR	-1.35	p < 0.01**	0.26
EN 1st	-1.08	p < 0.001***	0.35
EN 2nd	-0.81	p < 0.01**	0.45
TW 1st	-0.24	p = 0.36	0.8
TW 2nd	-1.72	p < 0.001***	0.19

Although this preference for stimuli with an infrequent stress pattern was also found for other stimulus groups, it was not as strong or consistent. For stimulus group 2-5, there was one or two groups of participants (e.g., native group and the second-year English-speaking group for stimulus group 2, the first-year, and the second-year Chinese-speaking group for stimulus group 3, first-year Chinese-speaking group for stimulus group 4 and 5) within each stimulus group that had a significant difference in accuracy. Notably, these groups of participants all consistently had significantly lower correctness of response for stimuli with a frequent stress pattern than stimuli with an infrequent stress pattern. In other words, these groups of participants were more successful in determining whether a word is real or not when the stimuli had infrequent stress, which again disagrees with the predication for the frequency effect.

Following the analyses conducted in the stress identification task, a one-way ANOVA with repeated measures was conducted for each stimulus group to examine the role of frequency in participants' performance in the task. The full summary table for all analyses conducted is given in Appendix 4.4.2. The results showed a rather mixed trend of estimate value. Although there seems to be a trend of negative estimate value, which would suggest shorter reaction times to stimuli with a frequent stress pattern than stimuli with an infrequent stress pattern, these differences were not significant for most stimulus groups. Even when significant differences were found, the trend was not immediately obvious. For example, the first-year and second-year Chinese-speaking participants had significantly shorter reaction times to stimuli with a frequent stress pattern for stimulus group 5, but similar significant differences were not found for other groups of participants. Therefore, the results do provide sufficiently decisive evidence for a frequency effect.

In sum, the analyses in this section show that there was no overall pattern for an effect of relative frequency on reaction time, as significant differences were only found for a few groups of participants within a rather small number of stimulus groups. In other words, a more frequent stress pattern did not help the participants determine whether a word is real or not more quickly. On the other hand, for response accuracy, the analyses found that the participants were significantly more successful in determining whether a stimulus was a real word or not when responding to stimuli with infrequent stress pattern than stimuli with frequent stress pattern, which is the opposite of what was predicted for frequency effect.

4.4.5 Summary of Results

To summarize this section, Table 4. 9 presents the findings from the lexical decision task. Based on the data examined in this section, the effect of native language for non-native groups was not significant, as no difference in either reaction time or correctness of response was found between the English-speaking groups and the Chinese-speaking groups. Comparing the non-native groups and the native group, the results show that the participants in the native group were more successful in determining whether stimuli were real words or not than both non-native groups. Nevertheless, such an advantage did not extend to reaction time, as the native group took longer to respond to the stimuli than did the non-native groups. The results also show that there was not a significant effect of proficiency on either reaction time or correctness of response. In general, higher proficiency did not help the non-native participants to respond faster or to accurately judge whether the nonsense words were real Arabic words or not.

Table 4.9 Summary Table for the Findings from Lexical Decision Task

	REACTION TIME	CORRECTNESS OF RESPONSE
Native Language	No effect	1. Higher response accuracy for native group than non-native groups
Proficiency	No effect	No effect
Frequency	No effect	1. Strong trend for lower CR for stimuli that had frequent stress pattern for stimulus group 1

The multiple comparisons confirm these findings, as a significant effect of native language was consistently found for all types of stress patterns. In contrast, the effect of proficiency was insignificant across nearly all types of stress patterns, which strongly suggests that the proficiency of participants had little effect on their performance. However, for the few instances when proficiency mattered, participants' responses did demonstrate the advantage of having higher proficiency when responding to stimuli with specific types of stress pattern.

Upon examining the role of frequency in performance, relative frequency was not found to be a significant indicator of reaction time. Only a few groups of participants were found to demonstrate a significant difference in reaction time, and these differences were scattered across a limited number of stimulus groups. On the other hand, relative frequency was significant for participants' responses to stimulus group 1. The participants were found to respond less correctly to the stimuli that had a frequent stress pattern than stimuli with an infrequent stress pattern. Such preference was consistently found in all groups of participants, regardless of their native language and proficiency in stimulus group 1. Considering that the stimuli examined here were all nonsense words, the participants were expected to reject them in their responses. Since real words were not considered in the analyses, instead of evaluating participants' ability to both accept real words and reject nonsense words, the analyses seemed to only evaluate their ability to reject nonsense words due to the absence of responses to real words in the analyses. This way, the fact that the participants were less correct in their responses means that they considered the nonsense words to be real Arabic words thus did not reject them. Our finding then should be interpreted as that the participants were significantly less inclined to reject nonsense words with frequent stress patterns than ones with infrequent stress patterns. Another way of seeing this result is that there is a higher chance for the participants to mistake a nonsense word as a real word if the word had a frequent stress pattern, and vice versa.

It is worth noting that further examination of participants' responses to both real and nonsense words found that the non-native participants had significantly higher response accuracy for nonsense words than for real words. Detailed analyses using d-primes scores also confirm this outcome, suggesting that non-native participants' higher response accuracy to nonsense word is more likely to be due to the fact that they simply responded to all stimuli with "no" (i.e., not a real word) without actually processing the stimuli, which might not completely reflect their actual lexical knowledge. In contrast, the native participants in the control group did not have such bias, and were consistently accurate for stimuli that are either real or nonsense words. The native-speaking participants' unbiased performance in the task renders the aforementioned finding of frequency effect on participants' misjudgment of stimuli that are nonsense words to be real words credible .

The main finding from the lexical decision task was the global effect of native language in that the participants in the native control group were significantly more accurate but not faster in their responses compared to the non-native groups in nearly all analyses. Moreover, the proficiency of the non-native-speaking participants generally did not help promote either fluency or accuracy of responses in the task, with few exceptions where higher proficiency help improve the accuracy of the Chinese-speaking group on the task.

The frequency effect seems to locally (i.e., for specific stimulus group) influence participants' response accuracy, but not the reaction time. Within these stimulus groups (stimulus group 1 in particular), participants were more successful in determining whether a given word was a real word or not when the word had an infrequent stress pattern. This preference for the infrequent, or lack thereof for the frequent, was somehow understandable as it is hard for listeners to reject familiar-sounding nonsense words due to their frequent stress patterns. This finding will be further discussed in Chapter 5.

Lastly, due to the biased identified from non-native participants' biased response accuracy towards nonsense word, the results presented above shall be taken with a grain of salt.

4.5 Stress Production Task

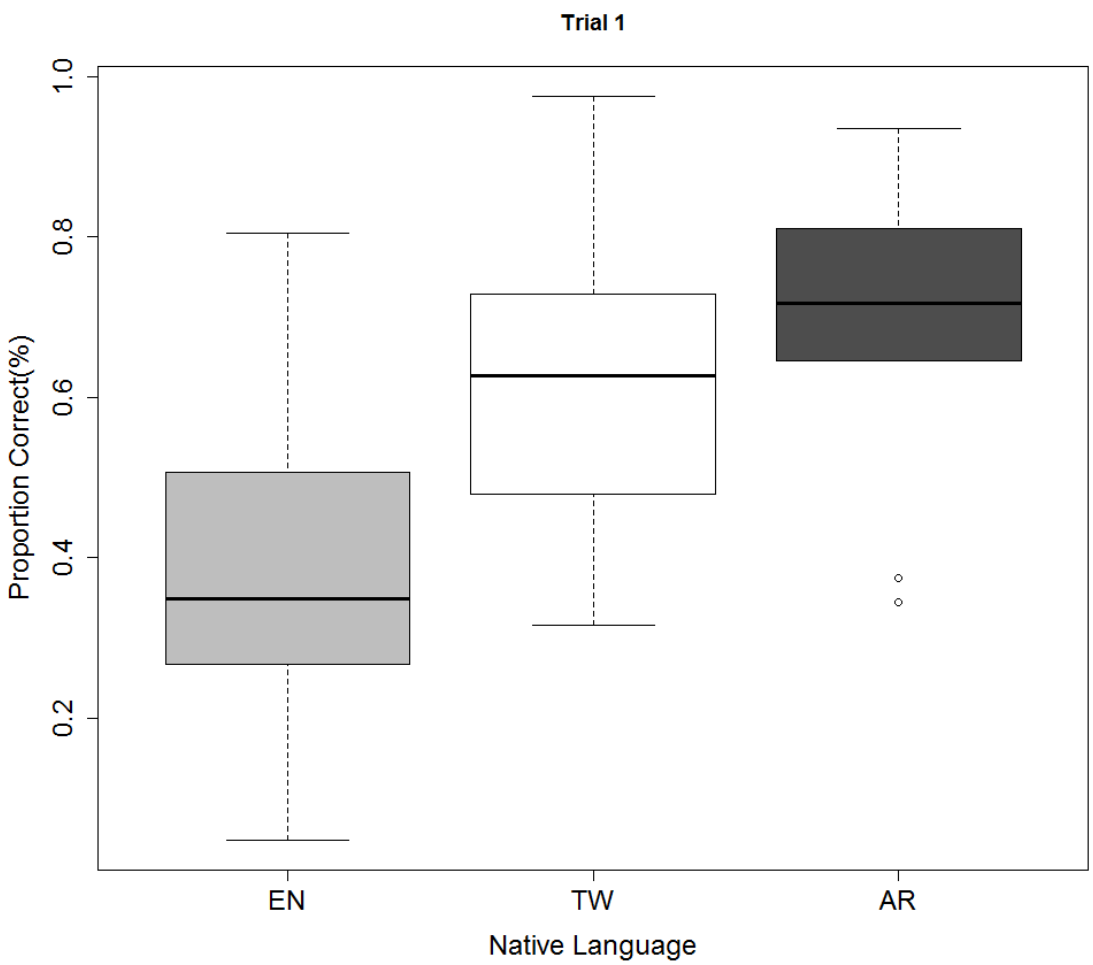
The stress production task was the third and the last main task in the experiment. The purpose of this task was to examine the production of the participants, particularly the participants' ability to place stress on the correct syllable. In this task, the participants were presented with visual stimuli on the computer screen, which were in Arabic script with full vocalization. The participants were prompted with a beeping sound by the testing program to read the word printed on the screen as quickly as they could. For every token in the task, the script for each stimulus appeared twice consecutively, requiring the participants to read out each word twice. In this section, the first attempt will be referred as trial 1 and the second attempt as trial 2 for the stimulus. Their production was recorded and analyzed. Following previous tasks, the measures used in this task were the reaction time of the response as well as the correctness of the response. The response time was determined by calculating difference between the time tag for each trial and the onset of their production (measured in milliseconds). The correctness of stress production was determined by the scoring scheme proposed in Chapter 3. The scheme extracted and compared the three correlates, i.e. F0, duration, and intensity, for each syllable of the production of the participants. The scoring schema evaluates and ranks each syllable of the production based on the three correlates. The syllable with the highest score is predicted to be the stressed syllable produced by the participants, which is then compared with the attested position of stress for the stimulus and determine whether the production was either correct (coded as 1 in the dataset), or incorrect (coded as 0 in the dataset). Recall that the analyses for this task were based on participants' responses to stimuli that were nonsense words, to control for any bias that might be introduced by an individual participant's previous familiarity with any words that might have been included in the stimuli.

To compare participants' performance in the two trials, two ANOVA tests were conducted with the native group and the non-native groups respectively. For the control group, a one-way ANOVA was conducted with trial number (two levels: trial 1 and trial 2) as the between-group variable and response accuracy/reaction time as dependent variable. The results show that trial number is not a significant indicator of the native speakers' performance. As for the non-native groups, a two-way ANOVA was conducted for each non-native group, with trial number as between-group variable, proficiency as within-group variable and response accuracy /reaction time as dependent variable. Again, the test did not find a significant main effect of trial number on performance either. Since the difference between the participants' first and second production attempts were in most cases not significant, the present study will focus on the analysis of the first attempt. Moreover, the first production attempt is arguably less rehearsed and may be more reflective of participants' typical performance.

4.5.1 Native Language and Performance

Following the sequence of the analyses presented in previous sections, this section first analyzes the influence of participants' native language on stress production. Figure 4.13 illustrates the response times according to participants' native language. As shown in Figure 4.14, the participants on the control group have the highest average response accuracy, which is above chance, suggesting that the native speakers, as expected, were able to place stress on the correct syllable of the word. The Chinese-speaking participants also demonstrated above-change response accuracy, even though the proportion correct is slightly lower than that of the native speakers. Lastly, the English-speaking participants seemed have difficulty in the task, as their response accuracy is relatively low.

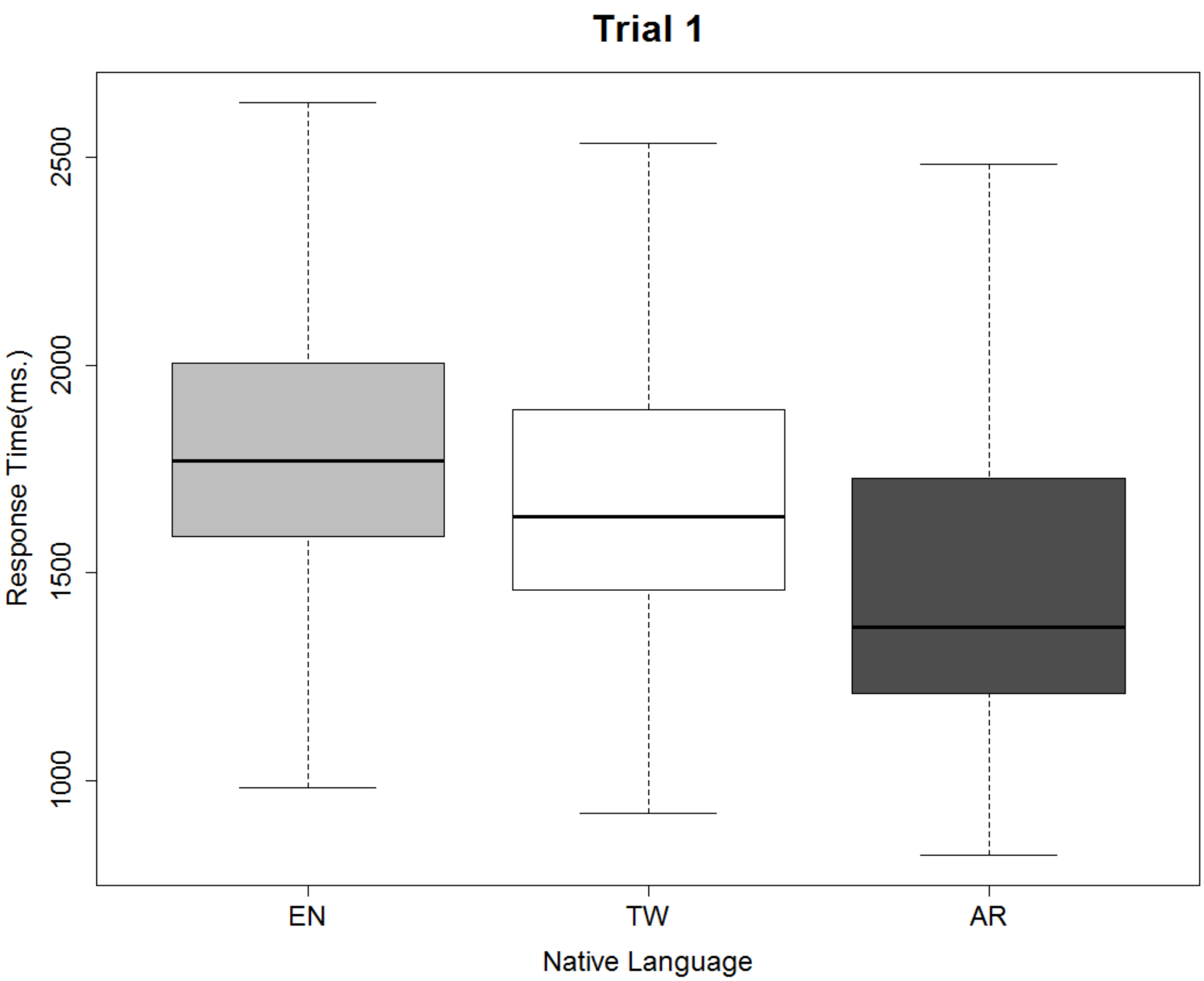
Figure 4.14. Boxplot for the Response Accuracy for ohe Stress Production Task According to Native Language



A one-way ANOVA was conducted with native language as a between-group variable (three levels: English, Mandarin Chinese and Arabic) and response accuracy as dependent variable to detect the extent of the differences observed in Figure 4.11. The ANOVA test found a significant main effect of native language on response accuracy ($F(2,73) = 27.09$, $p < 0.001$, $\eta^2 = 0.42$). The post-hoc Tukey tests further showed that both the Arabic-speaking group and the Chinese-speaking group were significantly more accurate ($p < 0.001$) than the English-speaking group in the stress production task. Moreover, no significant difference in response accuracy was found between the Arabic-speaking group and the Chinese-speaking group. In other words, the results suggested that the Arabic-speaking participants and the Chinese-speaking participants produced stress on the correct syllable more than the English-speaking participants did. The results also showed that the Chinese-speaking group were as accurate as the native speakers on the task.

Figure 4.15 provides the reaction times according to participants' native language. The figure shows that the native group took less time to initiate the production than the non-native groups. Between the non-native groups, the Chinese-speaking participants took less time to begin their production than the English-speaking group. To examine the extent of these differences, a similar one-way ANOVA was conducted with reaction time as the dependent variable. The test shows that there was a significant effect of native language on reaction time ($F(2,73) = 9.29$, $p < 0.001$, $\eta^2 = 0.20$). The post-hoc comparisons using Tukey tests showed that both Arabic-speaking and Chinese-speaking group had significantly shorter ($p < 0.001$) reaction times than than the English-speaking group, while no significant difference was found between the Chinese-speaking and the native group.

Figure 4.15. Boxplot for the Reaction Time for the Stress Production Task According to Native Language

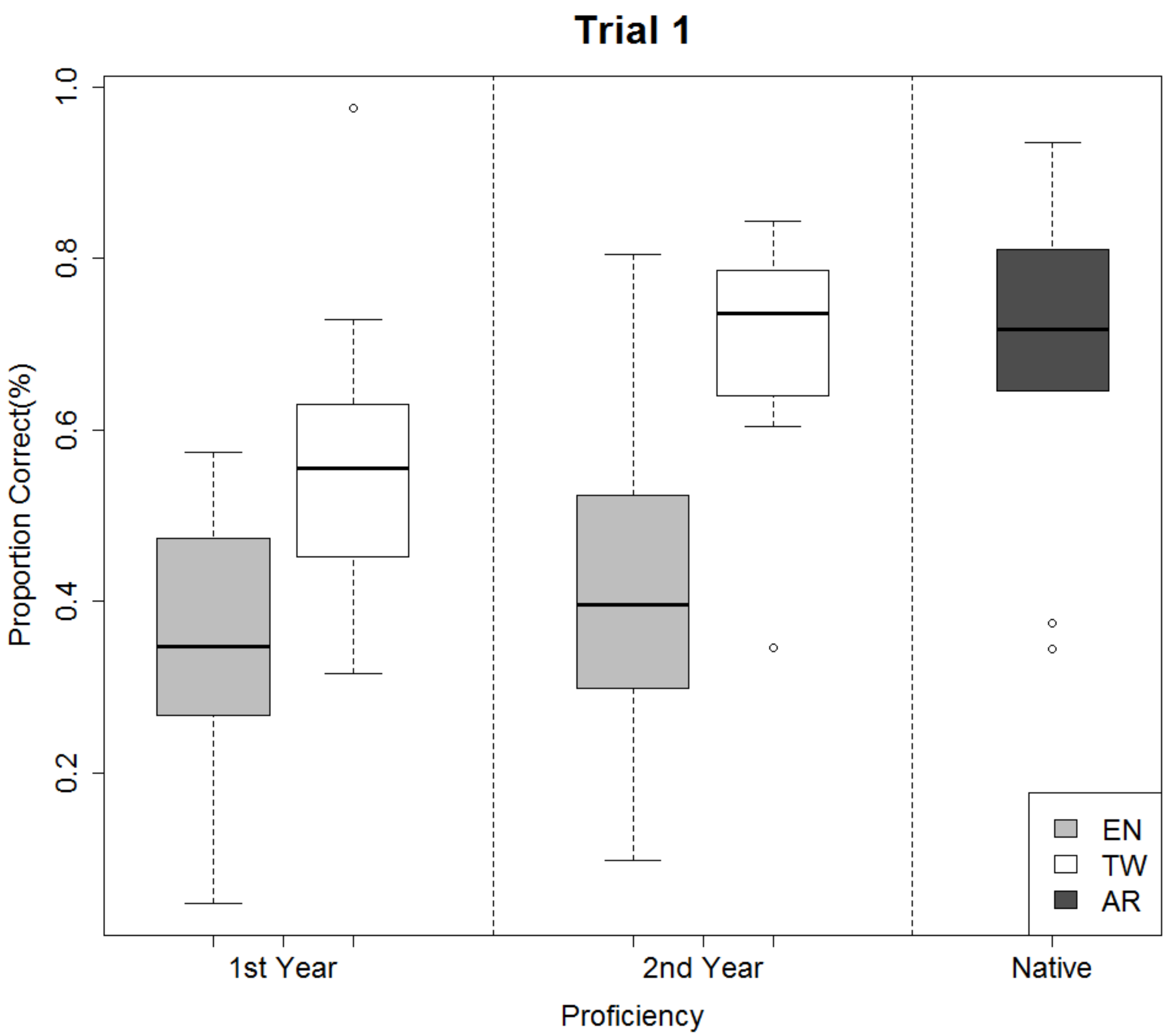


In sum, the analyses conducted in this section showed that participants in the control group had significantly better performance in the stress production task, in that the time they took to respond to the stimuli was significantly faster and their production was significantly more accurate than the English-speaking participants. While this advantage for native speakers in production should not be surprising, the analyses found that the Chinese-speaking groups also had an advantage in the production task compared to the English-speaking groups. It was revealed that the Chinese-speaking participants were significantly faster, just like their native counterparts, in reaction time and were more accurat in production than their English-speaking group counterparts.

4.5.2 Proficiency and Performance

To further investigate the role of proficiency, Figure 4.16 provides the response accuracy according to the native language and proficiency level of the participants. As can be seen in the figure, although no clear difference in average response accuracy is found between the first-year and second-year English-speaking participants, an increase in accuracy is exhibited as the proficiency level increases in the Chinese-speaking groups.

Figure 4.16. Boxplot for the Correctness of Response Grouped by Participants' L1 and Proficiency for the Stress Production Task (Trial 1)



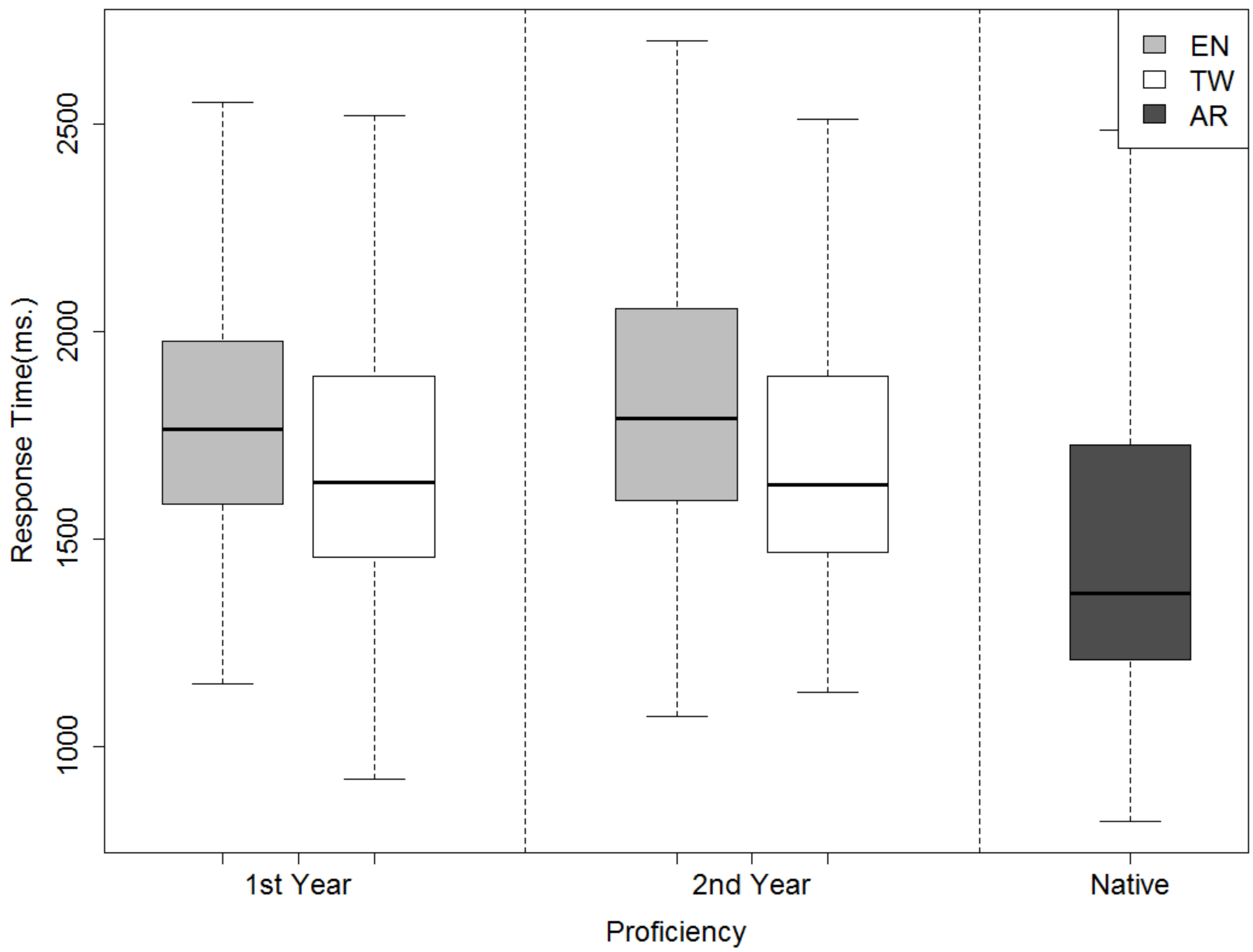
To examine the extent of such difference results from the proficiency of the non-native participants, a two-way ANOVA was conducted with native language as a between-group variable (two levels: English and Mandarin Chinese), proficiency (two levels: 1st year and 2nd year) as a within-group variable and the response accuracy as the dependent variable. The test found a significant main effect of proficiency ($F(1,63) = 4.97$, $p < 0.05$, $\eta^2 = 0.07$) and a near significant interaction between native language and proficiency ($F(2,63) = 2.88$, $p = 0.09$, $\eta^2 = 0.04$). The post-hoc Tukey tests showed that the Chinese-speaking participants were significantly more accurate in their production than the English-speaking group participants, regardless of their level proficiency. That is, both first- and second-year Chinese-speaking participants were significantly more accurate than both first- and second-year English-speaking participants. Lastly, the test showed that the second-year Chinese-speaking participants were significant more accurate in stress placement than their first-year counterpart, suggesting that higher proficiency does help the learners achieve better accuracy in stress production.

To compare response accuracy between participants of different proficiency levels and the control group, two additional one-way ANOVA tests were conducted, with native language as between-group variable and response accuracy as the dependent variable. The two tests showed that both the Chinese-speaking and native speakers were significantly more accurate than the first-year ($F(2,47) = 17.81$, $p < 0.001$, $\eta^2 = 0.43$) and second-year ($F(2,32) = 14.33$, $p < 0.001$, $\eta^2 = 0.47$) English-speaking participants, while no significant difference in response accuracy was found between the native group and both first-year and second-year Chinese-speaking groups.

Taken together, the effect of proficiency is evident for the Chinese-speaking groups while such an effect was not found for the English-speaking groups. Additionally, compared to other participants in the experiment, the English-speaking participants seemed to have difficulty placing stress on the correct syllable in their productions. On the other hand, all Chinese-speaking groups seem to have an advantage in terms of producing stress on the correct syllable, as their performance in the tasks was constantly found to be comparable to that of the native speakers, regardless of their proficiency.

Figure 4.16 illustrates the response time according to the native language and the proficiency of the participants. As shown in the figure, the native speakers again have a shorter average response time, seemingly followed by the Chinese-speaking participants then the English-speaking participants. To examine the extent to which participants' proficiency contribute the abovementioned differences in response time, a similar two-way ANOVA was conducted, with response time as the dependent variable, native language as a between-group variable (two levels: English and Mandarin Chinese), proficiency (two levels: 1st year and 2nd year) as a within-group variable. However, the test showed that there was not a significant effect of proficiency nor was there significant interaction between native language and proficiency, which suggest that the non-native participants took similar time to begin producingof the stimuli.

Figure 4.17. Boxplot for the Reaction Time Grouped by Participants' L1 and Proficiency for the Stress Production Task (Trial 1)



In short, the effect of proficiency was only identified in the response accuracy of the Chinese-speaking participants, as the second-year participants were significantly more accurate in their production than their first-year counterparts, while such effect was not detected for the English-speaking participants. Proficiency was not found to be a significant indicator of response time.

To summarize this section, the Chinese-speaking group exhibited some advantages that are comparable to that of the native speakers in stress production over their English-speaking counterparts. Both groups of Chinese-speaking participants begin the production of stimuli more quickly, and were significantly more accurate in the placement of stress than their English-speaking counterparts. The effect of proficiency was also relevant for the Chinese-speaking participants, as their response accuracy in the task was found to increase along with the level of their proficiency.

4.5.3 Stress Pattern and Performance

To delve more into the effect of native language and proficiency, more in-depth analyses were conducted with subgroups of the stimuli. These subgroups were based on the type of stress patterns of the stimuli. A similar set of ANOVAs as discussed in Section 4.3.3 was conducted for the participants' responses to the eight types of stress patterns. The full summary tables of the analyses are presented in Appendix 4.5.1 for response accuracy and in Appendix 4.5.2 for response time.

The analyses showed that there was a significant effect of native language on both response accuracy and reaction time for nearly all stress patterns whereas neither proficiency and the interaction between native language and proficiency were significant indicators of performance across all stress patterns. In terms of response accuracy, the Tukey tests generally found that the English-speaking participants were significantly less accurate than the native speakers on most stress patterns (i.e. stress patterns 3-8); the same holds for the Chinese-speaking participants on most stress patterns (i.e. stress patterns 1 and 3-8). The comparisons also showed that the Chinese-speaking participants' response accuracy was similar to that of as the native speakers for many stress patterns (i.e., stress patterns 1-5 and 8). Similar outcomes were found with respect to response time, as the native speakers took significantly less time to begin their production than the English-speaking participants on all stress patterns; the Chinese-speaking participants also began their productions more quickly on some of the stress patterns (i.e., stress patterns 4,5 and 8) than their English-speaking counterparts. Moreover, the Chinese-speaking group had similar responses time as the native speakers on some stress patterns (i.e., stress patterns 2,4 and 8)

In sum, the multiple comparisons confirmed the findings regarding the effect of native language and proficiency, but provided more detailed examination of the data. In addition to the generally better performance of the control group than the non-native groups, particularly the English-speaking groups, the multiple comparisons revealed that the Chinese-speaking groups had an advantage on the stress production task over the English-speaking groups in that they had more comparable response time to that of the native group. The Chinese-speaking groups were found to have significantly slower reaction time than the native group for only two stress pattern groups, i.e, stress patterns 6 and 7, while the English-speaking groups were found to have significantly slower reaction time for all stress patterns. Moreover, the production of stress by the Chinese-speaking groups was found to be comparable that of the native speakers, as no significant distinction in accuracy was found between the two groups for most stress pattern groups.

4.5.4 Relative Frequency and Performance

This section examines whether relative frequency of the stress pattern could be a predictor for both response accuracy and response time for each stimulus group. To examine the interaction between relative frequency and response accuracy, a repeated-measures logistic regression was conducted with relative frequency (two levels: frequent and infrequent) as a within-group variable, participant as the random variable to control for individual variability, and the correctness of response (two levels: correct and incorrect) as the dependent variable. The analysis structure was applied for all stimulus groups across all groups of participants in both trials. The full summary table for the analyses is presented in Appendix 4.6.1, and Table 4.10 presents a portion of the full summary table that is indicative of the effect of relative frequency.

Table 4.10 Summary of Frequency Effect

on Response Accuracy for Stress Production Task (Partial)

		Stimulus Group 1	
		Mean	SD
AR	Frequent	0.69	0.47
	Infrequent	0.64	0.48
EN 1st	Frequent	0.56	0.5
	Infrequent	0.3	0.46
EN 2nd	Frequent	0.58	0.5
	Infrequent	0.27	0.45
TW 1st	Frequent	0.63	0.49
	Infrequent	0.44	0.5
TW 2nd	Frequent	0.88	0.33
	Infrequent	0.64	0.48

	Group 1		
	Estimate	P Value	Odds Ratio
AR	0.31	$p = 0.48$	1.36
EN 1st	0.63	$p < 0.05$*	1.87
EN 2nd	0.3	$p = 0.39$	1.34
TW 1st	0.85	$p < 0.01$**	2.34
TW 2nd	1.44	$p < 0.001$***	4.2

A general trend of positive estimate was demonstrated for trisyllabic stimuli, especially for stimulus groups 1. Within this stimulus group, the main effect of relative frequency was significant for several groups of participants. In particular, for the responses in stimulus group 1, nearly all groups of participants, except for the native group and first-year English-speaking group, produced significantly more accurate stress placement for stimuli with frequent stress patterns than stimuli with infrequent stress patterns. This preference for frequent stress pattern was indicated by the positive estimate value.

For the stimuli in groups 2 to 5, the trend was rather mixed, which was exhibited by contrasting estimate values and scattered results. Even within the same stimulus group, some groups of participants had positive estimate values while some others had negative estimate values. In addition, although the main effect of relative frequency was significant for some groups of participants, these significant differences in performance were found for different stimulus groups. These scattered differences resulted in a less clear pattern regarding the effect of relative frequency.

Following the response time analyses conducted in the previous section, a one-way ANOVA with repeated measures was conducted, with relative frequency as a within-group variable (two levels: frequent or infrequent), participants as a random variable to control for individual variability, and reaction time as the dependent variable. The same ANOVA structure was used across all stimulus group for both trials. For trial 1 responses, the full summary table is presented in Appendix 4.6.2. Table 4.11 was extracted from the full summary table and presents the results which were more consistent and indicative of the main effect of relative frequency.

Table 4.11 Summary of Frequency Effect

on Reaction Time for Stress Production Task for Trial 1 (Partial)

		Stimulus Group 1		Stimulus Group 2		Stimulus Group 3	
		Mean	SD	Mean	SD	Mean	SD
AR	Frequent	1395	369	1490	443	1476	506
	Infrequent	1583	528	1625	532	1425	423
EN 1st	Frequent	1762	320	1806	329	1821	338
	Infrequent	1733	288	1838	310	1812	337
EN 2nd	Frequent	1741	343	1773	326	1883	375
	Infrequent	1902	373	1837	330	1870	371
TW 1st	Frequent	1676	342	1706	397	1666	386
	Infrequent	1668	406	1778	402	1653	316
TW 2nd	Frequent	1630	314	1732	347	1613	268
	Infrequent	1653	325	1732	324	1634	288

	Trisyllabic								
	Group 1			Group 2			Group 3		
	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value
AR	-188	F(1,134) = 8.98	p < 0.01**	-223	F(1,134) = 13.72	p < 0.001***	35	F(1,134) = 0.23	p = 0.63
EN 1st	29	F(1,297) = 0.94	p = 0.33	-102	F(1,297) = 8.71	p < 0.01**	131	F(1,297) = 17.04	p < 0.001***
EN 2nd	-162	F(1,194) = 16.21	p < 0.001***	-188	F(1,193) = 20.73	p < 0.001***	25	F(1,193) = 0.33	p = 0.57
TW 1st	11	F(1,312) = 0.1	p = 0.76	-128	F(1,311) = 12.25	p < 0.001***	139	F(1,313) = 11.1	p < 0.001***
TW 2nd	-24	F(1,194) = 0.39	p = 0.53	-203	F(1,194) = 29.52	p < 0.001***	180	F(1,194) = 19.52	p < 0.001***

As shown in Table 4.11, a consistent pattern seemed to emerge from participants' responses to stimuli in stimulus groups 1 and group 2. A trend of negative estimate value was prevalent in the two groups with significant main effects of relative frequency found for several groups of participants. The negative estimate value was indicative of shorter reaction time for stimuli with frequent stress patterns than stimuli with infrequent stress patterns. The trend was stronger for the responses to stimulus group 2 than stimulus group 1, as all groups had significantly shorter reaction times for stimuli with frequent stress patterns than stimuli with infrequent stress patterns. The responses in stimulus group 1 and 4 (not shown in Table 4.11) generally followed this preference for stimuli with frequent stress patterns, even though the trend was not as strong as in stimulus group 2. For group 1 and 4 stimuli, some groups of participants (the native group, the second-year English-speaking groups for stimulus group 1, and first-year English-speaking and Chinese-speaking groups for stimulus group 4) had significantly shorter reaction times for stimuli with frequent stress patterns than stimuli with infrequent stress patterns, while the remaining groups did not seem to have a significant preference towards stimuli with a specific frequency pattern.

However, the responses for stimulus group 3 had a reverse pattern. Although the trend was less strong than was found for stimulus group 2, all groups of participants had a positive estimate value and three groups of participants had significantly longer reaction times for stimuli with frequent than for stimuli with infrequent stress patterns.

In summary, the analyses found a few trends that might be indicative of a frequency effect on participants' performance. The onset of production times had a somewhat mixed trend as the participants responded significantly more quickly to stimuli with a frequent stress pattern for stimulus groups 1 and 2, whereas they responded significantly more slowly to stimuli with a frequent stress pattern for stimulus group 3. On the other hand, the analyses for response accuracy had a more consistent trend. Nearly all groups of participants produced stress more accurately for stimuli with a frequent stress pattern than stimuli for stimulus group 1. Lastly, the analyses showed that the trends found for the two trials were similar.

Consistent with the findings for the other tasks of the experiment, frequency of the stress pattern did not globally influence the performance of the participants, as significant differences in performance were not found across all subgroups of stimuli. Instead, significant differences in performance seemed to converge for specific stimulus groups, such as group 1. The trends displayed by these stimulus groups also support what was predicted for responses to stimuli with a frequent stress pattern, in that the productions were found to take less time to initiate and were more accurate for stimuli with a frequent stress pattern than stimuli with an infrequent stress pattern.

4.5.5 Summary of Results

Table 4.12 presents the trend found in all analyses conducted for the responses in the stress production task and their interaction with the variables identified in the study. In comparing the performance of participants from different native language groups, a global effect of native language was found, as the trend extended to nearly all the analyses and comparisons conducted. The results showed that the participants in the native group consistently outperformed the non-native groups. Specifically, they responded to the stimuli significantly faster than the English-speaking and Chinese-speaking groups (for some specific stress pattern) and had significantly more accurate placement of stress in their productions than the English-speaking groups. The results are not surprising as they are the native speakers of the language. The results also suggested that the scoring scheme used to determine the response accuracy of stress production is, albeit perhaps not perfect, adequate, as it gave the highest correctness rate to the native group at approximately 90 %, which adequately evaluated the production of native speakers as their production is supposed to be close to ceiling most of the time, if not all the time.

Table 4.12 Summary Table for Findings in Stress Production Task

	REACTION TIME	CORRECTNESS OF RESPONSE
Native Language	1. Shorter reaction time for native group than non-native groups (mostly the English-speaking group) 2. Shorter response time for the Chinese-speaking groups than the English-speaking groups	1. Higher accuracy for the native group and the Chinese-speaking groups than the English-speaking groups
Proficiency	No effect	1. Higher accuracy for the second-year Chinese-speaking group than the first-year group
Frequency	1. Shorter reaction time for frequent stress pattern in stimulus group 1 and 2 2. Trend for longer reaction time for frequent stress pattern in stimulus group 3	1. Higher accuracy for frequent trisyllabic stress pattern in stimulus group 1
Trial	No effect	No effect

The more interesting finding emerged from comparing the performance between the English-speaking and Chinese-speaking groups. The analyses demonstrated that the Chinese-speaking groups had better performance than their English-speaking group counterparts. Namely, the Chinese-speaking participants begin their production of the stimuli significantly more quickly and their placement of stress was more accurate than that of the English-speaking group. Furthermore, no significant difference in the correctness of stress was found even between the Chinese-speaking group and the native group. This result suggests that the location in which the Chinese-speaking group placed stress was similar to that of the native speakers. In addition, the multiple comparisons showed that, when the Chinese-speaking groups responded to trisyllabic stimuli, their reaction times were comparable to those of the native speakers, as no significant difference in reaction time was found between the Arabic-speaking group and the Chinese-speaking groups, as opposed to the English-speaking groups who had significant longer reaction time than the native group for nearly all stress pattern groups.

As for the interaction between proficiency and performance, the analyses did not find the proficiency of the participants to be a main predictor of reaction time and response accuracy. A significant effect of proficiency was only shown for the Chinese-speaking participants, as their response accuracy significantly improved for the second-year participants compared to their first-year counterparts. However, more often than not, the main effect of proficiency was not significant for either reaction time or correctness of response analyses and it was not significant in the multiple comparisons either. Together, the results suggested that within a language group, the non-native participants responded to the stimuli comparably fast and were comparably correct, regardless of their proficiency. In other words, higher proficiency had a limited effect on stress production; higher proficiency did not lead the participants to process the script they saw and then produce it more quickly, and it did not help the English-speaking groups to place stress on the correct syllable in their production.

Unlike the global effect that native language had on performance, the frequency effect seemed to exist more locally, as the effect was significant for some of the stimulus groups but not all of them. In the production task, trends that aligned well with the predictions for stimuli with frequent stress pattern were found. For specific stimulus groups (i.e. stimulus group 1 and 2 for reaction time and stimuli 1 and 3 for correctness of response), the participants responded more quickly and produced stress more accurately for the stimuli with frequent stress patterns than stimuli with infrequent stress patterns. At times, an unexpected trend was found for reaction time analyses, as the participants had shorter reaction times and more accurate responses when responding to stimuli with infrequent stress patterns than stimuli with frequent stress patterns for some of the stimulus groups.

4.6 Sub-Group Comparisons – Input Variance

There remain a few variables that deserve further scrutiny. The potential disparity in the amount of input received by English-speaking groups is discussed here. As mentioned in Section 4.1 of this chapter, there was a group of graduate students who also enrolled in the Arabic program at the English-speaking data collection site. They were required to enroll in a special section of the Arabic class giving two extra hours of contact time than the students enrolled in the regular sections of Arabic class. These extra hours were dedicated to the conversation hour where extra input might be available to the students, resulting in possibly significantly different performance from students enrolled in regular sections of Arabic class.

Six participants enrolled in the special section in the second-year group. To control for this variable, I will compare the analyses done including these six participants (i.e. the already reported analyses) and excluding them to examine whether the inclusion of these participants influenced the results and analyses that were established in previous sections.

For the responses obtained from the stress identification task, a two-way ANOVA with repeated measures was conducted with native language as a between-group variable, proficiency as a within-group variable and reaction time as the dependent variable. The analysis showed that native language had significant effect on reaction time whereas no significant difference was found in the proficiency of the participants. A post-hoc Tukey test further showed that the native group had significant longer reaction times than the non-native groups, which was identical to our previous findings. As for response accuracy, an identical two-way ANOVA was conducted with proportion correct as the dependent variable. The test found that native language had significant effect on the correctness rate, and the Tukey test further indicated that both non-native groups outperformed the native group with significant higher proportion correct in their responses, which was again identical to our previous finding.

For the responses from the lexical decision task, the same sequence of analyses follow. A two-way ANOVA with repeated measures was conducted with the same between-group variable, within-group variable, and reaction time as the dependent variable. The test indicated that neither native language nor proficiency had a significant influence on the reaction time of the responses, which perfectly matches our previous finding. For response accuracy, an identical two-way ANOVA test was conducted with proportion correct as the dependent variable. The test showed that native language had significant effect on the correctness rate, the Tukey test further showed the native group was significantly more accurate in their responses than the two non-native groups. The results found for the lexical decision task again completely aligned with our previous findings.

Lastly, for the results from the stress production task, a two-way ANOVA with repeated measures was performed, with identical between-group variable, within-group variable, and reaction time as the dependent variable. For trial 1, the ANOVA test found that native language had a significant effect on reaction time; in particular, it was found by the Tukey test that the native group had significantly shorter reaction times than both non-native groups. As for response accuracy, a similar two-way ANOVA with repeated measures was conducted with proportion correct as the dependent variable. The test found native language to have significant effect on the proportion correct in both trials; specifically, the Chinese-speaking group significantly outperformed the English-speaking groups with significantly higher response accuracy. The findings found in this section again perfectly align with previous ones.

Thus far, the inclusion or exclusion of the participants who enrolled in the special section for Arabic classes does not seem to make a significant difference, as nearly all results aligned perfectly with our previous findings. Next, I will examine whether the extra input that the participants obtained could result in any disparity in results for the frequency analysis.

The full results for the frequency analyses are given in Appendix 9. Although there were some disparities in performance for some stimulus groups, such instances were rare, as nearly all comparisons indicated no significant difference between the inclusion of the special group of participants (or the graduate students group for convenience) or the exclusion of them from the data used for statistical analyses. The difference in performance, i.e. the estimate value, between stimuli that had frequent stress pattern remained in the same direction between the two groups; that is, the estimate value for including the graduate students group is positive, and it is normally positive for excluding them. The significance level corresponds in most cases as well, which in turn supports the findings that were discovered in the previous sections. This invariance continued for both reaction time and response accuracy analyses, and was consistent across different tasks,

To conclude, the comparison between the inclusion of graduate students group or the exclusion of them did not indicate a significantly different pattern from the results that were previously identified. This result suggests that the extra two hours that the graduate students participants spent in conversation did not make their performance significantly better (or worse) than that of other participants who enrolled in the regular sections of Arabic class. Therefore, the analyses will continue to use the responses from all participants without excluding the graduate student participants, for the purpose of maintaining a larger sample size that would lead to sounder statistical analyses.

4.7 Sub-Group Comparisons – Type Of Nonsense Word

There were two types of nonsense words that were implemented in the stimuli of the experiment. Recall that half of the nonsense words were minimal pairs with a real word, which were constructed by replacing one consonant in the root of the word, one that is not attested to exist with the other two consonants in the root in the lexicon of Arabic. The other half of the stimuli were constructed by applying nonsense roots to the eight types of stress patterns used in the stimuli. The nonsense roots were identified by analyzing the frequency of co-occurrence of the root consonants from an Arabic lexicon. Consonants that do not often co-occur were chosen to form the nonsense root.

Recall that both groups of nonsense words were implemented to control for potential differences in the degree of naturalness. It is possible that the participants would perform differently with nonsense words derived from real words, since only one consonant was replaced from the root of the word and the derived word, albeit nonsense, might invokes certain degree of familiarity from the participants if they had encountered the real word before. Similarly, it is possible that the participants would react differently to nonsense words derived from nonsense root with templates, as they might be more uncertain about with nonsense words that had roots that they have never encountered before. Since both groups of nonsense words had their shortcoming and advantages, as discussed in Section 3.2.5, both groups of nonsense words were included in the experiment to control for the effect each group of nonsense words might introduce to the experiment

To examine whether a disparity in performance exists, a two-way ANOVA with repeated measures was conducted for each task in the experiment, with the participant group as a between-group variable (five levels: first-year, second-year Chinese-speaking groups, first-year, second-year, English-speaking groups and the native group) and the type of nonsense word as a within-group variable, participants as a random variable to control for individual variability and response time as the dependent variable. A repeated-measures logistic regression was conducted with the same set up as the two-way ANOVA with response accuracy as the dependent variable.

The results of the two-way ANOVA as well as the logistic regression did not reveal any significant differences in performance between two types of nonsense words. Rather, the participants seemed to respond to the two types of nonsense words similarly, without noticeable preference for either type of nonsense words. To further examine the effect of the type of nonsense words on the frequency analyses established in previous sections, a one-way ANOVA with repeated measures was conducted for each stimulus group across all groups of participants with type of nonsense words as a within-group variable and reaction time as the dependent variable. Additionally, a repeated-measures logistic was conducted with the same setting with correctness of response as the dependent variable. The results of the analyses are presented in Appendix 8. Although the analyses found that some groups had significant differences in performance for some stimulus groups, the differences were few and mostly scattered in participants from different native language groups as well as proficiency. No patterned difference seemed to emerge from the results.

4.8 External Factors from the Questionnaire

As stated in the participants Section 3.3, the participants completed a post-experiment questionnaire with questions that evaluate selected aspects of the participants' background that was not captured in experiment. The questionnaire aimed to assess four external factors that might be relevant to their performance in the task, which were: (1) their perception of the instruction that they have received from the instructor or the textbook for lexical stress in Arabic, (2) self-evaluation of their ability and performance on lexical stress in Arabic with questions such as “I always think about stress rules before I say anything”, (3) their explicit knowledge about the formal rule of stress in Arabic via a short quiz and (4) their extracurricular involvement with the target language; for example, questions asking whether they have friends who are native speakers of Arabic, and the extent to which they use the target language outside of class. The scores were presented as a ratio of the total points that they received from their answers and the maximal points attainable in the section, which gives the score a range between 0 to 1 (i.e., 0%-100%). The detailed break-down for the questions in each section, along with the scores that each group of participants received is given in Appendix 9.

The participants’ responses to questions were scored for each section. It has to be noted that the scores were only suggestive, and are not reflective of the extent to which participants were influenced by these external factors, as the number of questions in each section does not reach sufficient number to have more explaining power. What the questionnaire could offer is a rather rough assessment of the perception and the self-assessment related to these external factors.

In the remainder of this section, these scores are examined with a focus on the interaction between the scores that the participants obtained in each section and the native language as well as the proficiency of the participants. It has to be noted as well that the native group only took the quiz that assessed their explicit knowledge on stress in Arabic, as other questions did not apply to them.

4.8.1 Instruction and Textbook

The summary table for the responses to the "Instruction and Textbook" section of the questionnaire is presented in Appendix 9.1. The first question in this section of the questionnaire was "My Arabic teacher has introduced stress rules in Arabic". If the participants selected "Agree" or "Strongly Agree" for the question, they are prompted with an open-ended question "List the stress rules that your Arabic teacher has introduced" in which they could provide the details of the stress rules that were presented to them in class. Of the 76 participants who answered the question, 50 of them, i.e. approximately 62% of the participants, agreed or strongly agreed that their teacher had introduced stress rules in class. A more detailed breakdown of the participants who agreed or strongly agreed with the statement is presented in Table 4.13. The results seem to suggest an overall lower agreement for the first-year English-speaking group than the first-year Chinese-speaking group. In other words, fewer English-speaking participants perceived that that their Arabic teachers had introduced stress rules in the Arabic class than their Chinese-speaking counterparts did.

Table 4.13 Summary Table for Participants' Responses to Question 1

	Question: "My Arabic teacher has introduced stress rules in Arabic"			
	EN		TW	
	1st Year	2nd Year	1st Year	2nd Year
Agree	55%	69%	76%	69%
Disagree	45%	31%	24%	31%

The stress rules provided by these 50 participants were individually examined and labelled. Sample of stress rules provided by the participants are presented in Table 4.14. A total of 56 rules were provided by the participants, and the responses converged to 10 types of rules. The distribution of their responses is presented in Table 4.14. The most mentioned stress rules were “long vowel” and “gemination”, accounting for approximately 80% of their responses. These two rules meant to place stress on a syllable that has long vowel and to place stress on a syllable that has a geminated consonant, respectively.

Table 4.14 Perceived Stress Rule Introduced by The Arabic Teacher

Long Vowels	**56%**
Gemination	**23%**
The letter “alif”	4%
Second Long Vowel	4%
Nunation (Tanween)	2%
The Letter 'ain	2%
Second Syllable of Form 8 and 10 Verbs	2%
First Syllable	2%
Second Syllable	2%
The Letter Hamza	2%

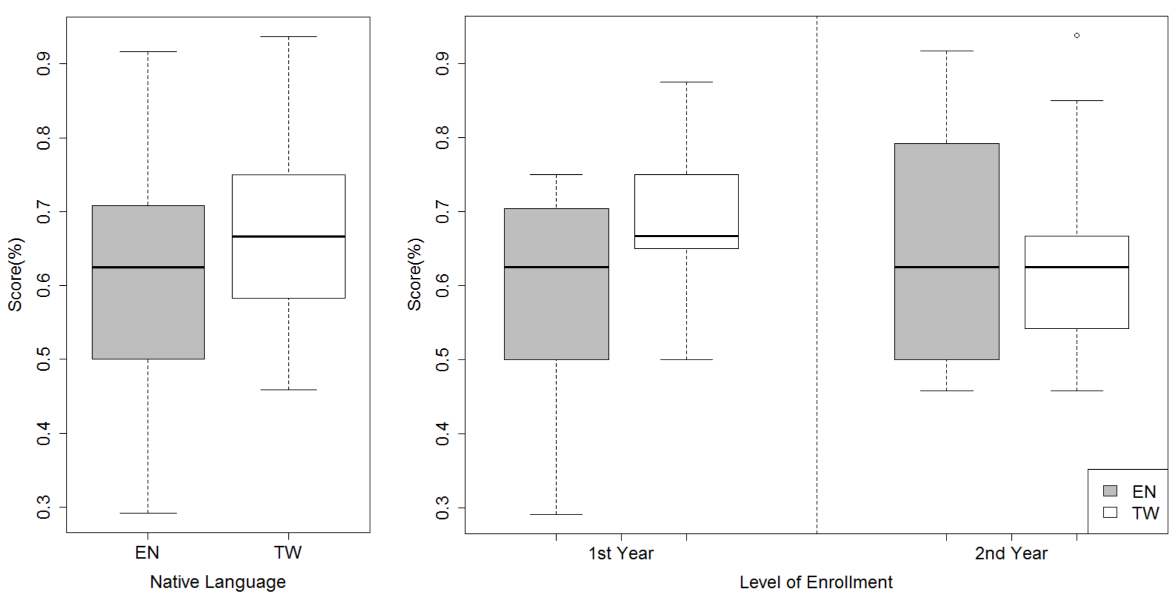
More interestingly, in a subgroup comparison, it was found that the two rules were not distributed evenly across participants of different proficiency. As presented in Table 4.15, both gemination and long vowels were mentioned by the English-speaking groups. However, the rule “gemination” was not mentioned at all by the first-year Chinese-speaking group, and was increasingly mentioned in higher level groups.

Table 4.15 Subgroup Comparison for the Most-Mentioned Stress Rules

	EN		**TW**	
	1st Year	2nd Year	1st Year	2nd Year
Gemination	29%	33%	0%	17%
Long Vowels	71%	17%	78%	67%

Figure 4.18 gives the Instruction and Textbook scores according to the participants' native language on the left, and by the proficiency on the right. A one-way ANOVA was conducted to examine the significance of the difference in section score between the participants of different native language. The ANOVA was conducted with native language as a between-group variable (three levels: Arabic, English and Mandarin Chinese) and section score as the dependent variable. The test found a nearly significant effect of native language on the Instruction and Textbook score ($F(1, 65) = 3.49$, $p < 0.06$, $\eta^2 = 0.05$). With follow-up Tukey tests, the result showed a nearly significant difference between the Chinese-speaking and the English-speaking groups where the former has higher score in this section of the questionnaire. In other words, the Chinese-speaking groups perceived that they received more instruction from the instructor and textbook regarding the stress rules in Arabic more than the English-speaking groups

Figure 4.18 Scores for Instruction and Textbook According to L1 and Proficiency



A two-way ANOVA was conducted to examine the interaction between the questionnaire scores and the proficiency of the participants. The two-way ANOVA was conducted with native language as a between-group variable (two levels: English and Mandarin Chinese), proficiency as a within-group variable (two levels: first-year and second-year) and Instruction and Textbook score as the dependent variable. However, there was not a significant effect of proficiency nor was there an interaction between native language and proficiency, suggesting that the English-speaking and the Chinese-speaking participants gave similar scores in the section regardless of their proficiency.

4.8.2 Self-Awareness of Stress

The questions in this section aimed to probe more into participants' awareness of stress when dealing with tasks that were related to stress in Arabic. Questions such as "I always think about stress rules before I say anything", or "I always pay attention to where I put my stress in speaking" were presented in this section of the questionnaire. The full summary table for their overall degree of agreement is presented in Appendix 9.2.

Figure 4.19 presents the overall Self-Awareness scores according to the native language of the participants on the left and the proficiency of the participants on the right. Although the average score seemed to be slightly higher for the Chinese-speaking groups, the one-way ANOVA with the Self-Awareness scores as dependent variable did not find a significant main effect of native language, meaning that the two non-native groups had similar scores.

Figure 4.19 Scores for Self-Evaluation Grouped by L1 and Proficiency

A two-way ANOVA was conducted to determine whether there was a difference in the scores between participants of different proficiency levels, with native language as a between-group variable (two levels: English and Mandarin Chinese), proficiency as a within-group variable (two levels: first-year and second-year) and the self-awareness score as the dependent variable. However, there was not a significant effect of proficiency nor an interaction between the two factors on the score.

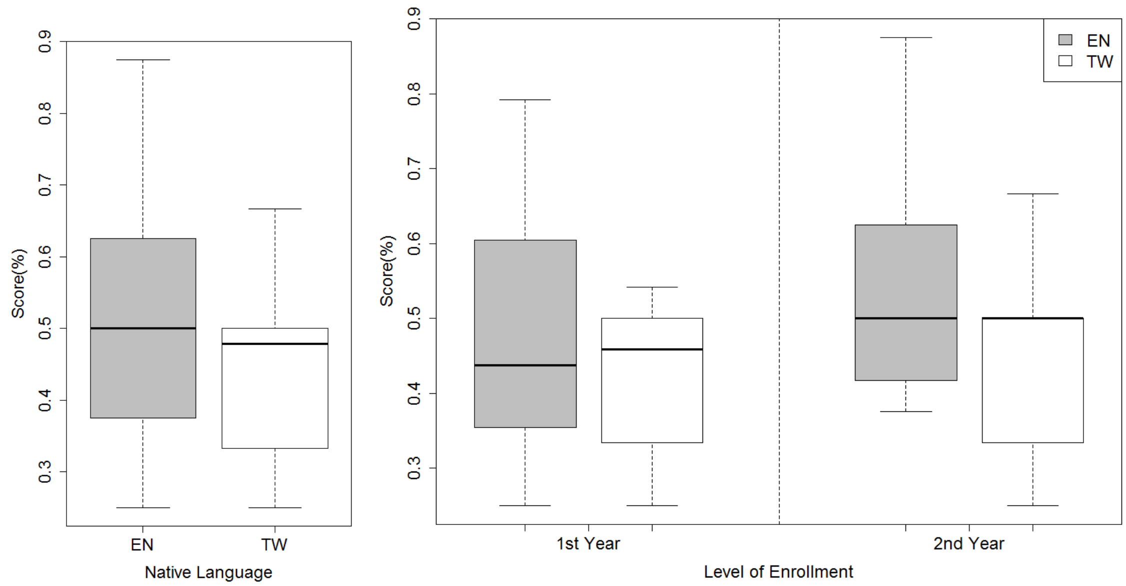
In sum, all participants, regardless of their native language and level of proficiency, were similarly aware of stress when performing stress-related tasks, and the degree of such awareness did not seem to increase or decrease significantly as they become more proficient in the language.

4.8.3 Extracurricular Engagement

The questions in this section of the questionnaire attempted to assess the participants' degree of contact with Arabic when they are not in the Arabic class. These questions inquired about their opportunity of exposure to the target language by interacting with native speakers with questions such as "I have a lot of opportunities to use Arabic outside of class "or "I have a lot of friends who are native speakers of Arabic". These questions were also intended to probe into the extent to which the participants used the target language outside of class and the specific language skills (e.g., speaking, listening, writing and reading) that they used, with questions such as "I find myself speaking Arabic all the time outside of class".

Figure 4.20 presents the overall score according to the participants' native language on the left and participants' proficiency on the right. Following previous analyses, a one-way ANOVA was conducted with native language (two levels: English and Mandarin Chinese) as a between-group variable and the Extracurricular Engagement score as dependent variable. The analysis found the main effect native language to be significant ($F(1, 65) = 4.66$, $p < 0.05$, $\eta^2 = 0.06$). The post-hoc Tukey tests further revealed that the English-speaking groups had significantly higher scores than the Chinese-speaking groups. The higher score suggested that the English-speaking participants were generally more engaged and had more opportunity to have contact with Arabic than the Chinese-speaking participants outside of class.

Figure 4.20 Scores for Extracurricular Engagement Grouped by L1 and Proficiency

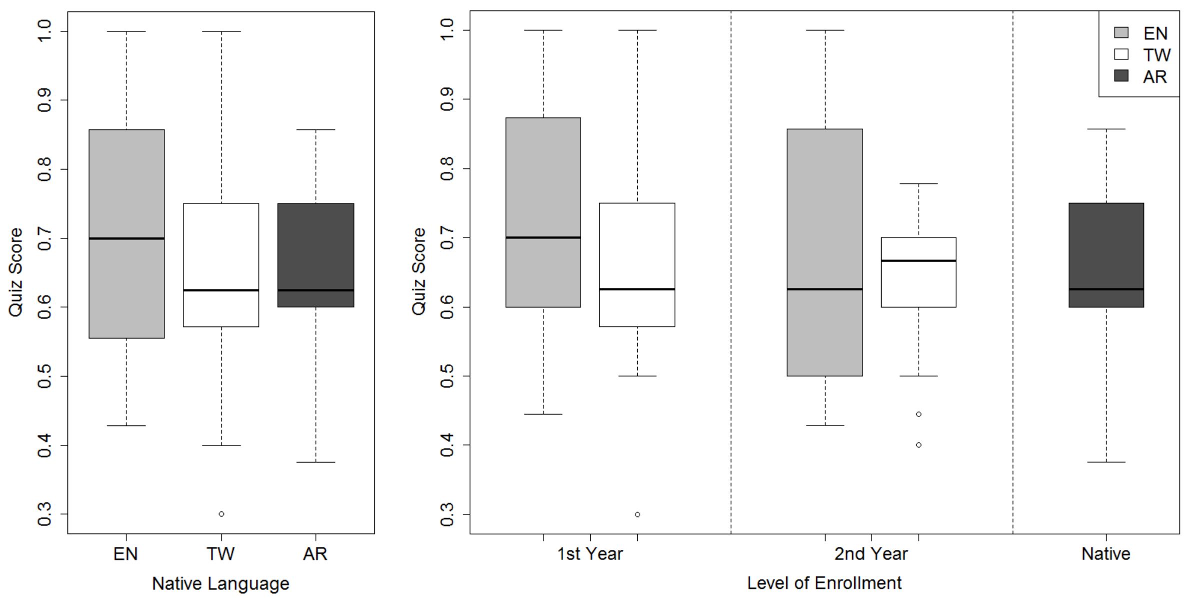


A follow-up two-way ANOVA was conducted to examine the main effect of proficiency on the Extracurricular Engagement score in this section, with native language as a between-group variable (two levels: English and Mandarin Chinese), proficiency as a within-group variable (two levels: first-year and second-year) and the overall extracurricular engagement score as the dependent variable. However, the analysis did not find a significant main effect of proficiency or an interaction between native language and proficiency, suggesting that the participants received similar scores regardless of their level of proficiency.

4.8.4 Explicit Knowledge

Finally, this last section of the questionnaire aimed to examine the extent to the participants explicitly know whether a statement about the stress rule in Arabic is true or false. This section of questionnaire took the format of a quiz and presented ten statements regarding the stress rules in Arabic. The participants needed to determine whether the statement was true or false by selecting “True” or “False” on the questionnaire. They received one point for each correct answer and zero for incorrect answer. There was a "I don't know" option as well. When the option "I don't know" is selected, the calculation will skip this question and use the other questions to calculate scores. Figure 4.21 presents the Explicit Knowledge score grouped by participants' native language on the left and by the proficiency of the participants on the right.

Figure 4.21 Explicit Knowledge Score According to L1 and Proficiency



Overall, the participants did not perform very well on the quiz. The average score for the quiz was approximately 0.7 for the English-speaking group and 0.6 for the Chinese-speaking groups, as well as for the participants who were native speakers of Arabic. In other words, the participants on average answered 6 questions correctly out of a total ten, which is slightly above chance. To determine whether the language groups differed significantly from each other, a one-way analysis of variance (ANOVA) was conducted with native language as a between-group variable (three levels: Arabic, English, Mandarin Chinese) and Explicit Knowledge score as the dependent variable. There was not a significant effect of native language on the quiz score, suggesting that the three groups received similar scores on the quiz.

A two-way ANOVA was followed to examine the effect of proficiency on the quiz score for the non-native groups. The ANOVA was conducted with native language as a between-group variable (two levels: English and Mandarin Chinese), proficiency as a within-group variable (two levels: first-year and second-year) and quiz score as the dependent variable. The analysis indicated that there was no significant effect of proficiency or an interaction between native language and proficiency. This result suggest that higher proficiency did not help the participants to answer the questions in the quiz more correctly, and all groups of participants received similar scores on the quiz, regardless of their native language and proficiency.

In sum, the results show that native language and proficiency did not inform their explicit knowledge about stress in Arabic. Besides, the generally low scores, especially for the Chinese-speaking participants, on the quiz suggested that although they had some applicable explicit knowledge about the stress rules in Arabic as presented in Appendix 10, a generally more sophisticated understanding of lexical stress in Arabic did not seem to be readily available to the participants.

4.8.5 External Factors and Experiment Results

In this section, I further examined whether these external factors interacted with the performance of the participants on the experiments. A correlation was conducted between average reaction time as well as the accuracy with the scores obtained from each section of the questionnaire. The correlation test was done for the participants in each language group, regardless of their proficiency since proficiency was not found to have a significant within-group effect on the scores as shown in the analyses conducted in the “External Factors” section. The full summary table for the correlation index (Pearson’s r) as well as the significance level is presented in Appendix 11.1 for the native group, 11.2 for the English-speaking groups, and 11.3 for the Chinese-speaking groups. Note again that the native speaker participants only completed the quiz section of the questionnaire as the other sections of the questionnaire that the non-native groups completed did not apply to native speakers of Arabic.

Overall, although the four external factors mostly did not significantly correlate with the responses in the experiment, the analyses showed some trends in the results that were worth mentioning. Firstly, the analyses found that participants’ explicit knowledge of stress which was evaluated by the quiz in the questionnaire was significantly and positively correlated with the participants’ accuracy in the stress identification task of the experiment. Specifically, this positive correlation was found for the native group and the English-speaking groups, but not the Chinese-speaking group. In other words, the more the participants (especially the native and English-speaking participants) were aware of the formal rules of stress in Arabic, the more they were able to identify the stressed syllable correctly.

Additionally, the analyses found a significant negative correlation between the scores obtained in the extracurricular engagement section of questionnaire and the response accuracy in the lexical decision task for the English-speaking groups. A negative correlation suggests that the more socially engaged the English-speaking participants were, the less correct they were in the lexical decision task. Recall that stimuli examined here were all nonsense words; therefore, to correctly respond to the stimuli in the lexical decision task means to reject nonsense word. In other words, the results showed that the more socially engaged the participants were with the language outside of class, the less willing there were to reject stimuli that were nonsense words. However, this tendency was not found for the Chinese-speaking groups.

Finally, response accuracy in the stress production task was found to be significantly and positively correlated with the scores obtained by the Chinese-speaking groups in the "Self-Awareness" section of the questionnaire. A positive correlation suggests that the higher the scores were, the accurate the participants were in the stress production task. Namely, the result suggests that when the Chinese-speaking participants were more aware of stress in production, the more successful they were in placing stress on the correct syllable in their stress production.

In addition, the scores obtained in the "Self-Awareness" section of the questionnaire was negatively correlated with the reaction time of the stress production task in the Chinese-speaking groups as well. A negative correlation suggests that the higher scores in the questionnaire, the shorter the reaction time. In other words, the results suggest that the Chinese-speaking participants have more processing advantage and a reduced reaction time in stress production when they perceived themselves to have better awareness and active notice of stress (assessed in the Self-Awareness section) when dealing with tasks that had to do with stress in Arabic. Nonetheless, the trends observed for the Chinese-speaking groups did not seem to apply to the English-speaking groups.

These trends have intriguing implications to reflect on participants' performance in the tasks. However, the trends obtained from the questionnaire shall be interpreted as suggestive at best, as the questionnaire was rather selectively shorter in length and required further validation (such as establishing face validity, piloting the questionnaire, and follow-up statistical tests) for it to be a sound questionnaire.

4.9 Summary of the Results

This section provides an overall summary for the experiment results and how they interacted with the main factors identified in the beginning of this chapter. These factors were (1) the native language of the participants, (2) the proficiency of the non-native speakers (3) relative frequency, and (4) external factors assessed in the questionnaire. The following sections presents the effect of these factor on the performance of the participants in the experiment.

4.9.1 Native Language

The main findings from the native language of the participants can be divided into two components, one of which is the performance of native speakers, and other is the comparison between the English-speaking and Chinese-speaking participants, who were non-native speakers of the target language.

As the control group, the native speakers were expected to excel in all tasks in the experiment, in that they were expected to have shorter reaction times and more accurate responses on all tasks of the experiment. However, the results found the native speakers did not meet this expectation in some parts of the experiment. The advantage of being a native speaker seemed to be in language production, and accuracy in tasks in which they could utilize their implicit knowledge of the language. The native speakers were excellent in determining whether a word was a real Arabic word or not. They were very successful in the stress production task, with significantly higher accuracy and shorter reaction times than some or all non-native groups. Nevertheless, their performance dropped when dealing with tasks that required more explicit knowledge that was not explicitly taught or available during the course of their education.

Precisely, in the stress identification task, they had significantly worse performance than non-native speakers, as they were significantly slower in their responses and were not able to accurately identify the stressed syllable in the stimuli as correctly as the non-native speakers.

The comparison between the English-speaking and Chinese-speaking groups did not yield many significant findings, as the participants in the two language groups performed rather similarly on two of three tasks. A significant difference between the two non-native groups was only found in the stress production task for nonsense words. In this task, the participants in the Chinese-speaking group significantly outperformed their English-speaking counterparts. The Chinese-speaking participants began to produce the stimuli more quickly and were more accurate in stress placement in their production than the English-speaking speakers. Additionally, the degree of accuracy of the production by the Chinese-speaking speakers was comparable to that of native speakers, as no significant difference in response accuracy was found between the two groups. Furthermore, the analyses also showed that the reaction times of the Chinese-speaking participants in the stress production task were comparable to these of the native speakers. Specifically, in the multiple comparisons for the participants' responses to stimuli with respective stress pattern, no significant difference in reaction time was found between the Chinese-speaking groups and the native group for trisyllabic stimuli.

4.9.2 Proficiency

Although the analyses have repeatedly examined the effect of proficiency, as measured by years of instruction, on the non-native participants' performance in the task, no significant effect of proficiency, except for few instances from the Chinese-speaking groups that might be suggestive of the advantage of higher proficiency in the performance. Precisely, the second-year

Chinese-speaking group responded more accurately to the stimuli in the stress production task than their first-year counterparts. Nevertheless, the results did not offer overly evidence to support the role of proficiency in improving the performance of the participants, at least not in the perception and production of stress.

4.9.3 Relative Frequency

The experimental results showed that relative frequency had a more local effect on participants' performance. Unlike the effect of native language, the effect of relative frequency was found in specific stimuli groups only where participants were more fluent and accuracy for stimuli with frequent stress patterns. This preference for frequent stress pattern was strong for trisyllabic stimulus groups, stimulus group 1 in particular. The reaction time was found to be significantly shorter for stimuli with frequent stress pattern in group1 in the stress identification task and in the stress production task. Additionally, accuracy of stress production was significantly higher for stimuli with frequent stress pattern. Although opposite trends in which stimuli with frequent stress pattern were disfavored were found, they were comparatively fewer than trends favoring frequent stress patterns. On a related note, the strong trend emerged in the lexical decision task favoring stimuli with infrequent stress pattern deserves a separate discussion in the upcoming chapter, Chapter 5, as the tasks seem to require other type of knowledge instead of the knowledge of stress.

4.9.4 External Factors

The questionnaire suggested several other interesting observations. The results show that the majority, i.e. approximately 60%, of the participants, expressed that their Arabic teacher has introduced stress rules in class, and the rules that were introduced seemed to heavily concentrate on either the length of the vowel or the consonant, without addressing other rules of stress in Arabic. As a result, all groups of participants had a rather low score on the quiz for stress rules in Arabic, suggesting a rather limited explicit knowledge about the stress rules in Arabic, as the correctness rate of their responses was only slightly above chance. Lastly, the questionnaire also found that the English-speaking groups were significantly more engaged with the target language and had more opportunity using it outside of class than the Chinese-speaking groups, even though they were less successful in placing stress on the correct syllable in their production. This result potentially highlight the role of native language in stress production as even with more opportunity to use the target language, the English-speaking groups could not reach the same level of accuracy in their stress production than the Chinese-speaking groups.

Several patterns emerged from examining the correlation between questionnaire scores and experiment results. The analyses seemed to imply a possible relation between the degree of awareness of the formal rules of stress in Arabic and the performance on the stress identification by the native group and the English-speaking groups. In the same vein, the results seem to suggest a possible relation between the degree of self-awareness of stress and the performance in the stress production task, as higher self-awareness and notice of stress seemed to correlate with higher accuracy and better fluency in stress production. However, the analyses also suggested that the English-speaking participants were less successful on the stress production task when they had better extracurricular engagement with the target language, which is somehow inconsistent with the advantages generally assumed to accrue with social engagement.

Overall, this study revealed several interesting patterns emerging from the experimental results and the questionnaire results. The implication and the explanation of these results are discussed in more detail in the next chapter, along with inconsistent trends that were found in the analyses. In particular, the poor performance of the native speakers on the stress identification task as well the prevalent preference for infrequent stress patterns in the lexical decision task will be further discussed in the next chapter.

Chapter 5

Discussion and Conclusion

The findings of the present study revealed a number of patterns regarding (1) the general trend of stress-related cues in Arabic and (2) L2 Arabic stress perception and production by the participants of the study. With regards to the general trends of stress-related cues in Arabic, the analysis of the frequency dictionary reveals that:

1. There is a negative correlation between the number of syllables and the frequency of a specific stress pattern.
2. The frequency within different types of stress-related cues (i.e., stress pattern, stress position, syllabic structure and conditional probability of the stressed syllable) is unevenly distributed.

These findings seem to reflect a cross-linguistic trend for the frequency distribution of stress-related cues. In addition to the higher frequency count for stress patterns with fewer number of syllables, the uneven frequency distribution for the stress-related cues is also attested in other languages. For example, English bisyllabic words tend to be more frequent than their trisyllabic and quadrisyllabic counterparts (Clopper, 2002). For words with a specific number of syllables, stress falls more frequently on a specific position than others (Murphy & Kandil, 2004); for instance, it is found that stress falls on the penultimate position more frequently than on either the antepenultimate or the final position in English trisyllabic words. This varying preference for the stress-related cues is not exclusive to English but also attested in other languages such as Brazilian Portuguese (Cristófaro-Silva & Fraga, 2005) and Russian (Jouravlev & Lupker, 2014).

The findings from the analysis of the frequency dictionary confirm the variability of stress-related cues in frequency, which in turn reaffirms the necessity of considering the frequency of stress-related cues as independent variables that should be either controlled for or manipulated when designing stress perception and production experiments – a practice that is rarely adopted in previous L2 stress perception and production studies. To this end, the present study incorporated the frequency of stress-related cues into the design of the stimuli and conducted a stress perception and production experiments, which reveal the following findings:

1. In the speed reading task, the non-native participants from the second-year group showed significantly better fluency than their first-year counterparts, while the controls outperformed all non-native groups.
2. In the stress identification task, the Chinese L1 and English L1 participants had similar accuracy (i.e. correct responses) and fluency (i.e. reaction time).
3. In the stress identification task, the Arabic controls showed significantly slower and less accurate perception of stress in nonsense words than the Arabic L2 learners.

4. In the lexical decision task, the Arabic controls had significantly higher accuracy in rejecting nonsense words than the Arabic L2 learners.
5. In the lexical decision task, the Arabic controls had similar fluency (reaction times) as the Arabic L2 learners.
6. In the stress production task, the Arabic native speakers had more accurate production of stress than the Arabic L2 learners.
7. In the stress production task, the Chinese L1 participants had more accurate and more fluent (rapid) production of stress than their English L1 counterparts.
8. In all tasks examined, proficiency of the L2 learners did not seem to significantly contribute to their accuracy and fluency in the experiment.
9. The frequency effect is only partially evident in participants' responses to specific set of stimuli within certain tasks.

5.1 Discussion

The above findings can be grouped into three clusters to do with: L1 transfer, proficiency, and frequency, which are discussed in the following three sections, respectively.

5.1.1 L1 Transfer

The first cluster of findings is related to the role of L1 transfer, or the extent to which learners' native language could influence (either positively or negatively) the participants' perception and production of stress. Precisely, the findings address the first research question stated in Chapter 1 in Section 1.3.

RQ 1. Does the native language of the learners influence their performance in stress perception and production, and, if so, how?

From the results of the experiment, the native language of the learners seems to be the most relevant indicator of their perception and production of stress in Arabic, as the effect from the native language of the learners is exhibited in more than one task of the experiment. In some tasks, the performance of the learners exhibited different patterns from that of the native speakers.

In the stress identification task, the experimental results indicate that the English-speaking and the Chinese-speaking participants were fairly capable of identifying the stressed syllable of the aural stimuli, as the accuracy of their responses is approximately 70 % for the English-speaking participants and approximately 65% for Chinese-speaking participants, which are both higher than chance. Additionally, the results also indicate that, although the English-speaking participants have a higher average of accuracy than their Chinese-speaking counterparts, this difference does not reach statistical significance. These results suggest that their perception of Arabic stress in the task are fairly similar in terms of the accuracy and fluency.

This finding is consistent with findings from previous studies on L2 stress perception. Studies on L2 English stress perception have indicated that L1 Chinese speakers are capable of perceiving stress and in some cases they perceive stress as well as their L1 English counterparts (e.g., Wang 2008, Yu and Andruski 2011). Similarly, L1 English speakers are also found to be capable of perceiving stress. In a previous study investigating whether L1 English learners of Spanish use syllable weight as a cue for stress perception in Spanish, the advanced Spanish L2 learners were able to identify the stressed syllable in the stimuli as well as their L1 Spanish counterparts. They were also able to utilize syllable weight to perceive stress (Face, 2005).

Although few studies have directly compared the perception of stress by L1 English and L1 Chinese learners of another foreign language, it does not seem to be far-fetched to suspect that the higher average in correct responses found in the L1 English participants in the present study might be due to the presence of lexical stress in their native language as opposed to the L1 Chinese participants, whose native language is generally considered to lack such a prosodic feature.

Another rather interesting finding from the stress identification task is the poorer performance of the L1 Arabic controls on the task compared to the non-native participants. The experimental results show that the participants in the control group took significantly longer time to respond to both real and non-word stimuli and were significantly less successful in identifying the stressed syllable in the stimuli than both the L1 English and L1 Chinese participants regardless of their proficiency level. It is worth noting that the accuracy average of their response is lower than chance with approximately 40%, which suggests that the Arabic L1 controls were, to an extent, unable to identify stress.

This finding might appear to run counter to the general expectation for a control group, as the performance of participants in any native-speaking group is normally expected to provide the upper bound for the range of performance in an experiment of this type. Apparently, the control group in the present study failed to achieve this expectation in this task; nevertheless, this finding should not be taken as a counter evidence to L1 transfer effect. On the contrary, the difficulty that native speakers of Arabic encountered in stress perception is taken as an evidence for effect of the participants' L1. This proposition aligns well with the findings from previous stress perception studies, especially ones that focus on the typology of stress perception and stress deafness. In particular, Altmann (2006) examined the perception of English stress by speakers of different L1s (including Arabic). The researcher reported that learners with predictable stress in their native language (such as French, Turkish and, more importantly, Arabic) were found to have problems perceiving the location of stress, in that the accuracy of their perception is significantly lower than learners with other native languages, particularly lower than learners with unpredictable stress (such as Spanish) or without word-level stress (such as Japanese and Chinese). This finding matches my finding from the perception task as the participants with predictable stress in their native language (i.e., the Arabic L1 participants) are indeed found to perform respectively worse on both real and non-word stimuli than the participants with unpredictable stress (i.e., the English L1 participants) and ones without word-level stress (i.e., the Chinese L1 participants) in their native language.

Similar contrast between speakers with predictable stress and ones with unpredictable stress or no word-level stress was reported in a series of studies on stress deafness, i.e., the inability to detect the stressed syllable in a given word by Sharon Peperkamp and Emmanuel Dupoux (e.g., Dupoux *et al.* 1997, 2001, 2008, to name a few). Rather than focusing on a specific language, the results of examination are expected to apply cross-linguistically in terms of stress perception. Thus, the stimuli are composed of sequences of unmarked syllables and with stress acoustically manipulated to fall on different syllables, bearing no phonological feature that is specific to a particular language. As a result, Peperkamp and Dupoux (e.g., 2002) found that speakers of Finnish and Hungarian, languages which, like Arabic, have predictable but not phonetically contrastive stress, experienced stress deafness like the French speakers as reported in previous research. This finding supports a generalization found in Altmann (2006), as speakers of languages with predictable stress seem to experience difficulty in perceiving stress not only in the target language but in languages in general. Relating these results back to the present study, it may be the case that Arabic speakers encountered the same difficulty identifying stress in the stress identification task, since stress in Arabic is predictable, a conclusion that points to the effect brought by the native language of the speaker. This conclusion also seems to be supported by statements from the Arabic native speaking participants in a brief follow-up interview after the experiment. When asked to reflect on their performance in the experiment, the Arabic speaking participants expressed that lexical stress in Arabic was something that "they have never thought about" and something that was not explicitly talked about when they studied at school; additionally, they generally expressed uncertainty about their performance in the tasks as well, especially their performance in the stress identification task. Their impression and the lack of confidence in their performance seem to also imply the difficulty that they encountered in the stress perception even in their native language, resulting in the worsened performance in the stress identification task.

The effect of participants' native language is also exhibited on the performance of the Chinese-speaking participants in the stress production task. The Chinese L1 participants were found to be generally more accurate in stressing the correct syllable in the stimuli and they did so more quickly than their English L1 counterparts. Not only did the second-year Chinese L1 participants outperform both of their first-year and second-year English L1 counterparts, the first-year Chinese L1 participants were furthermore found to have better accuracy than both first-year and second-year English-speaking participants as well. This finding excludes the possibility of the effect contributed from the slightly longer weekly hours at the Chinese-speaking data collection site than at the English-speaking site (i.e. 7 hours versus 5 hours). Additionally, the accuracy of both Chinese L1 groups is above chance, while that of the English L1 participants was near chance, suggesting that the English L1 participants encountered some difficulty having their production recognized by the phonetic analyses as to correct placement of stress. Furthermore, although the Arabic L1 participants in the control group generally performed better than their non-native counterparts, this advantage does not reach statistical significance when compared with the second-year Chinese L1 participants. That is, the higher level Chinese L1 participants are very similar to the Arabic L1 participants in stressing words in the stimuli.

The reason for this disparity in performance between the Chinese L1 learners and their English L1 counterparts may be related to the way in which they utilize different stress correlates when producing stress. In a study comparing the production of English stress by L1 English speakers and L1 Mandarin Chinese speakers, (Chen et al., 2001) found that although both the English L1 and L1 Chinese speakers utilize all three correlates of stress (i.e., F0, intensity, and duration) to signify stress, the Chinese L1 participants tend to have a greater contrast in F0 and intensity between the stressed and unstressed syllables compared to their English L1 counterparts, where the three correlates of stress do not differ as much in their production of stressed and unstressed syllables. In another study that also compared English stress production by English L1 and Chinese L1 speakers, Lai (Lai, 2008) reported that part of speech, i.e. whether a word is either noun, verb or adjective and so on, seems to influence the way in which stress is realized by participants of different native languages. The results of the experiment showed that both English L1 and Chinese L1 participants used all three correlates to contrast their stressed and unstressed syllables in their production of nouns in English, which align well with the findings from Chen *et al.* (2001). However, when producing verbs in English, the two groups varied in the correlates utilized: the Chinese L1 participants continued to utilize all three correlates to contrast their production of stressed and unstressed syllables, while the English L1 participants did not use F0. Thus, the relatively poor performance of the English L1 participants might be due to the lack of utilization of F0 (which is also found in the present study), and the tendency to have smaller contrast in stress correlates between stressed syllables and unstressed ones compared to the Chinese L1 counterparts. Consequently, this disparity in the phonetic realization of English stress seems to be carried over to the production of Arabic stress, resulting in poorer accuracy in stress production. On the other hand, the tonal contours that are active in Chinese seem to encourage the Chinese L1 participants to fully utilize all correlates of stress. They were able to form a more distinctive contrast between stressed and unstressed syllables, which aligns more closely with how stress is realized by the Arabic speakers.

Typologically speaking, speakers of variable stress languages (i.e. stress could fall in different positions such as English) are generally expected to perform better than speakers of languages which do not have stress. The results of Altmann (2006) seems to support this expectation, as participants of non-stress languages (e.g., Japanese, Chinese and Korean) were found to be more variable and generally less accurate in placing stress on English nonsense words when compared to participants of languages with unpredictable stress (e.g., English and Spanish). These findings do not seem to align with the results reported in the present study, which may be due to typological differences between the target languages, i.e., English and Arabic. As discussed in previous chapters, although English is to an extent quantity/weight-sensitive (see (Domahs, Plag, & Carroll, 2014)), typologically speaking, it remains a language with unpredictable stress as the stress position cannot be confidently and consistently determined by a single algorithm. Arabic, unlike English, is a weight-sensitive language with typologically predictable stress. The location of stress can be confidently determined by an algorithm that evaluates the weight of each syllable. For learners who already have some pre-established weight schema in their first language, such as the English L1 participants in the present study, learning a weight-sensitive language with predictable stress such as Arabic also means to learning the weight schema for the target language. This might require re-setting the pre-existent schema and parameter for stress in their L1, resulting in a slower acquisition curve. In contrast, for learners such as the Chinese L1 participants whose native language does not have stress, the acoustic realization of stress and weight schema might appear to be a more salient feature, promoting acquisition (Ellis, 2006a, 2006b). This is not to suggest that acquisition of such a feature is blocked; rather, re-setting the parameters for stress and re-learning the weight schema might require more time/higher levels of proficiency (see Garcia 2016 for an example of English L1 speakers learning extrametricity for stress assignment in L2 Portuguese and reaching similar preference as the native speakers at advanced level), as only first-year and second-year English L1 were recruited in the present study. Lastly, the relatively lower accuracy in stress production by the English L1 participants might be compounded by the general lack of attention to stress in Arabic, which will be discussed in more detail in the following section.

5.1.2 The Effect of Proficiency

The second cluster of findings is related to the second research question of the study, restated below.

> RQ 2. Does the proficiency of the learners influence their performance in stress perception and production, and, if so, how?

The experimental results suggest that proficiency is generally not a significant indicator for the perception and production of stress by L2 learners in nonsense words, as no significant effect was found in the three main tasks of the experiment. However, for real words, proficiency seems to inform stress perception and production, as a significant difference in performance is found in the warm-up speed reading drill.

Intuitively, one would expect learners with a higher level of proficiency (which roughly translate to more input exposure) to perform better in the target language, either with higher accuracy or better fluency. A trend toward improved performance, especially accuracy, was detected in the participants' responses as their proficiency level increased; however, the exhibited improvement did not reach statistical significance, which is found in other L2 stress studies as well (e.g., Post da Silveira 2011).

The part of the experiment where a proficiency effect was found was in the speed reading drill. Possible explanations for such an effect might be approached from two angles: the naturalness of the wordlist and the modality of the drill. In this warm-up drill, participants were required to read out a list of Arabic real words. One of the advantages of higher level of proficiency is the possession of larger vocabulary in the target language, which could positively contribute to the faster time completion exhibited by the second-year participants as they might have better familiarity with some of the words in the wordlist used for the drill. Additionally, in the main tasks, the stimuli were presented one at a time and a brief break was given between stimuli, whereas in the speed reading drill, the stimuli, i.e. the wordlist, were presented right away with no breaks in between. Since the participants were timed, there were no breaks in between drills. This might have required more on-line processing familiarity and better capability in consecutive word recognition capacities – skills at which more proficient learners are more adept, which in turn results in the differentiating performance in the drill between participants of different levels of proficiency.

A possible explanation for the general lack of significant proficiency effects on participants’ performance in the main tasks might be due to the proficiency level of the participants themselves. Both the first-year and second-year participants were recruited during the middle of their second semester, which could be roughly translated to 1.5 and 3.5 semesters worth of input, levels which may not be sufficient enough to develop an sufficient working knowledge of the language through which lexical learning becomes more robust (see Nation, 2001; see also Alhawary, 2013 for a similar explanation of acquiring L2 synonyms). Hence, to explore proficiency effects on Arabic L2 stress perception and production more effectively and reliably, future research should recruit Arabic L2 learners from higher levels.

5.1.3 Frequency Effect

The third cluster of findings address the following research questions, regarding the effect of frequency on stress perception and production:

RQ 3. Does the frequency of relevant structures for stress (i.e., position of the stressed syllable, syllabic structure of the stressed syllable, and stress pattern of the word) influence the participants’ performance on stress perception and production, and, if so, how?

RQ 4. Does frequency have a different effect on stress perception as compared to stress production?

RQ 5. Does frequency have a different effect on learners as compared to native speakers?

As the results suggest, frequency does influence stress perception and production, but only partially. It is more evident in the production than in perception, and it seems to influence both native and non-native speakers.

Additionally, the predicted frequency effect is that participants' fluency and accuracy will be significantly and positively associated with the frequency of the stimuli' stress patterns. However, overall, the present study found not a global effect, but rather a partial effect. With certain stress patterns, the effect of frequency could be evident, while with others the evidence somehow is mixed or even clearly contradicts the predictions, which is not uncommon for frequency/probability-based studies (see Jouravlev and Lupker 2015 for an example of modeling bisyllabic stress pattern in Russian).

While no consistent trend of frequency was found in the stress identification task, a frequency effect seems to be at play in the stress production task, as participants were able to produce stress more rapidly and accurately when the stress pattern of the stimuli was frequent. This trend is especially salient when the stimuli are short, as significant differences in performance are often found within the trisyllabic stimulus groups, and less so in the quadrisyllabic stimulus groups. Moreover, significant differences in performance that support the prediction made by frequency effect in the study is more often found within stimuli belonging to stimulus group 1. This may be due to the likelihood that the frequency contrast in stimulus group 1 is the largest, as the relatively-frequent stress pattern in the stimulus group [ˈCV.CV.CV] is more frequent in terms of the frequency of stress pattern and stress position than its relatively infrequent counterpart [CVC.ˈCVC.CVC]. Such preference for higher frequency contrast has been demonstrated in previous studies in L2 stress production. In modeling the stress position for bisyllabic words in Russian, Jouravlev and Lupker (2015) implemented a probabilistic model with three non-lexical factors to determine whether the stress pattern is either iambic or trochaic. The model showed strong predicting power with significantly positive correlation between the probability calculated by the model for a given stress pattern and the proportion of that stress pattern assigned by the Russian L1 speakers. However, the model seemed to perform best at extreme ends (100% and 0%) of the probability distribution and was comparatively inconsistent for stress pattern whose predicted probability is in between, say 30% or 60%. With respect to the results from the present study, it is understandable that the participants did relatively well/poorly for stress patterns whose frequency is on the extreme ends of the frequency distribution, which accordingly results in significant difference in performance. It is tempting to argue here that the degree of contrast in frequency, as well as the length of the stress pattern, seem to affect the extent to which frequency informs the perception and production of stress. Such a conclusion would undoubtedly require more empirical data from future studies to validate.

Finally, although it is less relevant to the focus of the present study, the experimental results also showed that frequency seems to inform the accuracy of responses in the lexical decision task. Precisely, the relative frequency of the stress pattern seems to bias the participants' ability, especially that of the native speakers, to reject a nonsense word: the participants were found to be less willing to reject nonsense words with frequent stress patterns than ones with infrequent stress patterns. It seems to be the case that a nonsense word would appear to be more like a real word to the participants when it has a frequent stress pattern which, in turn, would impede the participants' willingness, learners and native speakers alike, to reject the word as non-real.

5.1.4 Other External Factors

In addition to the main factors discussed in the above sections, the present study implemented a post-experiment questionnaire to speculate on other factors that are otherwise impossible to assess in the experiment in order to answer the last research question, as given below:

RQ 6. Do other external factors (specifically explicit knowledge of lexical stress, instruction, self-awareness and extracurricular engagement with the target language) influence learners' performance on stress perception and production, and, if so, how?

Although the questionnaire is limited in length, the results yielded from it are a great source to reflect on how these factors could potentially explain the experiment results and inform future stress perception and production studies.

In the questions about instruction and textbook, a good portion (approximately 70%) of the English L1 and Chinese L1 participants, regardless of their level of proficiency, agreed that the instructor has more or less introduced stress rules in Arabic to them. However, when asked about the specific stress rules that the instructor has introduced, most of the participants mentioned "long vowel" and significantly fewer of them responded "gemination". Although these two items do occasionally coincide with the stressed syllable in Arabic, relying on only these two rules could not resolve stress patterns that have neither long vowel nor gemination. It would be rather unlikely for the instructor to introduce the algorithmic description for stress in Arabic as well. In addition, explicit and comprehensive instruction of stress rules in Arabic is generally lacking in current Arabic curricula.

Upon a closer examination of the textbooks used in both data collection sites, none of the textbooks (i.e., *Al-Kitaab* and *Al-'Arabiyya Bayna Yadayk*) includes a section specifically dedicated to lexical stress in Arabic. This lack of inclusion in the textbook makes it challenging for the instructors to systematically introduce stress in Arabic. Besides, the curricula themselves have little room to include activities to help the learners practice the perception and production of stress, as there are already many drills and activities that the instructors need to get to in one session. Therefore, the instruction on lexical stress in Arabic, if any, is in most cases incidental, and often in the form of recast, i.e. by the teacher repeating the mispronounced word in the correct form. Lastly, since suprasegmental features such as stress are oftentimes marginal to the evaluation of standardized test, there is even less incentive for the instructor to focus on this very aspect of the language. On the other hand, with regard to the experiment of the study, even with the general lack of attention to lexical stress in the curricula arguably in both data collection sites, differences are still found among the participants, in that the Chinese L1 participants were found to outperform the English L1 participants in the accuracy of their production of stress, which could only be informed by the native language (since both groups received comparable instruction in terms of stress) of the participants, highlighting the effect of L1 transfer.

Another useful finding from the questionnaire is related to the short quiz that examined participants' explicit knowledge of stress in Arabic. All participants from the control group and the non-native groups scored above chance (in the 60% to 70% range), and the English L1 participants were found to have significantly higher scores on this section than the Chinese L1 participants, but not better than the native Arabic participants. Firstly, the disparity in quiz performance between the English L1 participants and Chinese L1 participants seems to align well with their performance in the stress identification task in the experiment: the former on average had better accuracy in stress perception than the latter, albeit not reaching statistical significance. However, with respect to the native Arabic participants, their performance in the quiz did not seem to align well with their performance in the stress identification task, where the accuracy of their perception was below chance, and was significantly worse than that of the non-native speakers. Comparing the results from the quiz and the stress identification task, it becomes clear that the native Arabic participants to a great extent knew where the stress should fall, as indicated by the relative higher score in the quiz, but they did not seem to hear it in the stress identification task. This comparison seems to suggest that the Arabic native speakers possessed some degree of explicit knowledge (such as higher tendency on long vowels and geminates but never the specific rules that makes the algorithm for stress) about lexical stress that is similar or even slightly better than that of the non-native participants. However, when presented with tasks that require somehow implicit processing of stress in both real and non-words, the explicit knowledge does not seem to inform this implicit processing, resulting in the lower accuracy in the task. All else being equal, the result seems to suggest that their L1 could influence their perception of stress, again highlighting the effect from the L1 even for native speakers of the target language.

5.2 Limitations of The Study

There are a number of limitations which should be taken into account in future related research. First, the dataset implemented to estimate the frequency distribution of stress-related cues involves only minimal inflectional morphology. As in many other dictionaries, various forms of a verb, noun, or adjective could not be exhaustively listed due to the length of the dictionary. Estimating frequency using only dictionary entries may not accurately reflect actual language use and stress distribution which, in turn, might have introduced biases toward the estimate of frequency distribution for stress in this study.

Second, inclusion of first and second year L2 learners of Arabic is not sufficient enough to test proficiency effect as well as frequency effect robustly and reliably. Future research should aim to both recruit more participants per level, as well as more participants from higher proficiency levels. Note that recruiting a sizable group of higher level participants remains a challenge if the same recruitment criteria adopted in the present study were to be met. As proficiency level increases, the study body decreases, which is accompanied by higher likelihood of additional exposure to the target language from travelling or studying in the Arabic-speaking world.

Third, the binary categorization of frequency may not adequately reflect the actual differences between one stress pattern and another. Rather than relying on raw frequency, frequency percentile or a normalized frequency index may be more reflective of how a stress pattern is positioned among other stress patterns.

Finally, other frequency cues, such as segmental cues for both vowels and consonants, and other non-lexical cues (as identified in Jouravlev and Lupker 2015) should also be considered in designing the stimuli. As mentioned above, a confound in vowel length (which resulted in one fewer stress pattern out of the eight identified patterns) could have been rectified if segmental cues had been considered as the variable of the stimuli. This would have allowed more space for comparison.

It will also be useful for future research examining frequency effects on Arabic L2 stress perception and production to incorporate other relevant frequency cues, such as root and part of speech, in addition to the ones identified in the present study. Future studies should also expand the pool of learners to include those who are speakers of L1s which have other distinctive prosodic systems, such as French and Japanese, which exhibit an invariable stress system and pitch-contour system respectively. Denser data of naturally-produced speech by L2 learners of Arabic will be useful in order to understand how stress is realized in context rather than in isolation.

5.3 Conclusion and Pedagogical Implications

To conclude, findings from the present study seem to only partially support frequency effect predictions, as no global effect for frequency on stress perception and production was found. The findings indicate that the effect of frequency is evident only when the contrast in frequency is sufficiently large and the input is shorter in length. On the other hand, the native language of the learners or the effect of L1 transfer is found to provide a better account of the similarity and disparity exhibited in stress perception and production experiment. The characteristics of the prosodic system in the learners' native language seem to continue to be evident in the learners' perception and production of Arabic L2 stress, resulting in differences in performance between learners of the two different language backgrounds: English L1 and Chinese L1. Nevertheless, the effect of native language alone cannot account for all the variability found in learners' perception and production of stress, in particular the variability in performance within the same group of participants.

Notwithstanding the limitations mentioned above, the present study does reveal several pedagogical implications relevant to Arabic L2 stress perception and production. As reported in the earlier section, the general observation from the post-experiment interaction with participants in the study was similar to what was noted in Ryding (2013), in that lexical stress in Arabic is a topic that is seldom mentioned and rarely explicitly activated in the Arabic language classrooms of the native speakers and L2 learners alike. In the foreseeable future when they become Arabic instructors, the L1 Arabic instructors would be less likely to incorporate stress in the curriculum because although “they know where to stress . . . [they] could not quite tell [when they were] hearing it.” For L1 Arabic instructors, lexical stress is something that “they have never thought about before”, “is not important in Arabic,” or something that “they have never been taught before.” L2 Arabic instructors would also be less inclined to incorporate lexical stress in the curriculum, as they were not explicitly taught lexical stress rules when learning the language. Both future sets of instructors are also constrained by time and by materials: the textbook does not provide sufficient materials for practicing listening and producing stress, and class time needs to be dedicated to other equally important materials as well.

These are all realistic considerations for not including lexical stress in the curriculum. Yet it is an integral part of pronunciation that relates to many other important aspects of second language acquisition. Lexical stress has a special place in Arabic, and systematically excluding it in classrooms and curricula perpetuates a vicious cycle of ignorance. As shown in the experimental results, this lack of awareness and activation of stress results in relatively low accuracy in the stress production task. There is a lack of any significant improvement from the first to second year of language instruction in the English L1 participants, who arguably are in desperate need for significantly more input to help them acquire a stress system that is predictive and weight-sensitive – one that is distinctively different from that of their native language. The experimental findings in turn call for more efforts from the instructor and textbook writer in the English-speaking world to integrate more content, drills, and activities aiming at developing the ability to perceive and produce stress from early on.

Another pedagogical implication from the study stems from the learners’ general performance on longer words. Recall the study’s findings that the participants generally performed better when the word was shorter (three rather than four syllables). From a usage-based perspective, and due to availability of words in the input, the better performance with short words is because stress patterns in trisyllabic words are a lot more frequent and are therefore better-entrenched in the minds of the learners than those of quadrisyllabic words. This observation implies that in order for the learners to perform better in stress perception and production, they need more input containing quadrisyllabic words, starting from the very beginning of learning. The analysis of the frequency distribution of stress patterns in Al-Kitaab curriculum echoes this proposition. In the analysis, it was found that the frequency distribution extracted from the frequency dictionary is significantly positively correlated with that of the curriculum, which might appear at first glance to be good for the learners but is in fact adverse for learning when taking usage-based view on language learning into consideration. A positive correlation with the frequency dictionary means that some stress patterns in the curriculum are frequent while some are not, just like in the frequency dictionary; however, this also means that the learners will do better with the frequent stress patterns as they are better entrenched in the minds of the learners and less so with the infrequent stress patterns, just like these participants. Ideally, the frequency distribution for the stress patterns in the curriculum should reach near zero correlation with the frequency distribution found in the frequency dictionary or at the very least increase the appearance of words that have less frequent stress patterns. This would give each stress pattern equal opportunity to be encountered and further acquired, resulting in a more balanced ability in stress perception and production. Since some stress patterns are inevitably more frequent, therefore, textbook writers and instructors should not refrain from including words with less frequent stress patterns just because learner's proficiency level is assumed to be too low. As long as the inclusion of such words is done judiciously and the words are relevant to the content and functions being introduced, such inclusion should be justified.

Finally, the results from the experiment are relevant to curricular planning at the institutional level, especially for institutions that plan to teach two varieties of Arabic at the same time. As generally acknowledged, stress is among the many aspects of L2 phonology which need to receive adequate instructional attention in the foreign language classroom at the early stages of learning. Lexical stress also requires considerable processing effort from the learners, from retrieving underlying structure, syllabifying the word, and then applying stress algorithm, each of which requires substantial endeavor from the speakers not to mention ones that are starting to learn the language. As discussed in the Chapters 2 and 3, there is more than one algorithm for stress in Arabic, and each variety of Arabic could have its unique algorithm for stress. The experimental result show that there is still a lot of room for the learners to improve on their perception and production of stress and they were only learning one variety of Arabic. Their performance could be expected to be yet poorer if they were learning two varieties of Arabic at the same time, especially ones that have different stress algorithms such as learning MSA and Egyptian at the same time. Moreover, the two varieties do not only vary in the stress algorithms, the rules for realizing surface structure to syllabification could vary drastically as well. From the perspective of usage-based language acquisition, learning another variety of Arabic means learning more new constructions. Accordingly, in order to learn more constructions, more input is needed for the constructions to be entrenched in the minds of the learners, which could require substantially more time from the class as well as instructional materials, as noted in Alhawary (2013). Relating this perspective back to Arabic curriculum planning, there is no doubt that teaching two varieties (MSA along with colloquial) of Arabic has several practical and educational merits; however, it also requires substantially more teaching materials and more input to activate the various aspects (such as syntactic, morphological and semantic aspects of the target variety) of the additional variety of Arabic, not just lexical stress alone. Simply put, more forms to be taught should translate into more input exposure and more teaching/instruction time.

Appendix 3.1 Task Description for the Warm-Up Activity

In English:

In this experiment, you will be given a list of Arabic words.

You have two tasks:

(1) Please read through the list quickly but clearly without any pause. The list should be read from right to left, one row at a time.

(2) Press [any key] when you have read through the entire list.

Let us practice with a shorter list first.

Press [any key] when you are ready to start.

In Chinese:

在這個實驗，你會看到數行阿拉伯文字

你有兩項任務:

(1) 請迅速並清楚地念出每行的所有的阿拉伯文字。請由右至左、一次一行地念。

(2) 當你念完所有的字之後，請按下 [任何鍵]。

接下來讓我們先練習一下。

請按 [任何鍵] 開始練習。

Appendix 3.2 Wordlists Used for the Warm-Up Activity

A. For practice

← Read this way

كابُوس	هٰؤُلاءِ	اِهْتَمَّ	كانَ	إنْفلَوَنْزا	بُروفِسور	آسِيَوِيّ	آخَر
مالِيَّة	بُؤْرَة	مُتَواصِل	أَسْمَى	هٰكَذا	مُتَمَنِّي	اِسْتَخْدَمَ	تَقَدُّم

B. For Testing

← Read this way

خِلالَ	اِسْتَطاعَ	مَأساة	عَرَفَ	مُؤَسَّسَة	شَغَب	رِئَة	جُغْرافِيا
إجازَة	اِعْتَقَلَ	تَسْتَلْزِم	اِفْتَتَحَت	قُرُب	مَلَأت	مُتَظاهِر	اِسْتِعانَة
مُتَعاقِب	اِسْتِقالَة	هَدَفَ	اِنْقَلَبَت	فَضِيلَة	اِكْتَشَفَ	تَسْتَقْبِل	وَلَدَت
مُتَواضِع	تَسْتَعْمِل	اِنْدَرَجَ	اِسْتِراحَة	نَطَرَت	دَعَمَ	اِنْتَظَرَت	قَبِيلَة
اِسْتَحالَ	ضَبَطَ	اِنْعَكَسَت	مُتَفاوِت	وَظِيفَة	اِحْتَفَظَ	وَقَفَت	تَسْتَوْرِد
أضْعَف	مُواطِن	فِيزِياء	عِنْدَئِذٍ	لِماذا	خِيرَة	هٰذِهِ	مُوبايِل

Appendix 3.3 Task Description for the Stress Identification Task and Lexical Decision

<u>In English:</u>

Welcome. In this experiment, you will hear a recording of words.

After hearing the word, you have two tasks:

A. Identify the stressed syllable of the word as quickly as possible by pressing:

- [1] if you hear the stress on the 1st syllable,
- [2] if you hear the stress on the 2nd syllable,
- [3] if you hear the stress on the 3rd syllable, or
- [4] if you hear the stress on the 4th syllable.

B. indicate if the word is a real Arabic word or not as quickly as possible by pressing:

- [Yes] to indicate that you think the word is a real Arabic word or
- [No] to indicate that you think the word is not a real Arabic word

You will be shown each word, broken into syllables. However, all long and short vowels are not included.

There are 148 trials (including practice ones) in all.

Let us practice with a few words first.

Press [any key] when you are ready to start.

In Chinese:

歡迎，在這實驗中您將會聽到數段錄音。每段錄音都是一個字。

每個字可能是實際存在於阿拉伯文中的字，也有可能是不存在於阿拉伯文中的字。

每聽完一段錄音之後，您有兩項任務:

1. 盡可能快速地指出您認為有重音的音節:

- 如果您認為重音在第一音節，請按 [1] ，

- 如果您認為重音在第二音節，請按 [2] ，

- 如果您認為重音在第三音節，請按 [3]，

- 如果您認為重音在第四音節，請按 [4]。

2. 盡可能快速地指認聽到的字是否為存在於阿拉伯文中的字

- 如果您認為該字實際存在於阿拉伯文中，請按 [Yes] ，

- 如果您認為該字不存在於阿拉伯文中，請按 [No]。

你看到的每個將會依音節分隔，然而，字沒有標音且長母音將不顯現

本段實驗含練習用的錄音總共有 148 段。

首先讓我們先練習一下。

請按 [任何鍵] 開始練習。

Appendix 3.4 Task Description for the Stress Production Task

In English:

Welcome. In this experiment, you will see a word in Arabic script twice.

Each time the word appears, please read the word as quickly as possible once.

Make sure that you read all the short vowels given.

There are 148 trials (including the ones for practice) in all.

Let us practice with a few words first.

Press [any key] when you are ready to start.

In Chinese:

歡迎! 在這個部份的實驗你將會看到一個用阿拉伯文字母寫成的字兩次。

每當字出現的時候，請盡可能快速並清楚地念出這個字。

請照著所給的標音念。

本段實驗含練習用的字總共有 148 個。

首先讓我們先練習一下。

請按 [任何鍵] 開始練習。

Appendix 3.5 Questionsf for Post-Experiment Questionnaire

3.5.1 Questions for Instruction and Textbook:

Question: To what extent do you agree with the following statements?
My Arabic teacher has introduced stress rules in Arabic
→ (if Strongly Agree or Somewhat Agree) List the stress rules that your Arabic teacher has introduced
My Arabic teacher regularly emphasizes the importance of putting stress on the right syllable
My Arabic teacher always corrects me when I don't put the stress on the correct position of the word
We regularly practice producing stress in class.
The textbook that I use has drills that help me learn how to produce stress correctly
The textbook that I use has detailed explanation of stress rules in Arabic
→ (if Strongly Agree or Somewhat Agree) List the stress rules that your Arabic teacher has introduced

3.5.2 Questions for Self-Awareness:

Question: To what extent do you agree with the following statements?
I completely understand the rules of stress in Arabic
I always think about stress rules before I say anything
I can identify where the stressed syllable is without saying it out loud
I always produce stress correctly for shorter words
I always produce stress correctly for longer words
I always pay attention to where I put my stress in speaking
l always read things out loud in Arabic
I always know when people do not produce stress correctly
I have read about stress rules in Arabic outside of class
→ (if Strongly Agree or Somewhat Agree) What are stress rules that you learned outside of class?

3.5.3 Questions for Explicit Knowledge of Stress Rules:

Question: Please indicate whether you think the statement is true or false
Stress always falls on the long vowels
Stress can occur anywhere in the word
Stress in Arabic has no rule
Stress does not always occur at the beginning of the word
Stress always falls on the end of the word
Stress always follow a specific type of vowel
Stress always follow a specific type of consonant.
Stress does not always fall on the same place whether the word is singular or plural
The feminine ending "ة" has an effect on where to put the stress
There is always more than one way to stress the same word.

3.5.4 Questions for Biographical Information and Language Background

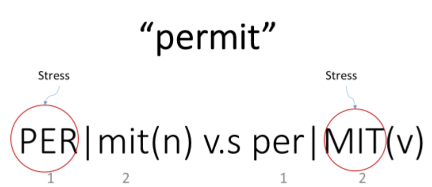
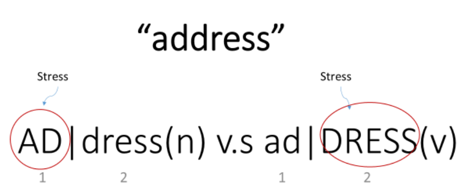
Question:
Gender
What is your age?
What is your major?
What year are you in at the university?
What language do you use at home with your parents?
What language do you use at school?
What language do you use with friends?
Do you also know other languages?

→ (if Yes) What language?
→ (if Yes) How many years have you studied it?
Have you studied Arabic before entering college?
How many semesters have you studied Arabic before college?
How many semesters have you studied Arabic since you entered college (excluding this semester)?
Have you visited any Arabic-speaking countries for more than 2 weeks?
→ (if Yes) Which country did you visit?
→ (if Yes) How long was the visit?
Did you study Arabic during this visit?
→ (if Yes) How long did you study Arabic during this visit?
→ (if Yes) How many hours did you go to Arabic class weekly during this visit?
Have you studied any Arabic dialect?
→ (if Yes) What was the dialect
→ (if Yes) How many years have you studied it?

3.5.5 Questions for Social Engagement with the Target Language:

Question: To what extent do you agree with the following statements?
I have a lot of friends who are native speakers of Arabic
I have a lot of opportunities to use Arabic outside of class
I find myself speaking Arabic all the time outside of class
I find myself listening to Arabic all the time outside of class
I find myself reading Arabic articles all the time outside of class
I find myself writing in Arabic all the time outside of class

Appendix 3.6 Introduction to Lexical Stress in English

# Slide	Screenshot of the Slide	Narrative
1		Hi, today we are going to work on lexical stress, as in stress in a word. First, we are going to use some examples from English to help you understand better what stress sound like in a word. In English, sometimes the same word could have two different pronunciations. The difference between the two pronunciations is usually the difference in stress position, that is, which part of the word that gets stressed.
2		For example, the word “permit” spelled p-e-r-m-i-t] has two pronunciations. Permit and permit. The word has two parts, for the word “permit” as in parking permit, the stress falls on the first part of the word, whereas for “permit” as in “I permit you to do something”, the stress falls on the second part of the word.
3		Another example is “address” spelled a-d-d-r-e-s-s. The word also has two pronunciations, address and address. For the word “address” as in “billing address”, the stress falls on the first part of the word whereas for the word “address” as in “the president addresses the congress”, the stress falls on the second part of the word.

4	Parts = Syllables	At this point, I'm sure that you have gotten a better sense about lexical stress. Bear in mind that it is important to know how many parts, or syllables, are there in a word, and which part of the word is stressed. Now let's move on to some longer examples.
5	"addition" ad | DI | tion 1 2 3	The next example is the word "addition". How many parts or syllables are there? [pause] Ad-di-tion, three parts, and which part of the word is stress? [pause] the second part because ad-DI-tion.
6	"ridiculous" ri | DI | cu | lous 1 2 3 4	How about the word "ridiculous"? How many parts are there in the word? [pause], ri-dialect-cu-lous, four parts, and which part of the word is stressed? ri-DI-cu-lous, the second part of it.

7	“Colorado” co|lo|RA|do 1 2 3 4	Last example, for the word “Colorado”, how many parts are there? [pause] co-lo-ra-do, four parts, and which part is stressed? [pause] co-lo-RA-do, the third part.

Appendix 3.7 Type of Co-Occurrence of Root Consonants and Its Conditional Probability

Table A.3.7 Conditional Probability of Arabic Roots

Type of Co-Occurrence	Frequency Count	Conditional Probability	Type of Co-Occurrence	Frequency Count	Conditional Probability	Type of Co-Occurrence	Frequency Count	Conditional Probability
ʔ|ħ	15	8.1%	r|ʕ	15	6.8%	r|b	26	8.9%
r|ħ	14	7.6%	n|ʕ	12	5.5%	l|b	14	4.8%
w|ħ	12	6.5%	b|ʕ	11	5.0%	b|b	12	4.1%
d|ħ	10	5.4%	l|ʕ	11	5.0%	d|b	11	3.8%
m|ħ	10	5.4%	j|ʕ	11	5.0%	s|b	11	3.8%
sˤ|ħ	9	4.9%	d|ʕ	10	4.5%	ʕ|b	10	3.4%
j|ħ	9	4.9%	m|ʕ	10	4.5%	t|b	10	3.4%
b|ħ	8	4.3%	q|ʕ	9	4.1%	tˤ|b	10	3.4%
f|ħ	8	4.3%	w|ʕ	9	4.1%	q|b	8	2.7%
n|ħ	8	4.3%	z|ʕ	8	3.6%	w|b	8	2.7%
s|ħ	7	3.8%	sˤ|ʕ	7	3.2%	j|b	8	2.7%
l|ħ	6	3.2%	dʒ|ʕ	6	2.7%	ħ|b	7	2.4%
q|ħ	5	2.7%	s|ʕ	6	2.7%	h|b	6	2.0%
dʒ|ħ	4	2.2%	t|ʕ	5	2.3%	k|b	6	2.0%
z|ħ	4	2.2%	θ|ʕ	5	2.3%	n|b	6	2.0%
ʃ|ħ	3	1.6%	f|ʕ	4	1.8%	θ|b	6	2.0%
t|ħ	3	1.6%	tˤ|ʕ	4	1.8%	ɣ|b	5	1.7%
tˤ|ħ	3	1.6%	h|ʕ	2	0.9%	ʔ|b	5	1.7%
θ|ħ	3	1.6%	ʃ|ʕ	2	0.9%	x|b	3	1.0%
k|ħ	2	1.1%	ʕ|ʕ	1	0.5%	sˤ|b	3	1.0%
ħ|ħ	1	0.5%	k|ʕ	1	0.5%	z|b	3	1.0%
						ʔ|b	1	0.3%
d|d	17	5.9%	r|f	23	10.4%	ðˤ|b	1	0.3%
r|d	17	5.9%	w|f	14	6.3%	ð|b	1	0.3%
j|d	17	5.9%	d|f	11	5.0%	dʒ|b	1	0.3%
m|d	12	4.2%	q|f	11	5.0%	ʃ|b	1	0.3%
n|d	11	3.8%	t|f	11	5.0%			
w|d	11	3.8%	l|f	10	4.5%	b|ðˤ	1	25.0%
ʕ|d	9	3.1%	j|f	9	4.1%	dʒ|ðˤ	1	25.0%
l|d	9	3.1%	f|f	8	3.6%	r|ðˤ	1	25.0%
h|d	7	2.4%	ʔ|f	6	2.7%	r|ð	2	14.3%
ʔ|d	7	2.4%	ʕ|f	5	2.3%	f|ð	1	7.1%
ħ|d	6	2.1%	n|f	5	2.3%	k|ð	1	7.1%

f|d	6	2.1%
s|d	6	2.1%
b|d	5	1.7%
x|d	4	1.4%
q|d	4	1.4%
ɣ|d	3	1.0%
θ|d	3	1.0%
k|d	2	0.7%
dʒ|d	1	0.3%
tˤ|d	1	0.3%
l|ɣ	10	12.8%
w|ɣ	8	10.3%
j|ɣ	8	10.3%
d|ɣ	7	9.0%
r|ɣ	7	9.0%
m|ɣ	5	6.4%
b|ɣ	4	5.1%
f|ɣ	3	3.8%
n|ɣ	3	3.8%
sˤ|ɣ	2	2.6%
s|ɣ	2	2.6%
tˤ|ɣ	2	2.6%
θ|ɣ	2	2.6%
z|ɣ	2	2.6%
ðˤ|ɣ	1	1.3%
ʃ|ɣ	1	1.3%
t|ɣ	1	1.3%
r|dʒ	15	9.9%
w|dʒ	13	8.6%
b|dʒ	11	7.2%
l|dʒ	11	7.2%
d|dʒ	8	5.3%
dʒ|dʒ	8	5.3%
m|dʒ	8	5.3%
s|dʒ	8	5.3%
n|dʒ	7	4.6%
f|dʒ	6	3.9%
z|dʒ	6	3.9%
h|dʒ	5	3.3%
ʕ|dʒ	4	2.6%
θ|dʒ	4	2.6%

s|f	5	2.3%
ħ|f	4	1.8%
k|f	4	1.8%
sˤ|f	4	1.8%
z|f	4	1.8%
dʒ|f	3	1.4%
θ|f	3	1.4%
h|f	2	0.9%
x|f	2	0.9%
ʃ|f	2	0.9%
ð|f	1	0.5%
tˤ|f	1	0.5%
j|h	12	10.9%
r|h	11	10.0%
l|h	10	9.1%
d|h	9	8.2%
m|h	8	7.3%
n|h	8	7.3%
b|h	6	5.5%
w|h	6	5.5%
dʒ|h	5	4.5%
z|h	5	4.5%
q|h	4	3.6%
ʔ|h	2	1.8%
ʃ|h	2	1.8%
t|h	2	1.8%
tˤ|h	2	1.8%
f|h	1	0.9%
θ|h	1	0.9%
r|z	11	8.8%
l|z	10	8.0%
n|z	8	6.4%
m|z	7	5.6%
j|z	7	5.6%
ʕ|z	6	4.8%
w|z	6	4.8%
b|z	5	4.0%
f|z	4	3.2%
dʒ|z	4	3.2%
z|z	4	3.2%
ħ|z	3	2.4%

l|ð	1	7.1%
m|ð	1	7.1%
n|ð	1	7.1%
j|ð	1	7.1%
l|ʔ	16	8.7%
b|ʔ	11	6.0%
m|ʔ	11	6.0%
s|ʔ	11	6.0%
j|ʔ	11	6.0%
n|ʔ	10	5.5%
r|ʔ	9	4.9%
d|ʔ	8	4.4%
f|ʔ	6	3.3%
w|ʔ	6	3.3%
θ|ʔ	5	2.7%
z|ʔ	5	2.7%
dʒ|ʔ	4	2.2%
k|ʔ	3	1.6%
x|ʔ	3	1.6%
sˤ|ʔ	3	1.6%
ð|ʔ	2	1.1%
ʃ|ʔ	2	1.1%
t|ʔ	2	1.1%
ħ|ʔ	1	0.5%
h|ʔ	1	0.5%
ʔ|ʔ	1	0.5%
q|ʔ	1	0.5%
tˤ|ʔ	1	0.5%
r|k	13	8.9%
b|k	11	7.5%
l|k	11	7.5%
ʔ|k	10	6.8%
s|k	9	6.2%
f|k	8	5.5%
w|k	8	5.5%
d|k	6	4.1%
k|k	6	4.1%
m|k	6	4.1%
n|k	6	4.1%
h|k	5	3.4%
θ|k	4	2.7%

<table>
<tr><td>j|dʒ</td><td>3</td><td>2.0%</td><td>h|z</td><td>3</td><td>2.4%</td><td>j|k</td><td>4</td><td>2.7%</td></tr>
<tr><td>ħ|dʒ</td><td>2</td><td>1.3%</td><td>q|z</td><td>3</td><td>2.4%</td><td>z|k</td><td>3</td><td>2.1%</td></tr>
<tr><td>ʔ|dʒ</td><td>2</td><td>1.3%</td><td>ɣ|z</td><td>2</td><td>1.6%</td><td>ʕ|k</td><td>2</td><td>1.4%</td></tr>
<tr><td></td><td></td><td></td><td>ʔ|z</td><td>2</td><td>1.6%</td><td>ʃ|k</td><td>2</td><td>1.4%</td></tr>
<tr><td>r|x</td><td>15</td><td>14.7%</td><td>k|z</td><td>2</td><td>1.6%</td><td>t|k</td><td>2</td><td>1.4%</td></tr>
<tr><td>l|x</td><td>11</td><td>10.8%</td><td>x|z</td><td>2</td><td>1.6%</td><td>ħ|k</td><td>1</td><td>0.7%</td></tr>
<tr><td>b|x</td><td>8</td><td>7.8%</td><td></td><td></td><td></td><td>sˤ|k</td><td>1</td><td>0.7%</td></tr>
<tr><td>w|x</td><td>8</td><td>7.8%</td><td>q|l</td><td>15</td><td>5.4%</td><td></td><td></td><td></td></tr>
<tr><td>n|x</td><td>6</td><td>5.9%</td><td>l|l</td><td>14</td><td>5.0%</td><td>r|q</td><td>17</td><td>8.1%</td></tr>
<tr><td>d|x</td><td>5</td><td>4.9%</td><td>m|l</td><td>13</td><td>4.7%</td><td>b|q</td><td>13</td><td>6.2%</td></tr>
<tr><td>m|x</td><td>5</td><td>4.9%</td><td>b|l</td><td>12</td><td>4.3%</td><td>d|q</td><td>13</td><td>6.2%</td></tr>
<tr><td>tˤ|x</td><td>5</td><td>4.9%</td><td>ħ|l</td><td>11</td><td>3.9%</td><td>m|q</td><td>12</td><td>5.7%</td></tr>
<tr><td>j|x</td><td>5</td><td>4.9%</td><td>f|l</td><td>11</td><td>3.9%</td><td>l|q</td><td>9</td><td>4.3%</td></tr>
<tr><td>f|x</td><td>4</td><td>3.9%</td><td>w|l</td><td>10</td><td>3.6%</td><td>n|q</td><td>9</td><td>4.3%</td></tr>
<tr><td>sˤ|x</td><td>4</td><td>3.9%</td><td>j|l</td><td>10</td><td>3.6%</td><td>j|q</td><td>8</td><td>3.8%</td></tr>
<tr><td>s|x</td><td>4</td><td>3.9%</td><td>ʕ|l</td><td>8</td><td>2.9%</td><td>f|q</td><td>7</td><td>3.3%</td></tr>
<tr><td>ʃ|x</td><td>3</td><td>2.9%</td><td>d|l</td><td>8</td><td>2.9%</td><td>sˤ|q</td><td>7</td><td>3.3%</td></tr>
<tr><td>ð|x</td><td>2</td><td>2.0%</td><td>ʔ|l</td><td>6</td><td>2.2%</td><td>tˤ|q</td><td>7</td><td>3.3%</td></tr>
<tr><td>t|x</td><td>2</td><td>2.0%</td><td>k|l</td><td>5</td><td>1.8%</td><td>w|q</td><td>7</td><td>3.3%</td></tr>
<tr><td>z|x</td><td>2</td><td>2.0%</td><td>h|l</td><td>4</td><td>1.4%</td><td>ʕ|q</td><td>6</td><td>2.9%</td></tr>
<tr><td>ʕ|x</td><td>1</td><td>1.0%</td><td>s|l</td><td>4</td><td>1.4%</td><td>s|q</td><td>6</td><td>2.9%</td></tr>
<tr><td>x|x</td><td>1</td><td>1.0%</td><td>ɣ|l</td><td>3</td><td>1.1%</td><td>t|q</td><td>5</td><td>2.4%</td></tr>
<tr><td>θ|x</td><td>1</td><td>1.0%</td><td>dʒ|l</td><td>3</td><td>1.1%</td><td>ð|q</td><td>3</td><td>1.4%</td></tr>
<tr><td></td><td></td><td></td><td>t|l</td><td>3</td><td>1.1%</td><td>q|q</td><td>3</td><td>1.4%</td></tr>
<tr><td>r|m</td><td>21</td><td>7.7%</td><td>n|l</td><td>2</td><td>0.7%</td><td>θ|q</td><td>3</td><td>1.4%</td></tr>
<tr><td>l|m</td><td>15</td><td>5.5%</td><td>sˤ|l</td><td>2</td><td>0.7%</td><td>ħ|q</td><td>2</td><td>1.0%</td></tr>
<tr><td>m|m</td><td>15</td><td>5.5%</td><td>tˤ|l</td><td>2</td><td>0.7%</td><td>h|q</td><td>2</td><td>1.0%</td></tr>
<tr><td>d|m</td><td>14</td><td>5.2%</td><td>θ|l</td><td>2</td><td>0.7%</td><td>ʃ|q</td><td>1</td><td>0.5%</td></tr>
<tr><td>s|m</td><td>12</td><td>4.4%</td><td>z|l</td><td>2</td><td>0.7%</td><td>z|q</td><td>1</td><td>0.5%</td></tr>
<tr><td>n|m</td><td>9</td><td>3.3%</td><td>r|l</td><td>1</td><td>0.4%</td><td></td><td></td><td></td></tr>
<tr><td>ħ|m</td><td>8</td><td>3.0%</td><td></td><td></td><td></td><td>l|t</td><td>8</td><td>8.0%</td></tr>
<tr><td>ʕ|m</td><td>8</td><td>3.0%</td><td>f|n</td><td>15</td><td>5.6%</td><td>r|t</td><td>8</td><td>8.0%</td></tr>
<tr><td>w|m</td><td>7</td><td>2.6%</td><td>w|n</td><td>15</td><td>5.6%</td><td>b|t</td><td>6</td><td>6.0%</td></tr>
<tr><td>j|m</td><td>7</td><td>2.6%</td><td>s|n</td><td>13</td><td>4.8%</td><td>q|t</td><td>6</td><td>6.0%</td></tr>
<tr><td>z|m</td><td>7</td><td>2.6%</td><td>k|n</td><td>12</td><td>4.4%</td><td>m|t</td><td>5</td><td>5.0%</td></tr>
<tr><td>ʔ|m</td><td>6</td><td>2.2%</td><td>d|n</td><td>11</td><td>4.1%</td><td>w|t</td><td>5</td><td>5.0%</td></tr>
<tr><td>k|m</td><td>6</td><td>2.2%</td><td>j|n</td><td>11</td><td>4.1%</td><td>n|t</td><td>4</td><td>4.0%</td></tr>
<tr><td>h|m</td><td>5</td><td>1.8%</td><td>q|n</td><td>10</td><td>3.7%</td><td>ʕ|t</td><td>3</td><td>3.0%</td></tr>
<tr><td>t|m</td><td>5</td><td>1.8%</td><td>ʕ|n</td><td>9</td><td>3.3%</td><td>j|t</td><td>3</td><td>3.0%</td></tr>
<tr><td>sˤ|m</td><td>4</td><td>1.5%</td><td>b|n</td><td>9</td><td>3.3%</td><td>ħ|t</td><td>2</td><td>2.0%</td></tr>
<tr><td>tˤ|m</td><td>4</td><td>1.5%</td><td>m|n</td><td>8</td><td>3.0%</td><td>d|t</td><td>2</td><td>2.0%</td></tr>
<tr><td>x|m</td><td>3</td><td>1.1%</td><td>n|n</td><td>8</td><td>3.0%</td><td>f|t</td><td>1</td><td>1.0%</td></tr>
<tr><td>dʒ|m</td><td>2</td><td>0.7%</td><td>z|n</td><td>8</td><td>3.0%</td><td>ɣ|t</td><td>1</td><td>1.0%</td></tr>
</table>

<table>
<tr><td>q|m</td><td>2</td><td>0.7%</td></tr>
<tr><td>ʃ|m</td><td>2</td><td>0.7%</td></tr>
<tr><td>θ|m</td><td>2</td><td>0.7%</td></tr>
<tr><td>ɣ|m</td><td>1</td><td>0.4%</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>r|sˤ</td><td>16</td><td>11.8%</td></tr>
<tr><td>f|sˤ</td><td>13</td><td>9.6%</td></tr>
<tr><td>l|sˤ</td><td>13</td><td>9.6%</td></tr>
<tr><td>b|sˤ</td><td>12</td><td>8.8%</td></tr>
<tr><td>d|sˤ</td><td>10</td><td>7.4%</td></tr>
<tr><td>w|sˤ</td><td>9</td><td>6.6%</td></tr>
<tr><td>j|sˤ</td><td>9</td><td>6.6%</td></tr>
<tr><td>m|sˤ</td><td>8</td><td>5.9%</td></tr>
<tr><td>ħ|sˤ</td><td>5</td><td>3.7%</td></tr>
<tr><td>n|sˤ</td><td>4</td><td>2.9%</td></tr>
<tr><td>sˤ|sˤ</td><td>4</td><td>2.9%</td></tr>
<tr><td>ʕ|sˤ</td><td>3</td><td>2.2%</td></tr>
<tr><td>ɣ|sˤ</td><td>2</td><td>1.5%</td></tr>
<tr><td>x|sˤ</td><td>2</td><td>1.5%</td></tr>
<tr><td>h|sˤ</td><td>1</td><td>0.7%</td></tr>
<tr><td>k|sˤ</td><td>1</td><td>0.7%</td></tr>
<tr><td>t|sˤ</td><td>1</td><td>0.7%</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>r|s</td><td>10</td><td>7.1%</td></tr>
<tr><td>s|s</td><td>9</td><td>6.4%</td></tr>
<tr><td>l|s</td><td>7</td><td>5.0%</td></tr>
<tr><td>m|s</td><td>7</td><td>5.0%</td></tr>
<tr><td>w|s</td><td>6</td><td>4.3%</td></tr>
<tr><td>d|s</td><td>5</td><td>3.5%</td></tr>
<tr><td>f|s</td><td>5</td><td>3.5%</td></tr>
<tr><td>q|s</td><td>5</td><td>3.5%</td></tr>
<tr><td>j|s</td><td>5</td><td>3.5%</td></tr>
<tr><td>ʕ|s</td><td>4</td><td>2.8%</td></tr>
<tr><td>n|s</td><td>4</td><td>2.8%</td></tr>
<tr><td>tˤ|s</td><td>4</td><td>2.8%</td></tr>
<tr><td>b|s</td><td>3</td><td>2.1%</td></tr>
<tr><td>x|s</td><td>3</td><td>2.1%</td></tr>
<tr><td>ʔ|s</td><td>2</td><td>1.4%</td></tr>
<tr><td>k|s</td><td>2</td><td>1.4%</td></tr>
<tr><td>ħ|s</td><td>1</td><td>0.7%</td></tr>
<tr><td>h|s</td><td>1</td><td>0.7%</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>r|w</td><td>21</td><td>7.4%</td></tr>
<tr><td>dʒ|n</td><td>7</td><td>2.6%</td></tr>
<tr><td>sˤ|n</td><td>7</td><td>2.6%</td></tr>
<tr><td>ħ|n</td><td>6</td><td>2.2%</td></tr>
<tr><td>ʔ|n</td><td>5</td><td>1.9%</td></tr>
<tr><td>tˤ|n</td><td>5</td><td>1.9%</td></tr>
<tr><td>ʃ|n</td><td>4</td><td>1.5%</td></tr>
<tr><td>h|n</td><td>3</td><td>1.1%</td></tr>
<tr><td>t|n</td><td>3</td><td>1.1%</td></tr>
<tr><td>θ|n</td><td>3</td><td>1.1%</td></tr>
<tr><td>x|n</td><td>2</td><td>0.7%</td></tr>
<tr><td>r|n</td><td>2</td><td>0.7%</td></tr>
<tr><td>ð|n</td><td>1</td><td>0.4%</td></tr>
<tr><td>ɣ|n</td><td>1</td><td>0.4%</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>d|r</td><td>20</td><td>4.7%</td></tr>
<tr><td>b|r</td><td>19</td><td>4.5%</td></tr>
<tr><td>j|r</td><td>17</td><td>4.0%</td></tr>
<tr><td>f|r</td><td>16</td><td>3.8%</td></tr>
<tr><td>dʒ|r</td><td>16</td><td>3.8%</td></tr>
<tr><td>q|r</td><td>15</td><td>3.5%</td></tr>
<tr><td>m|r</td><td>14</td><td>3.3%</td></tr>
<tr><td>r|r</td><td>14</td><td>3.3%</td></tr>
<tr><td>ʕ|r</td><td>13</td><td>3.1%</td></tr>
<tr><td>k|r</td><td>13</td><td>3.1%</td></tr>
<tr><td>ħ|r</td><td>11</td><td>2.6%</td></tr>
<tr><td>ʔ|r</td><td>9</td><td>2.1%</td></tr>
<tr><td>w|r</td><td>9</td><td>2.1%</td></tr>
<tr><td>h|r</td><td>7</td><td>1.6%</td></tr>
<tr><td>s|r</td><td>7</td><td>1.6%</td></tr>
<tr><td>sˤ|r</td><td>6</td><td>1.4%</td></tr>
<tr><td>t|r</td><td>6</td><td>1.4%</td></tr>
<tr><td>tˤ|r</td><td>6</td><td>1.4%</td></tr>
<tr><td>ɣ|r</td><td>4</td><td>0.9%</td></tr>
<tr><td>ʃ|r</td><td>4</td><td>0.9%</td></tr>
<tr><td>z|r</td><td>4</td><td>0.9%</td></tr>
<tr><td>θ|r</td><td>3</td><td>0.7%</td></tr>
<tr><td>ð|r</td><td>2</td><td>0.5%</td></tr>
<tr><td>x|r</td><td>2</td><td>0.5%</td></tr>
<tr><td>n|r</td><td>2</td><td>0.5%</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>r|ʃ</td><td>15</td><td>15.0%</td></tr>
<tr><td>j|ʃ</td><td>9</td><td>9.0%</td></tr>
<tr><td>f|ʃ</td><td>6</td><td>6.0%</td></tr>
<tr><td>ʔ|t</td><td>1</td><td>1.0%</td></tr>
<tr><td>dʒ|t</td><td>1</td><td>1.0%</td></tr>
<tr><td>k|t</td><td>1</td><td>1.0%</td></tr>
<tr><td>x|t</td><td>1</td><td>1.0%</td></tr>
<tr><td>s|t</td><td>1</td><td>1.0%</td></tr>
<tr><td>t|t</td><td>1</td><td>1.0%</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>r|tˤ</td><td>14</td><td>12.5%</td></tr>
<tr><td>w|tˤ</td><td>10</td><td>8.9%</td></tr>
<tr><td>f|tˤ</td><td>8</td><td>7.1%</td></tr>
<tr><td>l|tˤ</td><td>8</td><td>7.1%</td></tr>
<tr><td>m|tˤ</td><td>8</td><td>7.1%</td></tr>
<tr><td>j|tˤ</td><td>6</td><td>5.4%</td></tr>
<tr><td>b|tˤ</td><td>5</td><td>4.5%</td></tr>
<tr><td>ʕ|tˤ</td><td>4</td><td>3.6%</td></tr>
<tr><td>ʔ|tˤ</td><td>4</td><td>3.6%</td></tr>
<tr><td>n|tˤ</td><td>4</td><td>3.6%</td></tr>
<tr><td>tˤ|tˤ</td><td>4</td><td>3.6%</td></tr>
<tr><td>s|tˤ</td><td>3</td><td>2.7%</td></tr>
<tr><td>ħ|tˤ</td><td>2</td><td>1.8%</td></tr>
<tr><td>h|tˤ</td><td>2</td><td>1.8%</td></tr>
<tr><td>ʃ|tˤ</td><td>2</td><td>1.8%</td></tr>
<tr><td>ɣ|tˤ</td><td>1</td><td>0.9%</td></tr>
<tr><td>q|tˤ</td><td>1</td><td>0.9%</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>r|θ</td><td>15</td><td>14.3%</td></tr>
<tr><td>b|θ</td><td>10</td><td>9.5%</td></tr>
<tr><td>w|θ</td><td>8</td><td>7.6%</td></tr>
<tr><td>m|θ</td><td>6</td><td>5.7%</td></tr>
<tr><td>l|θ</td><td>5</td><td>4.8%</td></tr>
<tr><td>q|θ</td><td>5</td><td>4.8%</td></tr>
<tr><td>θ|θ</td><td>5</td><td>4.8%</td></tr>
<tr><td>ʔ|θ</td><td>4</td><td>3.8%</td></tr>
<tr><td>n|θ</td><td>3</td><td>2.9%</td></tr>
<tr><td>ʕ|θ</td><td>2</td><td>1.9%</td></tr>
<tr><td>h|θ</td><td>2</td><td>1.9%</td></tr>
<tr><td>x|θ</td><td>2</td><td>1.9%</td></tr>
<tr><td>j|θ</td><td>2</td><td>1.9%</td></tr>
<tr><td>dʒ|θ</td><td>1</td><td>1.0%</td></tr>
<tr><td>k|θ</td><td>1</td><td>1.0%</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>d|j</td><td>14</td><td>6.2%</td></tr>
<tr><td>b|j</td><td>10</td><td>4.4%</td></tr>
</table>

<table>
<tr><td>d|w</td><td>16</td><td>5.7%</td><td>m|ʃ</td><td>6</td><td>6.0%</td><td>n|j</td><td>10</td><td>4.4%</td></tr>
<tr><td>l|w</td><td>15</td><td>5.3%</td><td>tˤ|ʃ</td><td>6</td><td>6.0%</td><td>l|j</td><td>9</td><td>4.0%</td></tr>
<tr><td>b|w</td><td>14</td><td>5.0%</td><td>w|ʃ</td><td>6</td><td>6.0%</td><td>ʕ|j</td><td>8</td><td>3.5%</td></tr>
<tr><td>q|w</td><td>14</td><td>5.0%</td><td>ʕ|ʃ</td><td>5</td><td>5.0%</td><td>r|j</td><td>8</td><td>3.5%</td></tr>
<tr><td>dʒ|w</td><td>12</td><td>4.3%</td><td>h|ʃ</td><td>5</td><td>5.0%</td><td>m|j</td><td>7</td><td>3.1%</td></tr>
<tr><td>s|w</td><td>12</td><td>4.3%</td><td>k|ʃ</td><td>5</td><td>5.0%</td><td>ʔ|j</td><td>5</td><td>2.2%</td></tr>
<tr><td>j|w</td><td>11</td><td>3.9%</td><td>ħ|ʃ</td><td>3</td><td>3.0%</td><td>q|j</td><td>5</td><td>2.2%</td></tr>
<tr><td>m|w</td><td>9</td><td>3.2%</td><td>ʔ|ʃ</td><td>3</td><td>3.0%</td><td>s|j</td><td>5</td><td>2.2%</td></tr>
<tr><td>ʕ|w</td><td>8</td><td>2.8%</td><td>b|ʃ</td><td>2</td><td>2.0%</td><td>f|j</td><td>4</td><td>1.8%</td></tr>
<tr><td>f|w</td><td>8</td><td>2.8%</td><td>d|ʃ</td><td>2</td><td>2.0%</td><td>t|j</td><td>4</td><td>1.8%</td></tr>
<tr><td>t|w</td><td>8</td><td>2.8%</td><td>ɣ|ʃ</td><td>2</td><td>2.0%</td><td>w|j</td><td>3</td><td>1.3%</td></tr>
<tr><td>h|w</td><td>7</td><td>2.5%</td><td>dʒ|ʃ</td><td>2</td><td>2.0%</td><td>z|j</td><td>3</td><td>1.3%</td></tr>
<tr><td>n|w</td><td>7</td><td>2.5%</td><td>q|ʃ</td><td>2</td><td>2.0%</td><td>ħ|j</td><td>2</td><td>0.9%</td></tr>
<tr><td>sˤ|w</td><td>7</td><td>2.5%</td><td>t|ʃ</td><td>2</td><td>2.0%</td><td>ɣ|j</td><td>2</td><td>0.9%</td></tr>
<tr><td>z|w</td><td>7</td><td>2.5%</td><td>x|ʃ</td><td>1</td><td>1.0%</td><td>h|j</td><td>2</td><td>0.9%</td></tr>
<tr><td>ʔ|w</td><td>6</td><td>2.1%</td><td>l|ʃ</td><td>1</td><td>1.0%</td><td>dʒ|j</td><td>2</td><td>0.9%</td></tr>
<tr><td>k|w</td><td>6</td><td>2.1%</td><td>n|ʃ</td><td>1</td><td>1.0%</td><td>sˤ|j</td><td>2</td><td>0.9%</td></tr>
<tr><td>θ|w</td><td>6</td><td>2.1%</td><td>ʃ|ʃ</td><td>1</td><td>1.0%</td><td>ʃ|j</td><td>2</td><td>0.9%</td></tr>
<tr><td>ħ|w</td><td>5</td><td>1.8%</td><td>z|ʃ</td><td>1</td><td>1.0%</td><td>θ|j</td><td>2</td><td>0.9%</td></tr>
<tr><td>tˤ|w</td><td>5</td><td>1.8%</td><td></td><td></td><td></td><td>j|j</td><td>2</td><td>0.9%</td></tr>
<tr><td>ʃ|w</td><td>3</td><td>1.1%</td><td></td><td></td><td></td><td>k|j</td><td>1</td><td>0.4%</td></tr>
<tr><td>ð|w</td><td>1</td><td>0.4%</td><td></td><td></td><td></td><td>x|j</td><td>1</td><td>0.4%</td></tr>
<tr><td>ɣ|w</td><td>1</td><td>0.4%</td><td></td><td></td><td></td><td>tˤ|j</td><td>1</td><td>0.4%</td></tr>
<tr><td>w|w</td><td>1</td><td>0.4%</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Appendix 3.8 Stimuli

Table A.3.8.1 Nonsense Words Derived from Unattested Root and Template:

# block	# syl	stress pattern	Nonsen wd	Pronunciation	Root	template	Real Word	Pronunciation	Frequency of Stress Pattern	Frequency of Stress Position	Frequency of Syllabic Structure	Conditional Probability
1	3	ˈcv.cv.cv	فَقَزَ	ˈfa.qa.za	fqz	ˈC1a.C2a.C3a	وَثِقَ	ˈwa.θa.qa	**+ (5.9)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
1	3	ˈcv.cv.cvc	فَقَزَت	ˈfa.qa.zat	fqz	ˈC1a.C2a.C3at	جَعَلَت	ˈdʒa.ʕa.lat	**- (3.2)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
1	3	cv.ˈcvv.cv	فِقَازَ	fi.ˈqa:.za	fqz	C1i.ˈC2a:.C3a	نَتِيجَة	na.ˈti:.dʒa	**+ (5.8)**	**- (5.9)**	**+ (6.1)**	**+ (3.3)**
1	3	cvc.ˈcvc.cvc	مُسْتَفقَزْ	mus.ˈtaf.qaz	fqz	mus.ˈtaC1.C2aC3	مُسْتَوْرَد	mus.ˈtaw.rad	**- (4.0)**	**- (5.9)**	**+ (6.1)**	**+ (2.5)**
1	4	cv.cv.ˈcvv.cvc	مُتَفَاقِز	mu.ta.ˈfa:.qiz	fqz	mu.ta.ˈC1a:.C2iC3	مُتَتالٍ	mu.ta.ˈta:.lin	**+ (5.0)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
1	4	cvc.ˈcv.cv.cv	اِفتَقَزَ	ʔif.ˈta.qa.za	fqz	ʔiC1.ˈta.C2a.C3a	اِعْتَمَدَ	ʔiʕ.ˈta.ma.da	**- (2.6)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
1	4	cvc.ˈcv.cv.cvc	اِفتَقَزَتْ	ʔif.ˈta.qa.zat	fqz	ʔiC1.ˈta.C2a.C3at	اِرْتَكَزَت	ʔir.ˈta.ka.zat	**+ (4.7)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
1	4	cvc.cv.ˈcvv.cv	اِفتِقَازَة	ʔif.ti.ˈqa:.za	fqz	ʔiC1.ti.ˈC2a:.C3a	اِسْتَجابَ	ʔis.ta.ˈdʒa:.ba	**- (3.8)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
2	3	ˈcv.cv.cv	عَمَسَ	ˈʕa.ma.sa	ʕms	ˈC1a.C2a.C3a	سَمِعَ	ˈsa.mi.ʕa	**+ (5.9)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
2	3	ˈcv.cv.cvc	عَمَسَت	ˈʕa.ma.sat	ʕms	ˈC1a.C2a.C3at	سَمَحَ	ˈsa.ma.ħat	**- (3.2)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
2	3	cv.ˈcvv.cv	عِمَاسَ	ʕi.ˈma:.sa	ʕms	C1i.ˈC2a:.C3a	ثَلاثَة	θa.ˈla:.θa	**+ (5.8)**	**- (5.9)**	**+ (6.1)**	**+ (3.3)**
2	3	cvc.ˈcvc.cvc	مُسْتَعمَسْ	mus.ˈtaʕ.mas	ʕms	mus.ˈtaC1.C2aC3	يَسْتَمْتِع	jas.ˈtam.tiʕ	**- (4.0)**	**- (5.9)**	**+ (6.1)**	**+ (2.5)**
2	4	cv.cv.ˈcvv.cvc	مُتَعَامِس	mu.ta.ˈʕa:.mis	ʕms	mu.ta.ˈC1a:.C2iC3	يَتَناوَل	ja.ta.ˈna:.wal	**+ (5.0)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
2	4	cvc.ˈcv.cv.cv	اِعتَمَسَ	ʔiʕ.ˈta.ma.sa	ʕms	ʔiC1.ˈta.C2a.C3a	اِشْتَغَلَ	ʔiʃ.ˈta.ɣa.la	**- (2.6)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
2	4	cvc.ˈcv.cv.cvc	اِعتَمَسَتْ	ʔiʕ.ˈta.ma.sat	ʕms	ʔiC1.ˈta.C2a.C3at	اِلْتَحَقَت	ʔil.ˈta.ħa.qat	**+ (4.7)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
2	4	cvc.cv.ˈcvv.cv	اِعتِمَاسَة	ʔiʕ.ti.ˈma:.sa	ʕms	ʔiC1.ti.ˈC2a:.C3a	اِسْتِطاعة	ʔis.ti.ˈtˤa:.ʕa	**- (3.8)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
3	3	ˈcv.cv.cv	مَثَفَ	ˈma.θa.fa	mθf	ˈC1a.C2a.C3a	فَقَدَ	ˈfa.qa.da	**+ (5.9)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
3	3	ˈcv.cv.cvc	مَثَفَت	ˈma.θa.fat	mθf	ˈC1a.C2a.C3at	عَقَدَت	ˈʕa.qa.dat	**- (3.2)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
3	3	cv.ˈcvv.cv	مِثَافَ	mi.ˈθa:.fa	mθf	C1i.ˈC2a:.C3a	قِيادَة	qa.ˈdʒa:.da	**+ (5.8)**	**- (5.9)**	**+ (6.1)**	**+ (3.3)**
3	3	cvc.ˈcvc.cvc	مُسْتَمثَفْ	mus.ˈtam.θaf	mθf	mus.ˈtaC1.C2aC3	مُسْتَهْدَف	mus.ˈtah.daf	**- (4.0)**	**- (5.9)**	**+ (6.1)**	**+ (2.5)**
3	4	cv.cv.ˈcvv.cvc	مُتَمَاثِف	mu.ta.ˈma:.θif	mθf	mu.ta.ˈC1a:.C2iC3	مُتَسائِل	mu.ta.ˈsa:.ʔil	**+ (5.0)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
3	4	cvc.ˈcv.cv.cv	اِمتَثَفَ	ʔim.ˈta.θa.fa	mθf	ʔiC1.ˈta.C2a.C3a	اِنْطَلَقَ	ʔitˤ.ˈta.la.qa	**- (2.6)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
3	4	cvc.ˈcv.cv.cvc	اِمتَثَفَتْ	ʔim.ˈta.θa.fat	mθf	ʔiC1.ˈta.C2a.C3at	اِشْتَرَطَت	ʔiʃ.ˈta.ra.tˤat	**+ (4.7)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
3	4	cvc.cv.ˈcvv.cv	اِمتِثَافَة	ʔim.ti.ˈθa:.fa	mθf	ʔiC1.ti.ˈC2a:.C3a	اِنْتِفاضَة	ʔin.ti.ˈfa:.dˤa	**- (3.8)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
4	3	ˈcv.cv.cv	قَصَخَ	ˈqa.sˤa.xa	qsˤx	ˈC1a.C2a.C3a	بَدَأَ	ˈba.da.ʔa	**+ (5.9)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**

4	3	ˈCV.CV.CVC	قَصَخَت	ˈqa.sˤa.xat	qsˤx	ˈC1a.C2a.C3at	عَلِمَت	ˈʕa.li.mat	**- (3.2)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
4	3	CV.ˈCVV.CV	قِصَاخَ	qi.ˈsˤa:.xa	qsˤx	C1i.ˈC2a:.C3a	حُكُومَة	ħu.ˈku:.ma	**+ (5.8)**	**- (5.9)**	**+ (6.1)**	**+ (3.3)**
4	3	CVC.ˈCVC.CVC	مُسْتَقْصَخْ	mus.ˈtaq.sˤax	qsˤx	mus.ˈtaC1.C2aC3	يَستَقبِل	jas.ˈtaq.bil	**- (4.0)**	**- (5.9)**	**+ (6.1)**	**+ (2.5)**
4	4	CV.CV.ˈCVV.CVC	مُتَقَاصِخ	mu.ta.ˈqa:.sˤix	qsˤx	mu.ta.ˈC1a:.C2iC3	يَتَنافَس	ja.ta.ˈna:.fas	**+ (5.0)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
4	4	CVC.ˈCV.CV.CV	اِقتَصَخَ	ʔiq.ˈta.sˤa.xa	qsˤx	ʔiC1.ˈta.C2a.C3a	اِخْتَلَفَ	ʔix.ˈta.la.fa	**- (2.6)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
4	4	CVC.ˈCV.CV.CVC	اِقتَصَخَتْ	ʔiq.ˈta.sˤa.xat	qsˤx	ʔiC1.ˈta.C2a.C3at	اِنتَظَرَت	ʔin.ˈta.ðˤa.rat	**+ (4.7)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
4	4	CVC.CV.ˈCVV.CV	اِقتِصَاخَة	ʔiq.ti.ˈsˤa:.xa	qsˤx	ʔiC1.ti.ˈC2a:.C3a	اِستِضافَة	ʔis.ti.ˈdˤa:.fa	**- (3.8)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**

(): Log Frequency

Table A.3.8.2 Nonsense Words Derived from Real Words:

# block	# syl	stress pattern	Nonsen wd	Pronunciation	Root	template	Real wd	Pronunciation	Frequency of Stress Pattern	Frequency of Stress Position	Frequency of Syllabic Structure	Conditional Probability
5	3	ˈCV.CV.CV	نَغَرَ	ˈna.ɣa.ra	Minimal Pair	n.a.	نَشَرَ	ˈna.ʃa.ra	**+ (5.9)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
5	3	ˈCV.CV.CVC	طَثَعَتْ	ˈtˤa.θa.ʕat	Minimal Pair	n.a.	طَلَعَتْ	ˈtˤa.la.ʕat	**- (3.2)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
5	3	CV.ˈCVV.CV	خِلازَ	xi.ˈla:.za	Minimal Pair	n.a.	خِلالَ	xi.ˈla:.la	**+ (5.8)**	**- (5.9)**	**+ (6.1)**	**+ (3.3)**
5	3	CVC.ˈCVC.CVC	مُسْتَخْثَم	mus.ˈtax.θam	Minimal Pair	n.a.	مُسْتَخْدَم	mus.ˈtax.dam	**- (4.0)**	**- (5.9)**	**+ (6.1)**	**+ (2.5)**
5	4	CV.CV.ˈCVV.CVC	مُتَظانِر	mu.ta.ˈðˤa:.nir	Minimal Pair	n.a.	مُتَظاهِر	mu.ta.ˈðˤa:.hir	**+ (5.0)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
5	4	CVC.ˈCV.CV.CV	اِكْتَشَزَ	ʔik.ˈta.ʃa.za	Minimal Pair	n.a.	اِكْتَشَفَ	ʔik.ˈta.ʃa.fa	**- (2.6)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
5	4	CVC.ˈCV.CV.CVC	اِنْزَلَعَت	ʔin.ˈza.la.ʕat	Minimal Pair	n.a.	اِنْدَلَعَت	ʔin.ˈda.fa.ʕat	**+ (4.7)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
5	4	CVC.CV.ˈCVV.CV	اِسْتِقاخَة	ʔis.ti.ˈqa:.xa	Minimal Pair	n.a.	اِسْتِقالَة	ʔis.ti.ˈqa:.la	**- (3.8)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
6	3	ˈCV.CV.CV	سَهَقَ	ˈsa.ha.qa	Minimal Pair	n.a.	سَبَقَ	ˈsa.ba.qa	**+ (5.9)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
6	3	ˈCV.CV.CVC	رَأَلَت	ˈra.ʔa.lat	Minimal Pair	n.a.	سَأَلَت	ˈsa.ʔa.lat	**- (3.2)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
6	3	CV.ˈCVV.CV	إعاغَة	ʔi.ˈʕa:.ɣa	Minimal Pair	n.a.	إعادَة	ʔi.ˈʕa:.da	**+ (5.8)**	**- (5.9)**	**+ (6.1)**	**+ (3.3)**
6	3	CVC.ˈCVC.CVC	مُسْتَلْسَم	mus.ˈtal.sam	Minimal Pair	n.a.	مُسْتَلْزَم	mus.ˈtal.zam	**- (4.0)**	**- (5.9)**	**+ (6.1)**	**+ (2.5)**
6	4	CV.CV.ˈCVV.CVC	مُتَواغِع	mu.ta.ˈwa:.ɣiʕ	Minimal Pair	n.a.	مُتَواضِع	mu.ta.ˈwa:.dˤiʕ	**+ (5.0)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
6	4	CVC.ˈCV.CV.CV	اِرْتَكَحَ	ʔir.ˈta.ka.ħa	Minimal Pair	n.a.	اِرْتَكَبَ	ʔir.ˈta.ka.ba	**- (2.6)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
6	4	CVC.ˈCV.CV.CVC	اِمْتَبَأَت	ʔim.ˈta.ba.ʔat	Minimal Pair	n.a.	اِمْتَلَأَت	ʔim.ˈta.la.ʔat	**+ (4.7)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**

6	4	CVC.CV.ˈCVV.CV	اِسْتِجاذَة	ʔis.ti.ˈdʒaː.ða	Minimal Pair	n.a.	اِسْتِجابَة	ʔis.ti.ˈdʒaː.ba	**- (3.8)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
7	3	ˈCV.CV.CV	عَشَفَ	ˈʕa.ʃa.fa	Minimal Pair	n.a.	عَرَفَ	ˈʕa.ra.fa	**+ (5.9)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
7	3	ˈCV.CV.CVC	كَتَسَت	ˈka.ta.sat	Minimal Pair	n.a.	كَتَبَت	ˈka.ta.bat	**- (3.2)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
7	3	CV.ˈCVV.CV	مَذِينَة	ma.ˈðiː.na	Minimal Pair	n.a.	مَدِينَة	ma.ˈdiː.na	**+ (5.8)**	**- (5.9)**	**+ (6.1)**	**+ (3.3)**
7	3	CVC.ˈCVC.CVC	مُسْتَقْزَل	mus.ˈtaq.zal	Minimal Pair	n.a.	مُسْتَقْبَل	mus.ˈtaq.bal	**- (4.0)**	**- (5.9)**	**+ (6.1)**	**+ (2.5)**
7	4	CV.CV.ˈCVV.CVC	يَتَباجَل	ja.ta.ˈbaː.dʒal	Minimal Pair	n.a.	يَتَبادَل	ja.ta.ˈbaː.dal	**+ (5.0)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
7	4	CVC.ˈCV.CV.CV	اِسْتَخَعَ	ʔis.ˈta.xa.ʕa	Minimal Pair	n.a.	اِسْتَمَعَ	ʔis.ˈta.ma.ʕa	**- (2.6)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
7	4	CVC.ˈCV.CV.CVC	اِعتَبَفَت	ʔiʕ.ˈta.ba.ðat	Minimal Pair	n.a.	اِعتَبَرَت	ʔiʕ.ˈta.ba.rat	**+ (4.7)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
7	4	CVC.CV.ˈCVV.CV	اِسْتِراثة	ʔis.ti.ˈraː.θa	Minimal Pair	n.a.	اِسْتِراحة	ʔis.ti.ˈraː.ħa	**- (3.8)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
8	3	ˈCV.CV.CV	وَثَدَ	ˈwa.θa.da	Minimal Pair	n.a.	وَجَدَ	ˈwa.dʒa.da	**+ (5.9)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
8	3	ˈCV.CV.CVC	لَرَجَت	ˈla.ra.dʒat	Minimal Pair	n.a.	خَرَجَت	ˈxa.ra.dʒat	**- (3.2)**	**+ (6.2)**	**+ (5.9)**	**+ (3.6)**
8	3	CV.ˈCVV.CV	هُفاكَ	hu.ˈfaː.ka	Minimal Pair	n.a.	هُناكَ	hu.ˈnaː.ka	**+ (5.8)**	**- (5.9)**	**+ (6.1)**	**+ (3.3)**
8	3	CVC.ˈCVC.CVC	يَستَعكِل	jas.ˈtaʕ.kil	Minimal Pair	n.a.	يَستَعمِل	jas.ˈtaʕ.mil	**- (4.0)**	**- (5.9)**	**+ (6.1)**	**+ (2.5)**
8	4	CV.CV.ˈCVV.CVC	مُتَضايِث	mu.ta.ˈdˤaː.jiθ	Minimal Pair	n.a.	مُتَضايِق	mu.ta.ˈdˤaː.jiq	**+ (5.0)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
8	4	CVC.ˈCV.CV.CV	اِعْتَشَدَ	ʔiʕ.ˈta.ʃa.da	Minimal Pair	n.a.	اِعْتَقَدَ	ʔiʕ.ˈta.qa.da	**- (2.6)**	**+ (5.3)**	**+ (5.0)**	**+ (2.5)**
8	4	CVC.ˈCV.CV.CVC	اِفتَقَثَت	ʔif.ˈta.qa.θat	Minimal Pair	n.a.	اِفتَقَرَت	ʔif.ˈta.qa.rat	**+ (4.7)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**
8	4	CVC.CV.ˈCVV.CV	اِستَضاسَ	ʔis.ta.ˈdˤaː.sa	Minimal Pair	n.a.	اِستَضافَ	ʔis.ta.ˈdˤaː.fa	**- (3.8)**	**- (5.0)**	**+ (5.5)**	**+ (3.3)**

(): Log Frequency

Appendix 3.9 Samples of Recruitment Flyer in English and Chinese

Speak Arabic?

then, WE WANT YOU!!

We are conducting a study that investigates how students learn Arabic. We would like you to join us if you are:

- **NOT** a heritage speaker
- currently studying Arabic in 1st, 2nd and 3rd year classes
- have not lived in an Arabic speaking county for more than one semester

This study includes an experiment and a survey. The experiment is non-intrusive and is completely anonymous. The experiment takes about 40 minutes where you will hear and read some Arabic words . A **$30 gift card** will be given to completed participation. Limited spots available.

Please contact:
Cheng-Wei Lin arstresspp@gmail.com

會說阿拉伯文?

快來參與我們的實驗!

我們是美國密西根大學的博士生，目前正進行一項關於阿拉伯語習得的研究。如果您滿足下述的條件，我們竭誠歡迎您參與這項研究:

- 中文為您的母語
- 現為阿語系一、二、三年級學生
- 未曾在阿拉伯語系國家學習語言超過一學期

本研究包括了語言實驗以及問卷，實驗為非侵入性且完全匿名。實驗中您將會聽以及說一些阿拉伯文。全程約莫四十分鐘，完整參與實驗將獲得**五百元禮卷**作為酬謝。

名額有限，報名請聯絡:
林同學 arstresspp@gmail.com

Appendix 4.1 Summary Tables for Multiple Comparisons for Stress Identification

Table A.4.1.1 Summary Table for Response Accuracy (Stress Identification Task)

Main Effect

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value
Native Language	F(2,86) =9.63	p < 0.001***	F(2,86) =6.78	p < 0.001***	F(2,86) =2.15	p = 0.12	F(2,86) =5.59	p < 0.01**
Proficiency	F(2,77) =1.7	p = 0.19	F(2,77) =2.01	p = 0.14	F(2,77) =1.61	p = 0.21	F(2,77) =1.27	p = 0.29
Native Language x Proficiency	F(2,74) =0.52	p = 0.6	F(2,74) =0.16	p = 0.86	F(2,74) =0.53	p = 0.59	F(2,74) =0.3	p = 0.74

Multiple Comparisons – Native Language

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	Estimate	P Value	Estimate	P Value	Estimate	P Value	Estimate	P Value
EN-AR	0.43	p < 0.001***	0.39	p < 0.001***	n.a.	n.a.	0.27	p < 0.01**
TW-AR	0.38	p < 0.001***	0.31	p < 0.01**	n.a.	n.a.	0.16	p = 0.17
TW-EN	-0.06	p = 0.63	-0.09	p = 0.39	n.a.	n.a.	-0.11	p = 0.09

Main Effect

	Quadrisyllabic							
	STRESS PATTERN 5		STRESS PATTERN 6		STRESS PATTERN 7		STRESS PATTERN 8	
	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value
Native Language	F(2,86) =14.24	p < 0.001***	F(2,86) =7.04	p < 0.001***	F(2,86) =5.66	p < 0.01**	F(2,86) =5.76	p < 0.01**
Proficiency	F(2,77) =0.12	p = 0.89	F(2,77) =1.16	p = 0.32	F(2,77) =0.48	p = 0.62	F(2,77) =1.33	p = 0.27
Native Language x Proficiency	F(2,74) =0.04	p = 0.96	F(2,74) =1.36	p = 0.26	F(2,74) =0.65	p = 0.52	F(2,74) =0.26	p = 0.77

Multiple Comparisons – Native Language

	Quadrisyllabic							
	STRESS PATTERN 5		STRESS PATTERN 6		STRESS PATTERN 7		STRESS PATTERN 8	
	Estimate	P Value	Estimate	P Value	Estimate	P Value	Estimate	P Value
EN-AR	0.29	p < 0.001***	0.47	p < 0.001***	0.45	p < 0.001***	0.25	p < 0.001***
TW-AR	0.32	p < 0.001***	0.48	p < 0.001***	0.4	p < 0.01**	0.27	p < 0.001***
TW-EN	0.03	p = 0.72	0.01	p = 1	-0.06	p = 0.76	0.02	p = 0.87

Table A.4.1.2 Summary Table for Reaction Time (Stress Identification Task)

Main Effect

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value
Native Language	F(2,87) =2.34	p = 0.1	F(2,87) =7.17	p < 0.001***	F(2,87) =12.32	p < 0.001***	F(2,87) =5.25	p < 0.01**
Proficiency	F(2,78) =1.41	p = 0.25	F(2,78) =0.38	p = 0.69	F(2,78) =0.61	p = 0.55	F(2,78) =0.92	p = 0.4
Native Language x Proficiency	F(2,75) =0.28	p = 0.75	F(2,75) =1.22	p = 0.3	F(2,75) =3.34	p < 0.05*	F(2,75) =3.21	p < 0.05*

Multiple Comparisons – Native Language

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	Estimate	P_Value	Estimate	P_Value	Estimate	P_Value	Estimate	P_Value
EN-AR	n.a.	n.a.	-2030.62	p < 0.001***	-1671.28	p < 0.001***	-1807.29	p < 0.01**
TW-AR	n.a.	n.a.	-1485.05	p < 0.01**	-1898.82	p < 0.001***	820.32	p = 0.11
TW-EN	n.a.	n.a.	545.57	p = 0.22	-227.54	p = 0.58	628.05	p = 0.18

Multiple Comparisons – Proficiency (English-Speaking Groups)

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	Estimate	P_Value	Estimate	P_Value	Estimate	P_Value	Estimate	P_Value
1yr-2yr	n.a.	n.a.	n.a.	n.a.	494.34	p = 0.3	488.42	p = 0.54

Multiple Comparisons – Proficiency (Chinese-Speaking Groups)

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	Estimate	P_Value	Estimate	P_Value	Estimate	P_Value	Estimate	P_Value
1yr-2yr	n.a.	n.a.	n.a.	n.a.	-591.23	p < 0.05*	-1232.62	p < 0.05*

Main Effect

	Quadrisyllabic							
	STRESS PATTERN 5		STRESS PATTERN 6		STRESS PATTERN 7		STRESS PATTERN 8	
	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value
Native Language	$F(2,87) = 15.49$	$p < 0.001$***	$F(2,87) = 2.29$	$p = 0.11$	$F(2,87) = 1.88$	$p = 0.16$	$F(2,87) = 4.33$	$p < 0.01$**
Proficiency	$F(2,78) = 0.28$	$p = 0.76$	$F(2,78) = 0.16$	$p = 0.86$	$F(2,78) = 1.6$	$p = 0.21$	$F(2,78) = 1.28$	$p = 0.28$
Native Language x Proficiency	$F(2,75) = 1.89$	$p = 0.16$	$F(2,75) = 0.7$	$p = 0.5$	$F(2,75) = 0.71$	$p = 0.5$	$F(2,75) = 2.35$	$p = 0.1$

Multiple Comparisons – Native Language

	Quadrisyllabic							
	STRESS PATTERN 5		STRESS PATTERN 6		STRESS PATTERN 7		STRESS PATTERN 8	
	Estimate	P_Value	Estimate	P_Value	Estimate	P_Value	Estimate	P_Value
EN-AR	-2250.79	$p < 0.001$***	n.a.	n.a.	n.a.	n.a.	-1573.27	$p < 0.01$**
TW-AR	-2795.15	$p < 0.001$***	n.a.	n.a.	n.a.	n.a.	-1485.71	$p < 0.01$**
TW-EN	-544.36	$p = 0.16$	n.a.	n.a.	n.a.	n.a.	87.55	$p = 0.96$

Appendix 4.2 Summary Tables of Frequency Effect for Stress Identification Task

Table A.4.2.1. Summary for Response Accuracy (Stress Identification)

		Stimulus Group 1		Stimulus Group 2		Stimulus Group 3		Stimulus Group 4		Stimulus Group 5	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AR	Frequent	0.39	0.49	0.61	0.49	0.1	0.3	0.39	0.49	0.38	0.49
	Infrequent	0.58	0.5	0.65	0.48	0.12	0.33	0.38	0.49	0.58	0.5
EN 1st	Frequent	0.78	0.42	0.84	0.36	0.55	0.5	0.78	0.42	0.71	0.45
	Infrequent	0.81	0.39	0.94	0.24	0.58	0.49	0.71	0.45	0.81	0.39
EN 2nd	Frequent	0.83	0.38	0.88	0.32	0.52	0.5	0.83	0.38	0.82	0.39
	Infrequent	0.88	0.32	0.94	0.23	0.55	0.5	0.82	0.39	0.88	0.32
TW 1st	Frequent	0.73	0.44	0.84	0.37	0.42	0.5	0.73	0.44	0.65	0.48
	Infrequent	0.75	0.43	0.96	0.2	0.46	0.5	0.65	0.48	0.75	0.43
TW 2nd	Frequent	0.86	0.35	0.94	0.23	0.72	0.45	0.86	0.35	0.77	0.42
	Infrequent	0.74	0.44	0.98	0.14	0.62	0.49	0.77	0.42	0.74	0.44

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio
AR	-0.95	p < 0.01**	0.39	0.11	p = 0.81	1.12	-0.96	p < 0.01**	0.39	-0.33	p = 0.47	0.72	-0.29	p = 0.59	0.76
EN 1st	-0.22	p = 0.46	0.81	0.53	p = 0.07	1.69	-0.66	p < 0.01**	0.52	-1.3	p < 0.01**	0.28	-0.19	p = 0.49	0.83
EN 2nd	-0.58	p = 0.19	0.57	0.11	p = 0.82	1.12	-0.72	p = 0.12	0.5	-1.04	p = 0.1	0.36	-0.2	p = 0.61	0.83
TW 1st	-0.11	p = 0.68	0.9	0.54	p < 0.05*	1.72	-0.56	p < 0.05*	0.58	-1.91	p < 0.001***	0.15	0.22	p = 0.49	1.25
TW 2nd	0.93	p < 0.05*	2.52	0.87	p < 0.05*	2.38	0.21	p = 0.58	1.23	-1.16	p = 0.17	0.32	0.87	p < 0.05*	2.39

Table A.4.2.2. Summary for Reaction Time (Stress Identification)

		Stimulus Group 1		Stimulus Group 2		Stimulus Group 3		Stimulus Group 4		Stimulus Group 5	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AR	Frequent	3955	3609	4082	3629	4417	4407	3955	3609	4666	3625
	Infrequent	4579	3472	4309	3977	4312	3556	4666	3625	4579	3472
EN 1st	Frequent	2681	1985	2762	2255	3179	2622	2681	1985	2898	1958
	Infrequent	2980	2034	2353	2151	3146	2267	2898	1958	2980	2034
EN 2nd	Frequent	2048	1253	2366	1505	2965	2010	2048	1253	2383	1512
	Infrequent	2489	1689	1869	1545	3159	2406	2383	1512	2489	1689
TW 1st	Frequent	3573	10242	2334	1928	2913	2444	3573	10242	2978	2240
	Infrequent	2966	1804	1339	1440	3221	2956	2978	2240	2966	1804
TW 2nd	Frequent	2741	2044	3172	1970	3274	1782	2741	2044	3489	2446
	Infrequent	4198	3249	1881	1423	4345	6016	3489	2446	4198	3249

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value
AR	-625	$F(1,134) = 1.8$	$p = 0.18$	-712	$F(1,134) = 2.07$	$p = 0.15$	88	$F(1,134) = 0.04$	$p = 0.84$	-228	$F(1,134) = 0.25$	$p = 0.62$	106	$F(1,134) = 0.06$	$p = 0.82$
EN 1st	-297	$F(1,313) = 2.63$	$p = 0.11$	-215	$F(1,312) = 1.58$	$p = 0.21$	-85	$F(1,311) = 0.24$	$p = 0.63$	409	$F(1,312) = 4.45$	$p < 0.05$*	41	$F(1,312) = 0.05$	$p = 0.84$
EN 2nd	-444	$F(1,193) = 7.45$	$p < 0.01$**	-336	$F(1,192) = 5.38$	$p < 0.05$*	-111	$F(1,193) = 0.42$	$p = 0.52$	497	$F(1,194) = 9.26$	$p < 0.01$**	-180	$F(1,191) = 0.78$	$p = 0.38$
TW 1st	608	$F(1,314) = 0.6$	$p = 0.44$	596	$F(1,314) = 0.57$	$p = 0.45$	13	$F(1,314) = 0.01$	$p = 0.95$	995	$F(1,314) = 33.97$	$p < 0.001$***	-308	$F(1,314) = 1.6$	$p = 0.21$
TW 2nd	-1458	$F(1,194) = 22.19$	$p < 0.001$***	-749	$F(1,194) = 8.19$	$p < 0.01$**	-710	$F(1,194) = 4.43$	$p < 0.05$*	1291	$F(1,194) = 38.51$	$p < 0.001$***	-1072	$F(1,194) = 3.26$	$p = 0.07$

Appendix 4.3 Summary Tables for Multiple Comparisons for Lexical Decision

Table A.4.3.1 Summary Table for Response Accuracy (Lexical Decision Task)

Main Effect

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value
Native Language	F(2,86) =17.14	p < 0.001***	F(2,86) =31.65	p < 0.001***	F(2,86) =16.48	p < 0.001***	F(2,86) =13.23	p < 0.001***
Proficiency	F(2,77) =0.76	p = 0.47	F(2,77) =0.41	p = 0.67	F(2,77) =1.77	p = 0.18	F(2,77) =0.07	p = 0.93
Native Language x Proficiency	F(2,74) =1.89	p = 0.16	F(2,74) =2.12	p = 0.13	F(2,74) =0.94	p = 0.39	F(2,74) =0.28	p = 0.75

Multiple Comparisons – Native Language

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	Estimate	P Value	Estimate	P Value	Estimate	P Value	Estimate	P Value
EN-AR	-0.52	p < 0.001***	-0.61	p < 0.001***	-0.45	p < 0.001***	-0.48	p < 0.001***
TW-AR	-0.52	p < 0.001***	-0.66	p < 0.001***	-0.55	p < 0.001***	-0.5	p < 0.001***
TW-EN	-0.01	p = 0.99	-0.05	p = 0.58	-0.1	p = 0.2	-0.02	p = 0.95

Main Effect

	Quadrisyllabic							
	STRESS PATTERN 5		STRESS PATTERN 6		STRESS PATTERN 7		STRESS PATTERN 8	
	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value
Native Language	F(2,86) =18.46	p < 0.001***	F(2,86) =15.57	p < 0.001***	F(2,86) =17.6	p < 0.001***	F(2,86) =19.71	p < 0.001***
Proficiency	F(2,77) =0.79	p = 0.46	F(2,77) =2.15	p = 0.12	F(2,77) =2.91	p = 0.06	F(2,77) =1.88	p = 0.16
Native Language x Proficiency	F(2,74) =0.4	p = 0.67	F(2,74) =0.16	p = 0.85	F(2,74) =0.86	p = 0.43	F(2,74) =0.7	p = 0.5

Multiple Comparisons – Native Language

	Quadrisyllabic							
	STRESS PATTERN 5		STRESS PATTERN 6		STRESS PATTERN 7		STRESS PATTERN 8	
	Estimate	P Value	Estimate	P Value	Estimate	P Value	Estimate	P Value
EN-AR	-0.53	p < 0.001***	-0.53	p < 0.001***	-0.59	p < 0.001***	-0.51	p < 0.001***
TW-AR	-0.6	p < 0.001***	-0.59	p < 0.001***	-0.55	p < 0.001***	-0.59	p < 0.001***
TW-EN	-0.08	p = 0.4	-0.06	p = 0.58	0.04	p = 0.78	-0.08	p = 0.33

Table A.4.3.2 Summary Table for Reaction Time (Lexical Decision Task)

Main Effect

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value
Native Language	F(2,87) =0.2	p = 0.82	F(2,87) =0.23	p = 0.79	F(2,87) =0.55	p = 0.58	F(2,87) =0.82	p = 0.44
Proficiency	F(2,78) =0.09	p = 0.92	F(2,78) =0.23	p = 0.79	F(2,78) =0.19	p = 0.82	F(2,78) =1.11	p = 0.34
Native Language x Proficiency	F(2,75) =0.14	p = 0.87	F(2,75) =0.01	p = 0.99	F(2,75) =0.09	p = 0.92	F(2,75) =0.31	p = 0.73

Main Effect

	Quadrisyllabic							
	STRESS PATTERN 5		STRESS PATTERN 6		STRESS PATTERN 7		STRESS PATTERN 8	
	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value
Native Language	F(2,87) =2.21	p = 0.12	F(2,87) =2.17	p = 0.12	F(2,87) =0.24	p = 0.79	F(2,87) =2.18	p = 0.12
Proficiency	F(2,78) =0.55	p = 0.58	F(2,78) =0.14	p = 0.87	F(2,78) =0.83	p = 0.44	F(2,78) =0.32	p = 0.73
Native Language x Proficiency	F(2,75) =0.21	p = 0.81	F(2,75) =0.97	p = 0.38	F(2,75) =0.13	p = 0.88	F(2,75) =0.17	p = 0.85

Appendix 4.4 Summary Tables of Frequency Effect for Lexical Decision Task

Table A.4.4.1. Summary for Response Accuracy (Lexical Decision)

		Stimulus Group 1		Stimulus Group 2		Stimulus Group 3		Stimulus Group 4		Stimulus Group 5	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AR	Frequent	0.78	0.42	0.86	0.35	0.89	0.32	0.78	0.42	0.93	0.26
	Infrequent	0.92	0.28	0.89	0.32	0.93	0.26	0.93	0.26	0.92	0.28
EN 1st	Frequent	0.26	0.44	0.32	0.47	0.4	0.49	0.26	0.44	0.28	0.45
	Infrequent	0.44	0.5	0.35	0.48	0.38	0.49	0.28	0.45	0.44	0.5
EN 2nd	Frequent	0.27	0.45	0.4	0.49	0.42	0.5	0.27	0.45	0.42	0.5
	Infrequent	0.42	0.5	0.37	0.48	0.38	0.49	0.42	0.5	0.42	0.5
TW 1st	Frequent	0.31	0.46	0.2	0.4	0.31	0.46	0.31	0.46	0.3	0.46
	Infrequent	0.35	0.48	0.2	0.4	0.45	0.5	0.3	0.46	0.35	0.48
TW 2nd	Frequent	0.17	0.38	0.35	0.48	0.27	0.45	0.17	0.38	0.24	0.43
	Infrequent	0.5	0.5	0.41	0.49	0.31	0.46	0.24	0.43	0.5	0.5

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio
AR	-1.35	$p < 0.01$**	0.26	-1.49	$p < 0.01$**	0.23	0.2	$p = 0.75$	1.22	-0.26	$p = 0.61$	0.78	-0.62	$p = 0.35$	0.55
EN 1st	-1.08	$p < 0.001$***	0.35	-0.16	$p = 0.57$	0.86	-0.86	$p < 0.001$***	0.43	-0.22	$p = 0.42$	0.81	0.1	$p = 0.7$	1.11
EN 2nd	-0.81	$p < 0.01$**	0.45	-0.79	$p < 0.01$**	0.46	-0.06	$p = 0.86$	0.95	0.24	$p = 0.49$	1.26	0.25	$p = 0.47$	1.29
TW 1st	-0.24	$p = 0.36$	0.8	0.08	$p = 0.79$	1.08	-0.29	$p = 0.25$	0.76	-0.05	$p < 0.001$***	0.96	-0.97	$p < 0.001$***	0.38
TW 2nd	-1.72	$p < 0.001$***	0.19	-0.43	$p = 0.22$	0.66	-1.29	$p < 0.001$***	0.28	-0.32	$p = 0.3$	0.74	-0.22	$p = 0.51$	0.81

Table A.4.4.2. Summary for Reaction Time (Lexical Decision)

		Stimulus Group 1		Stimulus Group 2		Stimulus Group 3		Stimulus Group 4		Stimulus Group 5	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AR	Frequent	892	934	1111	1802	994	1235	892	934	958	1155
	Infrequent	1014	1211	1476	2486	838	1200	958	1155	1014	1211
EN 1st	Frequent	779	613	1085	1244	1013	883	779	613	912	902
	Infrequent	857	866	1111	1181	1027	1059	912	902	857	866
EN 2nd	Frequent	832	876	1286	1459	1142	1432	832	876	871	1010
	Infrequent	1004	1018	910	793	1078	1008	871	1010	1004	1018
TW 1st	Frequent	786	953	815	797	754	644	786	953	1038	4111
	Infrequent	718	713	913	1424	920	960	1038	4111	718	713
TW 2nd	Frequent	818	949	859	767	713	718	818	949	937	2089
	Infrequent	911	1016	866	804	1093	1434	937	2089	911	1016

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value
AR	-123	F(1,134) = 0.18	p = 0.67	-66	F(1,134) = 0.1	p = 0.75	-57	F(1,134) = 0.57	p = 0.46	-365	F(1,134) = 1.19	p = 0.28	156	F(1,134) = 0.65	p = 0.42
EN 1st	-81	F(1,314) = 3.55	p = 0.06	-138	F(1,314) = 0.52	p = 0.47	63	F(1,314) = 1.14	p = 0.29	-27	F(1,314) = 0.09	p = 0.77	-20	F(1,314) = 0.04	p = 0.86
EN 2nd	-168	F(1,194) = 0.09	p = 0.77	-39	F(1,194) = 0.98	p = 0.32	-128	F(1,194) = 1.82	p = 0.18	376	F(1,194) = 8.09	p < 0.01**	69	F(1,194) = 0.2	p = 0.66
TW 1st	69	F(1,314) = 0.6	p = 0.44	-253	F(1,314) = 0.99	p = 0.32	321	F(1,314) = 0.65	p = 0.42	-99	F(1,314) = 0.81	p = 0.37	-167	F(1,314) = 5.71	p < 0.01**
TW 2nd	-93	F(1,194) = 0.3	p = 0.59	-119	F(1,194) = 0.02	p = 0.91	27	F(1,194) = 0.54	p = 0.46	-8	F(1,194) = 0.01	p = 0.94	-380	F(1,194) = 7.61	p < 0.01**

Appendix 4.5 Summary Tables for Multiple Comparisons for Stress Production (Trial 1)

Table A.4.5.1 Summary Table for Response Accuracy (Stress Production Task)

Main Effect

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value
Native Language	F(2,83) =4.63	p < 0.01**	F(2,86) =1.77	p = 0.18	F(2,85) =18.98	p < 0.001***	F(2,86) =13.41	p < 0.001***
Proficiency	F(2,74) =3.81	p < 0.07	F(2,77) =0.99	p = 0.38	F(2,76) =0.75	p = 0.48	F(2,77) =5.51	p < 0.08
Native Language x Proficiency	F(2,71) =1.15	p = 0.32	F(2,74) =1.24	p = 0.29	F(2,73) =0.48	p = 0.62	F(2,74) =1.92	p = 0.15

Multiple Comparisons – Native Language

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	Estimate	P Value	Estimate	P Value	Estimate	P Value	Estimate	P Value
EN-AR	-0.17	p = 0.29	n.a.	n.a.	-0.25	p < 0.01**	-0.4	p < 0.001***
TW-AR	0.04	p = 0.94	n.a.	n.a.	0.11	p = 0.48	-0.14	p = 0.38
TW-EN	0.21	p < 0.01**	n.a.	n.a.	0.36	p < 0.001***	0.26	p < 0.001***

Main Effect

	Quadrisyllabic							
	STRESS PATTERN 5		STRESS PATTERN 6		STRESS PATTERN 7		STRESS PATTERN 8	
	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value
Native Language	F(2,86) =29.31	p < 0.001***	F(2,84) =28.62	p < 0.001***	F(2,86) =27.59	p < 0.001***	F(2,83) =8.01	p < 0.001***
Proficiency	F(2,77) =0.45	p = 0.64	F(2,75) =0.74	p = 0.48	F(2,77) =0.44	p = 0.64	F(2,74) =1.22	p = 0.3
Native Language x Proficiency	F(2,74) =0.48	p = 0.62	F(2,72) =2.81	p = 0.07	F(2,74) =3.72	p < 0.08	F(2,71) =0.13	p = 0.88

Multiple Comparisons – Native Language

	Quadrisyllabic							
	STRESS PATTERN 5		STRESS PATTERN 6		STRESS PATTERN 7		STRESS PATTERN 8	
	Estimate	P Value	Estimate	P Value	Estimate	P Value	Estimate	P Value
EN-AR	-0.31	p < 0.001***	-0.6	p < 0.001***	-0.61	p < 0.001***	-0.26	p < 0.05*
TW-AR	0.15	p = 0.28	-0.23	p < 0.01**	-0.36	p < 0.001***	0.03	p = 0.96
TW-EN	0.45	p < 0.001***	0.36	p < 0.001***	0.25	p < 0.001***	0.29	p < 0.001***

Table A.4.5.2 Summary Table for Response Accuracy (Stress Production Task)

Main Effect

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value
Native Language	F(2,86) =8.98	p < 0.001***	F(2,86) =4.5	p < 0.01**	F(2,86) =6.64	p < 0.01**	F(2,86) =5.37	p < 0.01**
Proficiency	F(2,77) =0.36	p = 0.7	F(2,77) =0.53	p = 0.59	F(2,77) =1.03	p = 0.36	F(2,77) =0.88	p = 0.42
Native Language x Proficiency	F(2,74) =0.09	p = 0.91	F(2,74) =0.27	p = 0.76	F(2,74) =0.79	p = 0.46	F(2,74) =1.92	p = 0.15

Multiple Comparisons – Native Language

	Trisyllabic							
	STRESS PATTERN 1		STRESS PATTERN 2		STRESS PATTERN 3		STRESS PATTERN 4	
	Estimate	P_Value	Estimate	P_Value	Estimate	P_Value	Estimate	P_Value
EN-AR	341.87	p < 0.001***	257.19	p < 0.01**	313.7	p < 0.001***	220.27	p < 0.05*
TW-AR	257.88	p < 0.01**	189.48	p = 0.08	249.72	p < 0.01**	68.71	p = 0.73
TW-EN	-83.98	p = 0.21	-67.71	p = 0.4	-63.99	p = 0.44	-151.56	p < 0.01**

Main Effect

	Quadrisyllabic							
	STRESS PATTERN 5		STRESS PATTERN 6		STRESS PATTERN 7		STRESS PATTERN 8	
	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value	F_statistics	P_Value
Native Language	F(2,86) =11.96	p < 0.001***	F(2,86) =5.62	p < 0.01**	F(2,86) =2.7	p = 0.07	F(2,86) =11.19	p < 0.001***
Proficiency	F(2,77) =0.43	p = 0.65	F(2,77) =0.94	p = 0.4	F(2,77) =0.45	p = 0.64	F(2,77) =0.53	p = 0.59
Native Language x Proficiency	F(2,74) =0.25	p = 0.78	F(2,74) =0.93	p = 0.4	F(2,74) =0.32	p = 0.73	F(2,74) =0.57	p = 0.57

Multiple Comparisons – Native Language

	Quadrisyllabic							
	STRESS PATTERN 5		STRESS PATTERN 6		STRESS PATTERN 7		STRESS PATTERN 8	
	Estimate	P_Value	Estimate	P_Value	n.a.	n.a.	Estimate	P_Value
EN-AR	377.86	p < 0.001***	281.12	p < 0.001***	n.a.	n.a.	333.05	p < 0.001***
TW-AR	211.63	p < 0.05*	207.93	p < 0.05*	n.a.	n.a.	161.43	p = 0.12
TW-EN	-166.23	p < 0.001***	-73.19	p = 0.33	n.a.	n.a.	-171.62	p < 0.001***

Appendix 4.6 Summary Tables of Frequency Effect for Stress Production Task (Trial1)

Table A.4.6.1. Summary for Response Accuracy (Stress Production)

		Stimulus Group 1		Stimulus Group 2		Stimulus Group 3		Stimulus Group 4		Stimulus Group 5	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AR	Frequent	0.69	0.47	0.7	0.46	0.6	0.49	0.69	0.47	0.69	0.47
	Infrequent	0.64	0.48	0.75	0.43	0.65	0.48	0.69	0.47	0.64	0.48
EN 1st	Frequent	0.56	0.5	0.11	0.32	0.27	0.45	0.56	0.5	0.52	0.5
	Infrequent	0.3	0.46	0.18	0.39	0.34	0.48	0.52	0.5	0.3	0.46
EN 2nd	Frequent	0.58	0.5	0.11	0.31	0.4	0.49	0.58	0.5	0.52	0.5
	Infrequent	0.27	0.45	0.14	0.35	0.34	0.48	0.52	0.5	0.27	0.45
TW 1st	Frequent	0.63	0.49	0.4	0.49	0.59	0.49	0.63	0.49	0.48	0.5
	Infrequent	0.44	0.5	0.35	0.48	0.76	0.43	0.48	0.5	0.44	0.5
TW 2nd	Frequent	0.88	0.33	0.51	0.5	0.69	0.47	0.88	0.33	0.71	0.46
	Infrequent	0.64	0.48	0.45	0.5	0.87	0.34	0.71	0.46	0.64	0.48

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio
AR	0.31	p = 0.48	1.36	0.05	p = 0.91	1.05	0.2	p = 0.65	1.22	-0.33	p = 0.48	0.73	-0.39	p = 0.41	0.69
EN 1st	0.63	p < 0.05*	1.87	0.46	p = 0.13	1.58	0.46	p = 0.13	1.58	-0.45	p = 0.23	0.65	-0.11	p = 0.73	0.91
EN 2nd	0.3	p = 0.39	1.34	0.36	p = 0.32	1.43	0.36	p = 0.32	1.43	-0.58	p = 0.24	0.57	0.09	p = 0.8	1.1
TW 1st	0.85	p < 0.01**	2.34	0.97	p < 0.01**	2.64	0.13	p = 0.63	1.14	0.22	p = 0.51	1.25	-0.97	p < 0.01**	0.38
TW 2nd	1.44	p < 0.001***	4.2	1.08	p < 0.05*	2.93	0.34	p = 0.37	1.4	0.31	p = 0.47	1.36	-1.69	p < 0.01**	0.19

Table A.4.6.2. Summary for Reaction Time (Stress Production)

		Stimulus Group 1		Stimulus Group 2		Stimulus Group 3		Stimulus Group 4		Stimulus Group 5	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AR	Frequent	1395	369	1490	443	1476	506	1395	369	1617	465
	Infrequent	1583	528	1625	532	1425	423	1617	465	1583	528
EN 1st	Frequent	1762	320	1806	329	1821	338	1762	320	1863	352
	Infrequent	1733	288	1838	310	1812	337	1863	352	1733	288
EN 2nd	Frequent	1741	343	1773	326	1883	375	1741	343	1926	401
	Infrequent	1902	373	1837	330	1870	371	1926	401	1902	373
TW 1st	Frequent	1676	342	1706	397	1666	386	1676	342	1807	429
	Infrequent	1668	406	1778	402	1653	316	1807	429	1668	406
TW 2nd	Frequent	1630	314	1732	347	1613	268	1630	314	1833	369
	Infrequent	1653	325	1732	324	1634	288	1833	369	1653	325

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value
AR	-188	F(1,134) = 8.98	p < 0.01**	-223	F(1,134) = 13.72	p < 0.001***	35	F(1,134) = 0.23	p = 0.63	-136	F(1,134) = 3.19	p = 0.08	52	F(1,134) = 0.59	p = 0.45
EN 1st	29	F(1,297) = 0.94	p = 0.33	-102	F(1,297) = 8.71	p < 0.01**	131	F(1,297) = 17.04	p < 0.001***	-32	F(1,296) = 1.01	p = 0.32	14	F(1,284) = 0.19	p = 0.67
EN 2nd	-162	F(1,194) = 16.21	p < 0.001***	-188	F(1,193) = 20.73	p < 0.001***	25	F(1,193) = 0.33	p = 0.57	-64	F(1,194) = 4.19	p < 0.05*	1	F(1,186) = 0	p = 1
TW 1st	11	F(1,312) = 0.1	p = 0.76	-128	F(1,311) = 12.25	p < 0.001***	139	F(1,313) = 11.1	p < 0.001***	-72	F(1,314) = 3.58	p < 0.05*	16	F(1,313) = 0.21	p = 0.65
TW 2nd	-24	F(1,194) = 0.39	p = 0.53	-203	F(1,194) = 29.52	p < 0.001***	180	F(1,194) = 19.52	p < 0.001***	-1	F(1,194) = 0.01	p = 0.99	-22	F(1,194) = 0.47	p = 0.5

Appendix 4.7 Comparisons between without and with Graduate Students

Table A.4.7.1. Reaction Time Comparisons with and without Graduate Students

Task One: Stress Identification Task

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value
EN 2nd (w/o Grads)	-444	F(1,193) = 7.45	p < 0.01**	-336	F(1,192) = 5.38	p < 0.05*	-111	F(1,193) = 0.42	p = 0.52	497	F(1,194) = 9.26	p < 0.01**	-180	F(1,191) = 0.78	p = 0.38
EN 2nd (w/ Grads)	-455	F(1,118) = 3.54	p = 0.06	-442	F(1,117) = 4.27	p < 0.05*	-17	F(1,118) = 0.01	p = 0.95	666	F(1,119) = 9.99	p < 0.001***	-289	F(1,119) = 0.98	p = 0.32

Task Two: Lexical Decision Task

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value
EN 2nd (w/o Grads)	-168	F(1,194) = 0.09	p = 0.77	-39	F(1,194) = 0.98	p = 0.32	-128	F(1,194) = 1.82	p = 0.18	376	F(1,194) = 8.09	p < 0.01**	69	F(1,194) = 0.2	p = 0.66
EN 2nd (w/ Grads)	-72	F(1,117) = 0.4	p = 0.53	-113	F(1,118) = 0.08	p = 0.79	42	F(1,118) = 0.24	p = 0.63	415	F(1,119) = 4.49	p < 0.05*	197	F(1,119) = 0.95	p = 0.33

Task Three: Stress Production Task (Trial 1)

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value	Estimate	F Statistics	P value
EN 2nd (w/o Grads)	-162	F(1,194) = 16.21	p < 0.001***	-188	F(1,193) = 20.73	p < 0.001***	25	F(1,193) = 0.33	p = 0.57	-64	F(1,194) = 4.19	p < 0.05*	1	F(1,186) = 0	p = 1
EN 2nd (w/ Grads)	-33	F(1,119) = 0.64	p = 0.43	-140	F(1,118) = 8.68	p < 0.01**	106	F(1,118) = 5.56	p < 0.05*	-46	F(1,119) = 1.86	p = 0.18	-21	F(1,115) = 0.25	p = 0.62

Table A.4.7.2. Response Accuracy Comparisons with and without Graduate Students

Task One: Stress Identification Task

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio
EN 2nd (w/o Grads)	-0.62	p = 0.21	0.55	0.39	p = 0.43	1.48	-0.99	p < 0.05*	0.38	-1.04	p = 0.17	0.36	-0.55	p = 0.28	0.59
EN 2nd (w/ Grads)	-0.58	p = 0.19	0.57	0.11	p = 0.82	1.12	-0.72	p = 0.12	0.5	-1.04	p = 0.1	0.36	-0.2	p = 0.61	0.83

Task Two: Lexical Decision Task

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio
EN 2nd (w/o Grads)	-0.48	p = 0.23	0.63	-0.25	p = 0.54	0.79	-0.26	p = 0.54	0.78	0.43	p = 0.3	1.54	0.29	p = 0.51	1.33
EN 2nd (w/ Grads)	-0.81	p < 0.01**	0.45	-0.79	p < 0.01**	0.46	-0.06	p = 0.86	0.95	0.24	p = 0.49	1.26	0.25	p = 0.47	1.29

Task Three: Stress Production Task (Trial 1)

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio
EN 2nd (w/o Grads)	1.75	p < 0.001***	5.71	0.26	p = 0.5	1.3	1.47	p < 0.001***	4.33	1.49	p = 0.12	4.42	0.78	p = 0.12	2.17
EN 2nd (w/ Grads)	1.57	p < 0.001***	4.77	0.41	p = 0.22	1.5	1.24	p < 0.001***	3.44	-0.31	p = 0.56	0.74	0.44	p = 0.23	1.55

Appendix 4.8 Comparisons Between Types of Nonsense Words

Table A.4.8.1 Comparisons between Types of Nonsense Words (Stress Identification Task)

REACTION TIME

	Trisyllabic						Quadrisyllabic			
	Group 1		Group 2		Group 3		Group 4		Group 5	
	F Statistics	P value	F Statistics	P value	F Statistics	P value	F Statistics	P value	F Statistics	P value
AR	F(1,134) = 0.01	p = 0.97	F(1,134) = 0.09	p = 0.78	F(1,134) = 0.19	p = 0.67	F(1,134) = 1.13	p = 0.29	F(1,134) = 3.37	p = 0.07
EN 1st	F(1,313) = 0.02	p = 0.91	F(1,312) = 0.88	p = 0.35	F(1,311) = 0.04	p = 0.86	F(1,312) = 0.06	p = 0.81	F(1,312) = 2.25	p = 0.14
EN 2nd	F(1,193) = 0.81	p = 0.37	F(1,192) = 0.9	p = 0.35	F(1,193) = 0.64	p = 0.42	F(1,194) = 5.22	p < 0.05*	F(1,191) = 3.4	p = 0.07
TW 1st	F(1,314) = 1.11	p = 0.29	F(1,314) = 1.04	p = 0.31	F(1,314) = 0.07	p = 0.81	F(1,314) = 3.77	p < 0.05*	F(1,314) = 1.94	p = 0.17
TW 2nd	F(1,194) = 0.39	p = 0.54	F(1,194) = 0.25	p = 0.62	F(1,194) = 0.08	p = 0.79	F(1,194) = 0.52	p = 0.47	F(1,194) = 0.22	p = 0.64

RESPONSE ACCURACY

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio
AR	0.14	p = 0.72	1.15	0.78	p < 0.001***	2.18	0.07	p = 0.86	1.07	-0.11	p = 0.81	0.9	-0.29	p = 0.59	0.76
EN 1st	0.32	p = 0.28	1.37	0.56	p < 0.05*	1.74	0.4	p = 0.15	1.49	-0.17	p = 0.68	0.86	0.35	p = 0.21	1.42
EN 2nd	0.2	p = 0.65	1.22	0.36	p = 0.45	1.43	0.11	p = 0.81	1.12	-2.63	p < 0.01**	0.08	0.11	p = 0.79	1.11
TW 1st	0.33	p = 0.22	1.39	0.23	p = 0.4	1.26	-0.49	p < 0.05*	0.62	-0.5	p = 0.23	0.61	0.55	p = 0.09	1.73
TW 2nd	0.15	p = 0.7	1.17	0.29	p = 0.52	1.33	-0.07	p = 0.85	0.94	0	p = 1	1	0.39	p = 0.31	1.48

Table A.4.8.2 Comparisons Between Types of Nonsense Words (Lexical Decision Task)

REACTION TIME

	Trisyllabic						Quadrisyllabic			
	Group 1		Group 2		Group 3		Group 4		Group 5	
	F Statistics	P value	F Statistics	P value	F Statistics	P value	F Statistics	P value	F Statistics	P value
AR	F(1,134) = 4.58	p < 0.05*	F(1,134) = 0.01	p = 0.97	F(1,134) = 0.04	p = 0.86	F(1,134) = 0.14	p = 0.72	F(1,134) = 0.01	p = 0.93
EN 1st	F(1,314) = 0.08	p = 0.78	F(1,314) = 1.1	p = 0.3	F(1,314) = 3	p = 0.08	F(1,314) = 0.43	p = 0.51	F(1,314) = 5.31	p < 0.05*
EN 2nd	F(1,194) = 6.21	p < 0.01**	F(1,194) = 0.16	p = 0.69	F(1,194) = 0.65	p = 0.42	F(1,194) = 4.79	p < 0.05*	F(1,194) = 0.57	p = 0.45
TW 1st	F(1,314) = 1.09	p = 0.3	F(1,314) = 0.92	p = 0.34	F(1,314) = 0.61	p = 0.44	F(1,314) = 0.18	p = 0.68	F(1,314) = 0.61	p = 0.44
TW 2nd	F(1,194) = 0.28	p = 0.6	F(1,194) = 0.34	p = 0.57	F(1,194) = 0.07	p = 0.81	F(1,194) = 2.37	p = 0.13	F(1,194) = 0.02	p = 0.9

RESPONSE ACCURACY

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio
AR	0.51	p = 0.32	1.66	0.89	p = 0.09	2.44	-0.2	p = 0.75	0.83	0.53	p = 0.31	1.7	-0.21	p = 0.75	0.82
EN 1st	0.62	p < 0.05*	1.86	-0.24	p = 0.39	0.79	-0.04	p = 0.9	0.97	0.29	p = 0.28	1.34	-0.37	p = 0.15	0.7
EN 2nd	0.59	p = 0.07	1.79	-0.12	p = 0.74	0.9	0.18	p = 0.6	1.2	-0.71	p < 0.05*	0.5	-0.13	p = 0.72	0.89
TW 1st	0.04	p = 0.9	1.04	0.15	p = 0.59	1.16	0.29	p = 0.25	1.34	0.23	p < 0.001***	1.26	0	p = 1	1
TW 2nd	0.28	p = 0.36	1.32	-0.56	p = 0.12	0.58	-0.32	p = 0.29	0.73	-0.59	p < 0.05*	0.56	-0.55	p = 0.1	0.58

Table A.4.8.3 Comparisons Between Types of Nonsense Words (Stress Production Task)

REACTION TIME

	Group 1		Group 2		Group 3		Group 4		Group 5	
	F Statistics	P value	F Statistics	P value	F Statistics	P value	F Statistics	P value	F Statistics	P value
AR	F(1,134) = 6.55	p < 0.01**	F(1,134) = 0.14	p = 0.71	F(1,134) = 1.26	p = 0.26	F(1,134) = 3.33	p = 0.07	F(1,134) = 0.84	p = 0.36
EN 1st	F(1,297) = 2.81	p = 0.1	F(1,297) = 2.54	p = 0.11	F(1,297) = 1.92	p = 0.17	F(1,296) = 0.28	p = 0.6	F(1,284) = 0.03	p = 0.88
EN 2nd	F(1,194) = 0.02	p = 0.91	F(1,193) = 0.18	p = 0.68	F(1,193) = 1.83	p = 0.18	F(1,194) = 0.22	p = 0.64	F(1,186) = 1.61	p = 0.21
TW 1st	F(1,312) = 0.3	p = 0.59	F(1,311) = 0.01	p = 0.99	F(1,313) = 2.43	p = 0.12	F(1,314) = 0.2	p = 0.66	F(1,313) = 10.03	p < 0.001***
TW 2nd	F(1,194) = 0.43	p = 0.51	F(1,194) = 0.01	p = 0.96	F(1,194) = 1.7	p = 0.19	F(1,194) = 2.05	p = 0.15	F(1,194) = 1.7	p = 0.19

RESPONSE ACCURACY

	Trisyllabic									Quadrisyllabic					
	Group 1			Group 2			Group 3			Group 4			Group 5		
	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio	Estimate	P Value	Odds Ratio
AR	-0.14	p = 0.75	0.88	-0.57	p = 0.18	0.57	0.18	p = 0.68	1.2	0.41	p = 0.39	1.5	0.02	p = 0.98	1.02
EN 1st	-0.1	p = 0.71	0.91	-0.44	p = 0.11	0.65	-0.34	p = 0.19	0.72	-0.2	p = 0.62	0.83	-0.38	p = 0.2	0.69
EN 2nd	0.03	p = 0.94	1.03	0.05	p = 0.9	1.05	0.18	p = 0.59	1.19	-0.2	p = 0.71	0.83	0.09	p = 0.82	1.09
TW 1st	-0.4	p = 0.16	0.68	-0.4	p = 0.24	0.68	-0.24	p = 0.38	0.79	0.59	p = 0.09	1.8	-0.52	p = 0.08	0.6
TW 2nd	-0.42	p = 0.27	0.66	-0.97	p < 0.05*	0.39	-0.6	p = 0.11	0.55	0.03	p = 0.95	1.03	-0.4	p = 0.48	0.68

Appendix 4.9 Questionnaire Results

Table A.4.9.1 Degree of Agreement for Teaching Factors Questions:

	EN		TW	
	1st Year	2nd Year	1st Year	2nd Year
Question:	N=20	N=13	N=21	N=13
My Arabic teacher has introduced stress rules in Arabic	60%	73%	73%	71%
My Arabic teacher regularly emphasizes the importance of putting stress on the right syllable	71%	79%	80%	71%
My Arabic teacher always corrects me when I don't put the stress on the correct position of the word	80%	81%	85%	79%
We regularly practice producing stress in class.	57%	71%	64%	58%
The textbook that I use has drills that help me learn how to produce stress correctly	47%	46%	66%	54%
The textbook that I use has detailed explanation of stress rules in Arabic	40%	40%	51%	45%
Average	**59%**	**65%**	**70%**	**63%**

Table A.4.9.2. Degree of Agreement for Personal Evaluation Questions:

	EN		TW	
	1st Year	2nd Year	1st Year	2nd Year
Question:	N=20	N=13	N=21	N=13
I completely understand the rules of stress in Arabic	62%	62%	66%	66%
I always think about stress rules before I say anything	65%	64%	76%	73%
I can identify where the stressed syllable is without saying it out loud	65%	71%	65%	68%
I always produce stress correctly for shorter words	71%	72%	71%	78%
I always produce stress correctly for longer words	60%	66%	69%	71%
I always pay attention to where I put my stress in speaking	75%	75%	79%	80%
I always read things out loud in Arabic	74%	75%	73%	76%
I always know when people do not produce stress correctly	69%	75%	71%	72%
I have read about stress rules in Arabic outside of class	56%	54%	59%	54%
Average	**66%**	**68%**	**70%**	**71%**

Table A.4.9.3. Degree of Agreement for Extracurricular Engagement Questions:

	EN		TW	
	1st Year	2nd Year	1st Year	2nd Year
Question:	N=20	N=13	N=21	N=13
I have a lot of friends who are native speakers of Arabic	46%	56%	39%	44%
I have a lot of opportunities to use Arabic outside of class	51%	60%	43%	42%
I find myself speaking Arabic all the time outside of class	36%	46%	42%	40%
I find myself listening to Arabic all the time outside of class	45%	50%	43%	42%
I find myself reading Arabic articles all the time outside of class	46%	63%	37%	44%
I find myself writing in Arabic all the time outside of class	60%	50%	50%	50%
Average	**48%**	**54%**	**42%**	**44%**

Table A.4.9.4. Degree of Correctness for Quiz on Stress Rules Questions:

	AR	EN		TW	
	Native	1st Year	2nd Year	1st Year	2nd Year
Question:	N=9	N=20	N=13	N=21	N=13
Stress always falls on the long vowels	56%	20%	0%	22%	8%
Stress can occur anywhere in the word	38%	40%	50%	64%	54%
Stress in Arabic has no rule	88%	100%	100%	75%	92%
Stress does not always occur at the beginning of the word	78%	89%	69%	85%	92%
Stress always falls on the end of the word	78%	100%	100%	89%	100%
Stress always follow a specific type of vowel	50%	60%	67%	36%	50%
Stress always follow a specific type of consonant.	75%	69%	80%	67%	78%
Stress does not always falls on the same place whether the word is singular or plural	71%	94%	75%	94%	100%
The feminine ending “ة” has an effect on where to put the stress	67%	53%	56%	56%	40%
There is always more than one way to stress the same word.	44%	57%	89%	50%	0%
Average	**64%**	**68%**	**69%**	**64%**	**61%**

Appendix 4.10 Sample of Stress Rules Provided by the Participants in the Questionnaire

Ayn is usually stressed / Letters with shudda are always stressed
Essentially that equal stress is applied unless there's a shadda or long vowel. She hasn't explicitly taught this, necessarily, but we've learned it from her corrections in our speech.
Follow the vowels (long and short). Other than that no specific rules, just lots of practice and habit forming
Just a lot of reading practice with short vowels--no formal lessons
Long vowels are stressed more than short vowels
long vowels take stress, otherwise it's usually the second syllable
No explicit rule, but pronunciation is a focus
Pronouncing the shaddha, and she enunciates very clearly so we can hear the stress on each word.
Second syllable in a word starting with alif-siin-taa or alif-nuun
Shadda, long vowels must be stressed
Shadda, long vowels, hamza
Stress with the different vowels, and in particular trying to distinguish long vowels.
Stressing the beginning of the word,
The long vowels in a word is where the stress should be when a long vowel is present.
The stress of a word usually falls with the long vowel of the word...I think
We didn't cover any specific rules, but he emphasizes "the music of the Arabic language," and we frequently have to repeat long lists of similarly stressed words one after the other.
We talked about short vowels in relation to where you'd put a hamza. Kesra is stronger than dhamma is stronger than fatha

Appendix 4.11 Correlation between External Factor Scores and Experiment Results

Table A.4.11.1 Correlation between External Factor Scores (Native Arabic Group)

	STRESS IDENTIFICATION			
	REACTION TIME		RESPONSE ACCURACY	
	Pearson's r	P value	Pearson's r	P value
Explicit Knowledge of Stress	-0.2	p = 0.6	0.3	p < 0.01**

	LEXICAL DECISION			
	REACTION TIME		RESPONSE ACCURACY	
	Pearson's r	P value	Pearson's r	P value
Explicit Knowledge of Stress	0.09	p = 0.81	0.04	p = 0.7

	STRESS PRODUCTION (Trial1)			
	REACTION TIME		RESPONSE ACCURACY	
	Pearson's r	P value	Pearson's r	P value
Explicit Knowledge of Stress	0.01	p = 0.98	-0.13	p = 0.21

Table A.4.11.2 Correlation between External Factor Scores (English-Speaking Group)

	STRESS IDENTIFICATION			
	REACTION TIME		RESPONSE ACCURACY	
	Pearson's r	P value	Pearson's r	P value
Instruction and Textbook	-0.09	p = 0.58	0.04	p = 0.79
Self-Evaluation	-0.19	p = 0.23	0.09	p = 0.56
Extracurricular Engagement	-0.29	p < 0.05*	-0.15	p = 0.33
Explicit Knowledge of Stress	0.22	p = 0.15	0.31	p < 0.05*

	LEXICAL DECISION			
	REACTION TIME		RESPONSE ACCURACY	
	Pearson's r	P value	Pearson's r	P value
Instruction and Textbook	-0.2	p = 0.18	0.16	p = 0.31
Self-Evaluation	0	p = 0.97	-0.04	p = 0.8
Extracurricular Engagement	-0.15	p = 0.34	-0.3	p < 0.05*
Explicit Knowledge of Stress	0.07	p = 0.64	0.24	p = 0.12

	STRESS PRODUCTION (Trial1)			
	REACTION TIME		RESPONSE ACCURACY	
	Pearson's r	P value	Pearson's r	P value
Instruction and Textbook	-0.06	p = 0.69	-0.06	p = 0.68
Self-Evaluation	-0.05	p = 0.75	-0.08	p = 0.59
Extracurricular Engagement	-0.2	p = 0.19	0.12	p = 0.44
Explicit Knowledge of Stress	0.24	p = 0.12	0.01	p = 0.93

Table A.4.11.3 Correlation between External Factor Scores (Chinese-Speaking Group)

	STRESS IDENTIFICATION			
	REACTION TIME		RESPONSE ACCURACY	
	Pearson's r	P value	Pearson's r	P value
Instruction and Textbook	0.02	p = 0.9	0.15	p = 0.38
Self-Evaluation	-0.3	p = 0.07	0.38	p < 0.05*
Extracurricular Engagement	0.05	p = 0.76	-0.07	p = 0.69
Explicit Knowledge of Stress	0.03	p = 0.85	0.15	p = 0.39

	LEXICAL DECISION			
	REACTION TIME		RESPONSE ACCURACY	
	Pearson's r	P value	Pearson's r	P value
Instruction and Textbook	0.09	p = 0.6	-0.04	p = 0.84
Self-Evaluation	0.27	p = 0.11	0.17	p = 0.31
Extracurricular Engagement	-0.11	p = 0.54	-0.15	p = 0.38
Explicit Knowledge of Stress	0.05	p = 0.76	-0.14	p = 0.41

	STRESS PRODUCTION (Trial1)			
	REACTION TIME		RESPONSE ACCURACY	
	Pearson's r	P value	Pearson's r	P value
Instruction and Textbook	-0.28	p = 0.09	0.04	p = 0.84
Self-Evaluation	-0.31	p = 0.06	0.39	p < 0.01**
Extracurricular Engagement	-0.12	p = 0.5	-0.05	p = 0.79
Explicit Knowledge of Stress	-0.18	p = 0.29	0.01	p = 0.95

BIBLIOGRAPHY

Al-Ani, S. H. (1992). Stress variation of the construct phrase in Arabic: a spectrographic analysis. *Anthropological Linguistics*, 256–276.

Alfano, I., Llisterri, J., & Savy, R. (2007). The perception of Italian and Spanish lexical stress: A first cross-linguistic study. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1793–1796). Saarbrücken, Germany.

Alfano, I., Schwab, S., Savy, R., & Llisterri, J. (2010). Cross-language speech perception: Lexical stress in Spanish with Italian and Francophone subjects. In *AISV 2009. La dimensione temporale del parlato. Atti del 5o convegno nazionale AISV - Associazione Italiana di Scienze della Voce.* (pp. 455–474). Università di Zurigo: Torriana: EDK Editore.

Alhawary, M. T. (2011). *Modern standard Arabic grammar: A learner's guide*. John Wiley & Sons.

Alhawary, M. T. (2013). Arabic second language acquisition research and second language teaching: what the teacher, textbook writer, and tester need to know. *Al-'Arabiyya*, 23–35.

Almbark, R., Bouchhioua, N., & Hellmuth, S. (2014). Acquiring the Phonetics and Phonology of English Word Stress: Comparing Learners from Different L1 Backgrounds. In *Proceedings of the International Symposium on the Acquisition of Second Language Speech, Concordia Working Papers in Applied Linguistics (Vol. 5).* (Vol. 5, pp. 19–35). Retrieved from http://doe.concordia.ca/copal/documents/3_Almbark_etal_Vol5.pdf

Altmann, H. (2006). *The Perception and Production of Second Language Stress: A crosslinguistic experimental study* (Ph.D Dissertation). University of Delaware. Retrieved from http://www.linguistlist.org/pubs/diss/browse-diss-action.cfm?DissID=12720&RequestTimeout=500

Anani, M. (1989). Incorrect Stress Placement in the Case of Arab Learners of English. *IRAL*, *27*(1), 15–22.

Angoujard, J.-P. (1990). *Metrical structure of Arabic* (Vol. 35). Walter de Gruyter GmbH & Co KG.

Archibald, J. (1993). *Language Learnability and L2 Phonology: The Acquisition of Metrical Parameters*. Springer Science & Business Media.

Archibald, J. (1997). *The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress*. Retrieved from https://www.degruyter.com/view/j/ling.1997.35.issue-1/ling.1997.35.1.167/ling.1997.35.1.167.xml

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340.

Barry, C., & Seymour, P. H. (1988). Lexical priming and sound-to-spelling contingency effects in nonword spelling. *The Quarterly Journal of Experimental Psychology*, *40*(1), 5–40.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). MIT Press Cambridge, MA.

Bouchhioua, N. (2008). The Role of Vowel Quality in Cuing Stress and Accent in Tunisian Arabic, Native English, and L2 English. In *Speech Prosody 2008: Fourth Conference on Speech prosody* (pp. 539–542).

Bouchhioua, N. (n.d.). *The Role of Vowel Quality in Cuing Stress and Accent in Tunisian Arabic, Native English*.

Boudelaa, S., & Meftah, M. (1996). Cross-language effects of lexical stress in word recognition: the case of Arabic English bilinguals. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* (Vol. 1, pp. 121–124). IEEE.

Brooks, P. J., Braine, M. D., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language*, *32*(1), 76–95.

Chahal, D. (2001). *Modeling the intonation of Lebanese Arabic using the autosegmental-metrical framework: a comparison with English*. University of Melbourne.

Chen, Y., Robb, M. P., Gilbert, H. R., & Lerman, J. W. (2001). A study of sentence stress production in Mandarin speakers of American English. *The Journal of the Acoustical Society of America*, *109*(4), 1681–1690.

Clopper, C. (2002). Frequency of stress patterns in English: A computational analysis. *IULC Working Papers Online*, *2*(2).

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*(4), 589.

Cowell, M. W. (1964). *A reference grammar of Syrian Arabic: based on the dialect of Damascus*. Washington: Georgetown University Press. Retrieved from http://hdl.handle.net/2027/[u]:mdp.39015002674748

Cristófaro-Silva, T., & Fraga, T. (2005). ASPA: a formulação de um banco de dados de referência da estrutura sonora do português contemporâneo. In *Proceedings XXV Congress of Brazilian Society of Computing Science* (pp. 2268–2277).

de Jong, K., & Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration. *Journal of Phonetics*, *30*(1), 53–75. https://doi.org/10.1006/jpho.2001.0151

de Jong, K., & Zawaydeh, B. A. (1999). Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics*, *27*(1), 3–22. https://doi.org/10.1006/jpho.1998.0088

Domahs, U., Plag, I., & Carroll, R. (2014). Word stress assignment in German, English and Dutch: Quantity-sensitivity and extrametricality revisited. *The Journal of Comparative Germanic Linguistics*, *17*(1), 59–96. https://doi.org/10.1007/s10828-014-9063-9

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, *36*(3), 406–421.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness." *The Journal of the Acoustical Society of America*, *110*(3), 1606–1618. https://doi.org/10.1121/1.1380437

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": the case of French learners of Spanish. *Cognition*, *106*(2), 682–706. https://doi.org/10.1016/j.cognition.2007.04.001

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*(02), 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, N. C. (2006a). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*(1), 1–24.

Ellis, N. C. (2006b). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*(2), 164–194.

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2014a). Second language verb-argument constructions are sensitive to form, function, frequency, contingency, and prototypicality. *Linguistic Approaches to Bilingualism*, *4*(4), 405–431.

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2014b). The processing of verb-argument constructions is sensitive to form, function, frequency, contingency and prototypicality. *Cognitive Linguistics*, *25*(1), 55–98.

Face, T. L. (2000). The role of syllable weight in the perception of Spanish stress. *Hispanic Linguistics at the Turn of the Millenium*, 1–13.

Face, T. L. (2005). Syllable Weight and the Perception of Spanish Stress Placement by Second Language Learners. *Journal of Language and Learning*, *3*, 90–103.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627–635.

Foster, K. (1976). Accessing the mental lexicon. *New Approaches to Language Mechanisms*, 257–287.

Frisch, S. A., Large, N. R., Zawaydeh, B., Pisoni, D. B., & others. (2001). Emergent phonotactic generalizations in English and Arabic. *Typological Studies in Language*, *45*, 159–180.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, *1*(2), 126–152.

Garcia, G. D. (2016). Extrametricality and second language acquisition. *Proceedings of the Annual Meetings on Phonology*, *3*(0). https://doi.org/10.3765/amp.v3i0.3678

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178–186.

Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, *20*(2), 229–252.

Guma, F. (2003). *Perception and production of syllable structure and stress by adult Libyan Arabic speaker acquiring English in the UK* (Ph.D.). Durham University. Retrieved from http://etheses.dur.ac.uk/1329/

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.

Heliel. (1982). Stress Timing In Modern Literary Arabic. *Al-'Arabiyya*, *15*(1), 90–107.

Jangjamras, J. (2011). *Perception and Production of English Lexical Stress by Thai Speakers*. ProQuest LLC.

Jouravlev, O., & Lupker, S. J. (2014). Stress consistency and stress regularity effects in Russian. *Language, Cognition and Neuroscience*, *29*(5), 605–619.

Jouravlev, O., & Lupker, S. J. (2015). Lexical stress assignment as a problem of probabilistic inference. *Psychonomic Bulletin & Review*, *22*(5), 1174–1192.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. MIT Press.

Kager, R. W. J. (1996). On affix allomorphy and syllable counting. *Studia Grammatica*, *41*, 155–171.

Kenstowicz, M. (1983). Parametric Variation and Accent in the Arabic Dialects in Papers from the Nineteenth Regional Meeting, Chicago 21-22 April 1983. In *Papers from the... Regional Meeting. Chicago Linguistic Society Chicago, Ill.* (pp. 205–213).

Ladefoged, P., & Johnson, K. (2010). *A Course in Phonetics* (6 edition). Boston, MA: Cengage Learning.

Lai, Y. (2008). Acoustic Realization and Perception of English Lexical Stress by Mandarin Learners. Retrieved from https://kuscholarworks.ku.edu/handle/1808/4201

Martin, W. A., Church, K. W., & Patil, R. S. (1987). Preliminary analysis of a breadth-first parsing algorithm: Theoretical and experimental results. In *Natural language parsing systems* (pp. 267–328). Springer.

McCarthy, J. J. (1979). On stress and syllabification. *Linguistic Inquiry*, *10*(3), 443–465.

McCarthy, J. J., & Prince, A. S. (1990). Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language & Linguistic Theory*, *8*(2), 209–283.

Mitchell, T. F. (1960). Prominence and Syllabication in Arabic. *Bulletin of the School of Oriental and African Studies*, *23*(2), 369–389.

Mitchell, T. F. (1975). *Principles of Firthian linguistics*. Prentice Hall Press.

Murphy, J., & Kandil, M. (2004). Word-level stress patterns in the academic word list. *System*, *32*(1), 61–74.

Nation, I. S. (2001). *Learning Vocabulary in Another Language Google eBook*. Cambridge University Press.

Nguyen, T. A. T. (2004). Prosodic transfer : the tonal constraints of Vietnamese acquisition of English stress & rhythm. Retrieved from https://espace.library.uq.edu.au/view/UQ:107093

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress 'deafness.' Retrieved from https://books.google.com/books?hl=en&lr=&id=BvkgTd5pxmsC&oi=fnd&pg=PA203&ots=ws_GcZBIta&sig=_YYBv96HmbsOGdil3GUUabQ32Rk

Peperkamp, S., Dupoux, E., & Sebastián-Gallés, N. (1999). Perception Of Stress By French, Spanish, And Bilingual Subjects. In *Proc. EUROSPEECH '99* (pp. 2683–2686).

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*(3), 422–430. https://doi.org/10.1016/j.wocn.2010.04.001

Post da Silveira, A. (2011). Frequency as (dis) advantage to word stress acquisition. In *Proceedings of the ICPhS XVII 2011* (pp. 1634–1637). Hong Kong.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

Ryding, K. C. (2013). *Teaching and learning Arabic as a foreign language: A guide for teachers*. Georgetown University Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Sheikh, I. a. M. (1987). The Interlanguage Stress Phonology of Arab Learners of English: Influence of Universals as Well as Dialect, Distance, and Task (psychotypology, Proficiency).

Shen, X. S. (1990). Ability of learning the prosody of an intonational language by speakers of a tonal language: Chinese speakers learning French prosody. *IRAL-International Review of Applied Linguistics in Language Teaching*, *28*(2), 119–134.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 145.

Ueyama, M. (2000). Prosodic Transfer: An Acoustic Study of L2 English vs. L2 Japanese. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.127.5239

Versteegh, K. (1997). *The Arabic language*. New York: Columbia University Press.

Wang, Q. (2008a). L2 Stress Perception: The reliance on different acoustic cues.

Wang, Q. (2008b). *Perception of English stress by Mandarin Chinese learners of English: An acoustic study* (Thesis). Retrieved from https://dspace.library.uvic.ca//handle/1828/1282

Wang, Q., & Yoon, T.-J. (2008). The transfer of L1 acoustic cues in the perception of L2 lexical stress. *Canadian Acoustics*, *36*(3), 126–127.

Watson, J. C. E. (2007). *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.

Younes, M. A. (1995). On vowel shortening in Palestinian Arabic. *Amsterdam Studies in the Theory and History Of Linguistic Science Series 4*, 157–157.

Yu, V. Y., & Andruski, J. E. (2011). The effect of language experience on perception of stress typicality in English nouns and verbs. *The Mental Lexicon*, *6*(2), 275–301. https://doi.org/10.1075/ml.6.2.03yu

Zuraiq, W. (2005). The production of lexical stress by native speakers of Arabic and English and by Arab learners of English. *University of Kansas*.